

**THE GENETICS OF HUMAN  
C-REACTIVE PROTEIN AND COMPLEMENT  
C3 EXPRESSION IN HEALTH AND DISEASE**

**BENJAMIN RHODES**

Submitted for the degree of Doctor of Philosophy (PhD)

Section of Molecular Genetics and Rheumatology

Imperial College London, 2009

## ABSTRACT

C-reactive protein (CRP) and complement component C3 are both plasma proteins involved in antigen recognition by the innate immune system. They may play a role in the pathogenesis of immune-mediated diseases including systemic lupus erythematosus (SLE) and coronary heart disease. CRP is heritable, but there is uncertainty over the key genetic variants influencing expression. C3 heritability has not been demonstrated, although a single report suggests C3 variation influences serum C3 level and susceptibility to SLE.

This thesis presents an association analysis of *CRP* genotype and serum CRP from four ethnically diverse cohorts. Two cohorts were disease based (UK European and Filipino SLE family cohorts) allowing the additional analysis of *CRP* genotype/disease association, and one (UK rheumatoid arthritis cases) consisted of individuals with active inflammatory disease. *CRP* genotype was associated with serum CRP in all cohorts. Analysis of an African-American cohort (which had an informative pattern of linkage disequilibrium at *CRP*) highlighted a single *CRP* SNP, *rs3091244*, likely to be a key regulator of expression at this locus. No association was observed between *CRP* genotype and SLE.

The UK European SLE cohort was also used to demonstrate C3 heritability and explore associations between C3 genotype, serum C3 and SLE. *Cis*-acting associations were identified with both serum C3 and SLE. The recognition that serum CRP and C3 levels are, in part, genetically determined traits has implications for their use in the monitoring and diagnosis of disease. For neither protein does a single *cis*-acting variant explain most of the variation in serum level: this may be determined by multiple interacting variants, many of which have yet to be identified.

# CONTENTS

|                                                                                 |    |
|---------------------------------------------------------------------------------|----|
| <b>ABSTRACT</b>                                                                 | 2  |
| <b>CONTENTS</b>                                                                 | 3  |
| <b>LIST OF TABLES</b>                                                           | 8  |
| <b>LIST OF FIGURES</b>                                                          | 11 |
| <b>ACKNOWLEDGEMENTS</b>                                                         | 12 |
| <b>STATEMENT OF CONTRIBUTION</b>                                                | 13 |
| <b>ABBREVIATIONS</b>                                                            | 14 |
| <b>CHAPTER 1 - INTRODUCTION</b>                                                 | 17 |
| <b>1.1. The Innate Immune System – An Introduction</b>                          | 18 |
| <b>1.2. Complement</b>                                                          | 19 |
| 1.2.1. <i>Complement – An outline</i>                                           | 20 |
| 1.2.2. <i>Complement – Physiological functions</i>                              | 23 |
| 1.2.3. <i>Complement and systemic lupus erythematosus</i>                       | 24 |
| <b>1.3. Complex Trait Genetics – An Introduction</b>                            | 30 |
| 1.3.1. <i>Genes, genetic variation and phenotype</i>                            | 30 |
| 1.3.2. <i>Confirming the genetic basis of a trait – Family studies</i>          | 32 |
| 1.3.3. <i>SNP discovery in complex genetic disease</i>                          | 33 |
| 1.3.4. <i>Methodological issues in genetic association studies</i>              | 39 |
| <b>1.4. C-reactive Protein</b>                                                  | 42 |
| 1.4.1. <i>CRP – Structure and physiological function</i>                        | 42 |
| 1.4.2. <i>CRP – Gene expression and genetic architecture</i>                    | 45 |
| 1.4.3. <i>CRP – Genetic variation and its relation to gene expression</i>       | 51 |
| 1.4.4. <i>CRP – Functional and genetic role in systemic lupus erythematosus</i> | 60 |
| 1.4.5. <i>CRP – Functional and genetic role in cardiovascular disease</i>       | 61 |
| <b>1.5. The C3 Component of Complement</b>                                      | 65 |
| 1.5.1. <i>C3 – Structure and physiological function</i>                         | 65 |
| 1.5.2. <i>C3 - Gene expression and genetic architecture</i>                     | 68 |
| 1.5.3. <i>C3 – Genetic variation, expression and disease susceptibility</i>     | 71 |
| <b>1.6. Summary and Aims of Thesis</b>                                          | 73 |

|                                                                             |               |
|-----------------------------------------------------------------------------|---------------|
| <b>CHAPTER 2 - MATERIALS AND METHODS</b>                                    | <b>76</b>     |
| <b>2.1 Study Cohorts</b>                                                    | <b>77</b>     |
| 2.1.1. <i>UK white European SLE families (Cohort 1)</i>                     | 77            |
| 2.1.2. <i>Filipino SLE families (Cohort 2)</i>                              | 78            |
| 2.1.3. <i>African-American population cohort (Cohort 3)</i>                 | 79            |
| 2.1.4. <i>UK rheumatoid arthritis cases (Cohort 4)</i>                      | 80            |
| <b>2.2 Genotyping</b>                                                       | <b>81</b>     |
| 2.2.1. <i>Selection of SNPs for genotyping</i>                              | 81            |
| 2.2.2. <i>Genotyping by mass spectrometry – Theoretical aspects</i>         | 82            |
| 2.2.3. <i>Genotyping by mass spectrometry – Practical aspects</i>           | 83            |
| 2.2.4. <i>Genotyping – quality control</i>                                  | 86            |
| <b>2.3 Immunological Assays</b>                                             | <b>86</b>     |
| 2.3.1. <i>CRP quantification – Immunoturbimetry</i>                         | 86            |
| 2.3.2. <i>C3 quantification – Radial immunodiffusion assay</i>              | 87            |
| <b>2.4 Statistical Analysis</b>                                             | <b>88</b>     |
| 2.4.1. <i>Data storage and management</i>                                   | 88            |
| 2.4.2. <i>Heritability estimates</i>                                        | 89            |
| 2.4.3. <i>Data imputation</i>                                               | 90            |
| 2.4.4. <i>Haplotype imputation</i>                                          | 91            |
| 2.4.5. <i>Association analyses</i>                                          | 91            |
| 2.4.6. <i>Admixture analysis</i>                                            | 93            |
| 2.4.7. <i>Model choice analyse</i>                                          | 94            |
| <br><b>CHAPTER 3 - CRP ASSOCIATIONS IN UK SLE FAMILIES</b>                  | <br><b>96</b> |
| <b>3.1. CRP Associations in UK SLE Families – Study Aims</b>                | <b>97</b>     |
| <b>3.2. UK SLE Cohort – Demographics, CRP Distribution and Heritability</b> | <b>98</b>     |
| <b>3.3. CRP Genetic Architecture</b>                                        | <b>100</b>    |
| <b>3.4. Genetic Associations with basal CRP</b>                             | <b>105</b>    |
| 3.4.1. <i>Association of CRP SNPs with basal CRP</i>                        | 105           |
| 3.4.2. <i>Trans-acting genetic effects</i>                                  | 107           |
| 3.4.3. <i>Mixed genetic models</i>                                          | 107           |
| 3.4.4. <i>Influence of SLE affection on CRP associations</i>                | 108           |
| 3.4.5. <i>Haplotypic association with basal CRP</i>                         | 110           |
| <b>3.5. Genetic Contribution to Total CRP Variability</b>                   | <b>111</b>    |
| <b>3.6. CRP Genetic Association with SLE</b>                                | <b>113</b>    |

|                                                                     |         |
|---------------------------------------------------------------------|---------|
| <b>3.7. CRP Genetics in UK SLE Families – Discussion</b>            | 113     |
| 3.7.1. <i>Summary of findings</i>                                   | 113     |
| 3.7.2. <i>A comparison with existing studies – cis-effects</i>      | 114     |
| 3.7.3. <i>A comparison with existing studies – trans-effects</i>    | 115     |
| 3.7.4. <i>A comparison with existing studies – SLE association</i>  | 115     |
| 3.7.5. <i>Key study features and potential limitations</i>          | 116     |
| 3.7.6. <i>Conclusions</i>                                           | 121     |
| <br><b>CHAPTER 4 - CRP ASSOCIATIONS IN FILIPINO SLE FAMILIES</b>    | <br>122 |
| <b>4.1. CRP Associations in Filipino SLE Families – Study Aims</b>  | 123     |
| <b>4.2. Filipino SLE Cohort – Demographics and CRP Distribution</b> | 124     |
| <b>4.3. Filipino CRP Genetic Architecture</b>                       | 125     |
| <b>4.4. Genetic Associations of with Basal CRP</b>                  | 131     |
| 4.4.1. <i>Association of CRP SNPs with basal CRP</i>                | 131     |
| 4.4.2. <i>Trans-acting genetic effects</i>                          | 131     |
| 4.4.3. <i>Mixed genetic models</i>                                  | 134     |
| 4.4.4. <i>Haplotypic association with basal CRP</i>                 | 135     |
| <b>4.5. CRP Genetics in Filipino SLE Families – Discussion</b>      | 137     |
| 4.5.1. <i>Summary of findings</i>                                   | 137     |
| 4.5.2. <i>A comparison with existing studies – cis-effects</i>      | 138     |
| 4.5.3. <i>A comparison with existing studies – trans-effects</i>    | 139     |
| 4.5.4. <i>Key study features and potential limitations</i>          | 141     |
| 4.5.5. <i>Conclusions</i>                                           | 142     |
| <br><b>CHAPTER 5 - CRP GENETICS IN AN AFRICAN-AMERICAN COHORT</b>   | <br>144 |
| <b>5.1. CRP Genetics in an African-American Cohort – Study Aims</b> | 145     |
| <b>5.2. African-American CRP Genetic Architecture</b>               | 146     |
| <b>5.3. African-American Phenotype Models</b>                       | 147     |
| <b>5.4. African-American Genetic models</b>                         | 153     |
| 5.4.1. <i>Primary Bayesian analysis</i>                             | 153     |
| 5.4.2. <i>Secondary Bayesian analyses</i>                           | 154     |
| 5.4.3. <i>Haplotype Association analysis</i>                        | 156     |
| <b>5.5. CRP Genetics in an African-American Cohort – Discussion</b> | 158     |
| 5.5.1. <i>Summary of Findings</i>                                   | 158     |
| 5.5.2. <i>A comparison with existing studies</i>                    | 160     |
| 5.5.3. <i>Key study features and potential limitations</i>          | 161     |
| 5.5.4. <i>Conclusions</i>                                           | 165     |

|                                                                             |                |
|-----------------------------------------------------------------------------|----------------|
| <b>CHAPTER 6 - GENETIC OF ACUTE-PHASE CRP EXPRESSION</b>                    | <b>166</b>     |
| <b>6.1. Genetic of Acute-Phase CRP Expression – Study Aims</b>              | <b>167</b>     |
| <b>6.2. Rheumatoid Arthritis Cohort – Demographics and CRP Distribution</b> | <b>167</b>     |
| <b>6.3. CRP Genetic Architecture</b>                                        | <b>168</b>     |
| <b>6.4. Genetic associations with acute-phase CRP</b>                       | <b>170</b>     |
| <b>6.5. Genetics of acute-phase CRP Expression – Discussion</b>             | <b>174</b>     |
| 6.5.1 <i>Summary of findings</i>                                            | 174            |
| 6.5.2 <i>A comparison with existing studies</i>                             | 175            |
| 6.5.3 <i>Key study features and potential limitations</i>                   | 176            |
| <br><b>CHAPTER 7 - COMPLEMENT C3 GENETICS IN UK SLE FAMILIES</b>            | <br><b>181</b> |
| <b>7.1. Complement C3 Genetics in UK SLE Families – Study Aims</b>          | <b>182</b>     |
| <b>7.2. C3 Distribution and Heritability</b>                                | <b>182</b>     |
| 7.2.1 <i>C3 heritability – Regression estimate</i>                          | 182            |
| 7.2.2 <i>C3 heritability – Replication using Bayesian methodology</i>       | 183            |
| <b>7.3. C3 Genetic Architecture</b>                                         | <b>184</b>     |
| <b>7.4. Genetic Associations with Serum C3</b>                              | <b>187</b>     |
| 7.4.1 <i>Association of C3 SNPs with serum C3</i>                           | 187            |
| 7.4.2 <i>Trans-acting genetic effects</i>                                   | 188            |
| <b>7.5. C3 Genetic Association with SLE</b>                                 | <b>189</b>     |
| 7.5.1 <i>Analysis of genotyped SNPs</i>                                     | 189            |
| 7.5.2 <i>Refinement and replication of association</i>                      | 189            |
| <b>7.6. Complement C3 Genetics in UK SLE Families – Discussion</b>          | <b>190</b>     |
| 7.6.1 <i>Summary of findings</i>                                            | 190            |
| 7.6.2 <i>Associations with serum C3 – discussion</i>                        | 191            |
| 7.6.3 <i>Associations with SLE – discussion</i>                             | 192            |
| 7.6.4 <i>A comparison with existing studies</i>                             | 193            |
| 7.6.5 <i>Key study features and potential limitations</i>                   | 194            |
| 7.6.6 <i>Conclusions</i>                                                    | 197            |
| <br><b>CHAPTER 8 – CONCLUSIONS</b>                                          | <br><b>200</b> |
| <b>8.1. Background to the Work Presented in this Thesis</b>                 | <b>201</b>     |
| <b>8.2. Summary of Study Findings</b>                                       | <b>202</b>     |

|                   |                                                                   |     |
|-------------------|-------------------------------------------------------------------|-----|
| <b>8.3.</b>       | <b>Serum CRP and C3 – General Considerations and Heritability</b> | 203 |
| 8.3.1             | <i>Basal vs. acute-phase levels</i>                               | 203 |
| 8.3.2             | <i>Tissue specific expression</i>                                 | 205 |
| 8.3.3             | <i>Influence of age and gender on serum levels</i>                | 206 |
| 8.3.4             | <i>Heritability</i>                                               | 207 |
| <b>8.4</b>        | <b>Cis-acting Effects on Basal CRP – Discussion</b>               | 209 |
| 8.4.1             | <i>Power</i>                                                      | 210 |
| 8.4.2             | <i>Strong LD across CRP in European populations</i>               | 210 |
| 8.4.3             | <i>Ethnic differences in CRP gene structure</i>                   | 211 |
| 8.4.4             | <i>Problems with a tri-allelic SNP</i>                            | 212 |
| 8.4.5             | <i>Potential for multiple functional SNPs at CRP</i>              | 213 |
| 8.4.6             | <i>CRP functional SNPs – A hypothesis based on thesis data</i>    | 214 |
| 8.4.7             | <i>Cis-effects explain a small proportion of CRP variance</i>     | 218 |
| <b>8.5</b>        | <b>Trans-acting Effects on Serum CRP – Discussion</b>             | 220 |
| 8.5.1             | <i>Trans-acting effects presented in this thesis</i>              | 220 |
| 8.5.2             | <i>Recent GWA studies</i>                                         | 223 |
| 8.5.3             | <i>Trans-acting effects – the current picture and future work</i> | 227 |
| <b>8.6</b>        | <b>CRP, C3 and Disease</b>                                        | 228 |
| <b>8.7</b>        | <b>Concluding Remarks and Future Work</b>                         | 229 |
| <b>REFERENCES</b> |                                                                   | 233 |
| <b>APPENDICES</b> |                                                                   | 270 |
| <b>A</b>          | <b>Sequenom Reagents and Reaction Conditions</b>                  | 271 |
| <b>B</b>          | <b>Radial Immunodiffusion Reagents</b>                            | 275 |
| <b>C</b>          | <b>UK SLE Cohort Patient Questionnaire</b>                        | 277 |
| <b>D</b>          | <b>Publications Resulting from Work Presented in Thesis</b>       | 281 |

## LIST OF TABLES

| CHAPTER 1                                                                                                                                  | PAGE |
|--------------------------------------------------------------------------------------------------------------------------------------------|------|
| <b>Table 1.1</b> Familial clustering of complex genetic diseases                                                                           | 32   |
| <b>Table 1.2</b> Summary of reported <i>P</i> -values for the association of <i>CRP</i> SNPs with basal CRP levels in publications to 2006 | 54   |
| <b>Table 1.3</b> Summary of published <i>CRP</i> haplotype association data                                                                | 55   |
| <br><b>CHAPTER 2</b>                                                                                                                       |      |
| <b>Table 2.1</b> Jeffreys' interpretation of Bayes Factors                                                                                 | 90   |
| <br><b>CHAPTER 3</b>                                                                                                                       |      |
| <b>Table 3.1</b> Effect of age, sex and SLE affection on CRP                                                                               | 99   |
| <b>Table 3.2</b> Typed SNPs demonstrating alleles and rare allele frequencies                                                              | 101  |
| <b>Table 3.3</b> Pairwise linkage disequilibrium between <i>CRP</i> SNPs, calculated in the unrelated parents.                             | 103  |
| <b>Table 3.4</b> Single SNP associations with CRP                                                                                          | 106  |
| <b>Table 3.5</b> Single SNP associations with serum CRP with genetic covariates                                                            | 108  |
| <b>Table 3.6</b> Single SNP associations in the unaffected, unrelated founders                                                             | 109  |
| <b>Table 3.7</b> The association of CRP with <i>CRP</i> haplotypes                                                                         | 110  |
| <b>Table 3.8</b> Pairwise comparison of CRP levels by <i>CRP</i> haplotype                                                                 | 110  |
| <b>Table 3.9</b> The association of CRP with <i>CRP</i> haplotypes (unaffected founders)                                                   | 111  |
| <b>Table 3.10</b> Pairwise comparison of CRP levels by <i>CRP</i> haplotype (unaffected founders).                                         | 112  |
| <br><b>CHAPTER 4</b>                                                                                                                       |      |
| <b>Table 4.1</b> Effect of age, sex and SLE affection on CRP                                                                               | 124  |

|                  |                                                                                               |     |
|------------------|-----------------------------------------------------------------------------------------------|-----|
| <b>Table 4.2</b> | Typed SNPs demonstrating alleles and rare allele frequencies                                  | 126 |
| <b>Table 4.3</b> | Pairwise linkage disequilibrium between <i>CRP</i> SNPs, calculated in the unrelated parents. | 130 |
| <b>Table 4.4</b> | Single SNP associations with CRP.                                                             | 132 |
| <b>Table 4.5</b> | <i>Trans</i> -acting effects with a significant CRP association                               | 133 |
| <b>Table 4.6</b> | Single SNP associations with serum CRP with genetic covariates                                | 135 |
| <b>Table 4.7</b> | The association of CRP with <i>CRP</i> haplotypes                                             | 136 |

## CHAPTER 5

|                  |                                                                             |     |
|------------------|-----------------------------------------------------------------------------|-----|
| <b>Table 5.1</b> | Typed SNPs demonstrating alleles and rare allele frequencies                | 149 |
| <b>Table 5.2</b> | <i>CRP</i> SNPs typed in one cohort only                                    | 150 |
| <b>Table 5.3</b> | Pairwise linkage disequilibrium between <i>CRP</i> SNPs                     | 152 |
| <b>Table 5.4</b> | Bayesian models with > 5% posterior probability.                            | 154 |
| <b>Table 5.5</b> | Models with > 5% posterior probability (vague priors)                       | 155 |
| <b>Table 5.6</b> | Models with > 5% posterior probability, (with genotype*cohort interactions) | 156 |

## CHAPTER 6

|                  |                                                                                    |     |
|------------------|------------------------------------------------------------------------------------|-----|
| <b>Table 6.1</b> | Effect of age sex and smoking on acute-phase CRP                                   | 168 |
| <b>Table 6.2</b> | Typed SNPs demonstrating alleles and rare allele frequencies                       | 169 |
| <b>Table 6.3</b> | Frequency of common <i>CRP</i> haplotypes                                          | 169 |
| <b>Table 6.4</b> | Single SNP associations with CRP                                                   | 170 |
| <b>Table 6.5</b> | The association of CRP with <i>CRP</i> haplotype                                   | 171 |
| <b>Table 6.6</b> | Single SNP associations with CRP, controlling for ESR or DAS                       | 173 |
| <b>Table 6.7</b> | The association of CRP with <i>CRP</i> haplotype, controlling for ESR              | 173 |
| <b>Table 6.8</b> | The association of acute-phase CRP with <i>CRP</i> haplotype, controlling for DAS. | 174 |

## CHAPTER 7

|                  |                                                                                                |     |
|------------------|------------------------------------------------------------------------------------------------|-----|
| <b>Table 7.1</b> | Effect of age, sex and SLE affection on serum C3                                               | 183 |
| <b>Table 7.2</b> | Typed SNPs demonstrating alleles and rare allele frequencies                                   | 184 |
| <b>Table 7.3</b> | Pairwise linkage disequilibrium between <i>C3</i> SNPs, calculated from the unrelated parents. | 185 |
| <b>Table 7.4</b> | Significant <i>C3</i> SNP associations with serum C3.                                          | 187 |
| <b>Table 7.5</b> | Single <i>IL1B</i> SNP associations with serum C3.                                             | 188 |

## CHAPTER 8

|                  |                                                                                       |     |
|------------------|---------------------------------------------------------------------------------------|-----|
| <b>Table 8.1</b> | Current CRP heritability estimates.                                                   | 208 |
| <b>Table 8.2</b> | Proportion of CRP variance explained by <i>cis</i> -acting effects in current studies | 219 |
| <b>Table 8.3</b> | Genes associated with CRP expression in current GWA studies ( <i>P</i> -values shown) | 224 |
| <b>Table 8.4</b> | Most significant SNPs at <i>CRP</i> in the current GWA studies                        | 226 |

## LIST OF FIGURES

| CHAPTER 1            |                                                                                  | PAGE |
|----------------------|----------------------------------------------------------------------------------|------|
| <b>Figure 1.1</b>    | The complement cascade summarising three pathways of activation.                 | 21   |
| <b>Figure 1.2</b>    | A classical approach to the identification of functional genetic variants.       | 34   |
| <b>Figure 1.3</b>    | <i>CRP</i> in its genomic context on chromosome 1.                               | 46   |
| <b>Figure 1.4</b>    | Cross-species DNA sequence conservation at <i>CRP</i> .                          | 48   |
| <b>Figure 1.5</b>    | Transcription factor binding sites at the <i>CRP</i> promoter.                   | 49   |
| <b>Figure 1.6</b>    | The position of <i>CRP</i> SNPs in the reported literature.                      | 53   |
| <b>Figure 1.7</b>    | <i>C3</i> in its genomic context on chromosome 19.                               | 69   |
| <br><b>CHAPTER 3</b> |                                                                                  |      |
| <b>Figure 3.1</b>    | <i>CRP</i> haplotypes and LD structure.                                          | 102  |
| <b>Figure 3.2</b>    | Cladogram of <i>CRP</i> haplotypes.                                              | 104  |
| <br><b>CHAPTER 4</b> |                                                                                  |      |
| <b>Figure 4.1</b>    | Filipino <i>CRP</i> haplotypes.                                                  | 127  |
| <b>Figure 4.2</b>    | Graphical representation of LD at the Filipino <i>CRP</i> locus.                 | 128  |
| <b>Figure 4.3</b>    | Cladogram of <i>CRP</i> haplotypes.                                              | 129  |
| <br><b>CHAPTER 5</b> |                                                                                  |      |
| <b>Figure 5.1.</b>   | African-American <i>CRP</i> haplotypes.                                          | 151  |
| <b>Figure 5.2.</b>   | Cladogram of the phylogenetic relationships between African-American haplotypes. | 157  |
| <br><b>CHAPTER 7</b> |                                                                                  |      |
| <b>Figure 7.1</b>    | <i>C3</i> haplotype structure.                                                   | 186  |

## **ACKNOWLEDGEMENTS**

Firstly I thank my PhD supervisor, Professor Tim Vyse, for his endless enthusiasm and unfailing support throughout the duration of the work presented in this thesis.

I would like to thank a number of colleagues for their help and advice with specific aspects of the research presented here: Deborah Cunningham Graham for teaching me genotyping using Sequenom; Peter Norsworthy for teaching me the radial immunodiffusion assays; John Meek and his colleagues in the department of clinical chemistry at Hammersmith Hospital for facilitating the quantification of serum CRP samples in their laboratory and John Whittaker, David Morris, Dorothea Nitsch and Saowalak Hunnangkul for their endless support with all the statistical aspects of this work. I would also like to acknowledge collaborators who assisted with this work by allowing access to patient/DNA cohorts: Philip de Jager, Jorge Oksenburg, Carlos Mendes de Leon and Sophia Steer

I thank the Arthritis Research Campaign for funding this research through a Clinical Research Fellowship (Reference 17127)

Finally I would like to thank my wife Joanne for her ongoing support, and my daughters Amelia and Jessica, both of whom arrived during the time of this research.

## STATEMENT OF CONTRIBUTION

This thesis is the result of my own work. Some of the data collection and analysis was, however, performed in collaboration with colleagues as follows:

The genotyping presented in chapter 3 was run concurrently with other projects within the laboratory. The work of processing these hME reactions was shared between me and a fellow PhD student, Michelle Fernando. The final Clean Resin and SpectroCHIP stages were conducted by Imperial College Genome Centre technical staff.

Serum C3 data utilised in the heritability pilot study, presented in section 7.2.1, were already available as a result of previous work in the laboratory. Further C3 quantification for the analyses presented in sections 7.2.2 and 7.3.1, was performed as part of this project, with assays shared between myself and an MSc student, Lien-Cheng Hsiao.

The more complex analyses presented in this thesis were performed in close collaboration with statistical colleagues. Saowalak Hunnangkul, a PhD student at the Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, performed the Bayesian variance components models of heritability and association presented in chapter 7. David Morris, a post-doctoral statistician in our laboratory performed the C3 data imputation presented in section 7.4.7. His expertise in using WinBUGs was also essential in enabling me to perform the relevant association and model choice analyses run using this statistical package.

Other minor contributions to this thesis are acknowledged in the text.

## ABBREVIATIONS

|       |                                                                                          |
|-------|------------------------------------------------------------------------------------------|
| ACR   | American College of Rheumatology                                                         |
| AIC   | Akaike Information Criteria                                                              |
| AMD   | Age Related Macular Degeneration                                                         |
| ANCA  | Anti-Neutrophil Cytoplasmic Antibody                                                     |
| ANOVA | Analysis of Variance                                                                     |
| BIC   | Bayesian Information Criteria                                                            |
| BMI   | Body Mass Index                                                                          |
| bp    | Base pairs                                                                               |
| C/EBP | CCAAT/Enhancer Binding Protein                                                           |
| CEPH  | Centre d'Etude du Polymorphisme Humain (HapMap population of Northern European ancestry) |
| CFT   | Complement Fixation Test                                                                 |
| CHAP  | Chicago Health and Aging Project                                                         |
| CHB   | HapMap population of Han Chinese origin                                                  |
| CHD   | Coronary Heart Disease                                                                   |
| CI    | Confidence Interval                                                                      |
| CNV   | Copy Number Variation                                                                    |
| CR1-4 | Complement receptor 1-4                                                                  |
| CRP   | C-reactive Protein                                                                       |
| DAS28 | Disease Activity Score (28 joints)                                                       |
| df    | Degrees of freedom                                                                       |
| DNA   | Deoxyribonucleic acid                                                                    |
| dNTP  | Deoxyribonucleotide triphosphate                                                         |

|               |                                                            |
|---------------|------------------------------------------------------------|
| dsDNA         | Double Stranded DNA                                        |
| EMSA          | Electrophoretic Mobility Shift Assay                       |
| ESR           | Erythrocyte Sedimentation rate                             |
| Fc $\gamma$ R | Gamma Immunoglobulin Receptor                              |
| FXR           | Farnesoid X Receptor                                       |
| GWA           | Genome-Wide Association                                    |
| $h^2$         | Heritability (strict sense)                                |
| hME           | Homogeneous massEXTEND reaction                            |
| HNF           | Hepatocyte Nuclear Factor                                  |
| HPLC          | High Performance Liquid Chromatography                     |
| HWP           | Hardy-Weinberg Equilibrium Probability                     |
| IFN           | Interferon                                                 |
| IL            | Interleukin                                                |
| kb            | kilobases                                                  |
| $\lambda_s$   | Sibling Risk Ratio                                         |
| LD            | Linkage Disequilibrium                                     |
| LDL           | Low density lipoprotein                                    |
| MALDI-TOF     | Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight |
| MASP          | Mannose-binding lectin associated serine protease          |
| MBL           | Mannose-binding lectin                                     |
| MCMC          | Markov Chain Monte Carlo                                   |
| MG            | Macroglobulin                                              |
| MHC           | Major Histocompatibility Complex                           |
| MI            | Myocardial Infarction                                      |
| MPGN          | Membranoproliferative Glomerulonephritis                   |

|                   |                                                                                 |
|-------------------|---------------------------------------------------------------------------------|
| MS                | Multiple Sclerosis                                                              |
| NOD               | Nucleotide-binding oligomerisation domain                                       |
| $P_{\text{marg}}$ | Marginal Probability (Bayesian)                                                 |
| $P_{(M/D)}$       | Probability of a Model Given the Data (i.e. the Bayesian Posterior Probability) |
| PR3               | Proteinase 3                                                                    |
| qTDT              | Quantitative Transmission Disequilibrium Test                                   |
| RA                | Rheumatoid Arthritis                                                            |
| RFLP              | Restriction Fragment Length Polymorphism                                        |
| RID               | Radial Immunodiffusion Assay                                                    |
| SAP               | Shrimp Alkaline Phosphatase                                                     |
| s.d.              | Standard Deviation                                                              |
| SLE               | Systemic Lupus Erythematosus                                                    |
| Sm                | Smith antigen                                                                   |
| SNP               | Single Nucleotide Polymorphism                                                  |
| snRNP             | Small Nuclear Ribonucleoprotein                                                 |
| SPSS              | Statistical Package for the Social Sciences                                     |
| T1D               | Type 1 Diabetes Mellitus                                                        |
| TDT               | Transmission Disequilibrium Test                                                |
| TGF               | Transforming Growth Factor                                                      |
| TNF               | Tumour Necrosis Factor                                                          |
| TE                | Tris-EDTA                                                                       |
| UK                | United Kingdom                                                                  |
| 3'UTR             | 3-Untranslated Region                                                           |
| YRI               | HapMap population of Yoruban (West African origin)                              |

## **CHAPTER 1**

# **INTRODUCTION**

## 1.1: The Innate Immune System – An Introduction

Classically, the components of the human immune system are divided into two broad categories: the adaptive immune system and the innate immune system. The adaptive immune system evolved relatively recently in chordate species (Pancer & Cooper 2006). It is characterised by high specificity receptors generated by the somatic recombination and diversification of receptor gene segments. Following antigen recognition a sophisticated response is orchestrated, resulting either in the activation of highly effective antigen clearance mechanisms or the down-regulation of immune responses resulting in antigen tolerance. A key function of the adaptive immune system is the generation of immunological memory: the production of antigen-specific memory cells that enable a more rapid and robust response when specific antigen is encountered for a second time (Murphy *et al* 2008).

In contrast, the basic components of the innate immune system are phylogenetically ancient, with human molecule homologues widespread in protostome species (Danilova 2006). Foreign antigen is recognised through germ-line encoded receptors which, unlike those of the adaptive immune system, are often non-specific and capable of recognising numerous antigens found on the surface of pathogenic micro-organisms or altered human cells. They are expressed on prevalent immune cells (such as neutrophils and macrophages) or as circulating plasma proteins, and they trigger a rapid immune response (Murphy *et al* 2008).

The innate and adaptive immune systems do not, however, function in isolation, but work in tandem: recognition of antigen by the innate immune system can recruit and prime an adaptive response, and in turn adaptive components can trigger highly effective innate bacterial killing mechanisms (Murphy *et al* 2008).

For convenience, the innate immune system can be divided into cellular and humoral arms. Recognition of antigen by the cellular branch is mediated by cell-surface or intracellular receptors recognising unique pathogen-associated molecular patterns, including protein, carbohydrate and nucleic acid antigens. The Toll-like receptor family is the best described, but there are others including the NOD/NALP receptor family, NK receptors and Scavenger receptors (Beutler 2008; Papadimitraki *et al* 2007; Peiser *et al* 2002; Petrilli *et al* 2007; Vivier *et al* 2008). These receptors are among the first components of the immune system to recognise foreign antigen. Cellular responses triggered in response to receptor binding, for example the release of type 1 interferons (IFN) or interleukin-1 $\beta$  (IL-1 $\beta$ ), are critical in directing the course of the subsequent innate and adaptive immune response (Lee & Kim 2007). The humoral branch of the innate immune system will be discussed in the next section.

## **1.2: Complement**

A key component of the humoral arm of the innate immune system is a network of proteins comprising the complement system, so called because they “complement” the function of antibody (Walport 2001). The central components of this system are zymogen proteins that circulate in the plasma in an inactive state. Activation of the early complement components initiates a cascade of protein activation, with each component in turn activated by proteolytic cleavage, resulting in the generation of molecules that are biologically active and contribute to host defence.

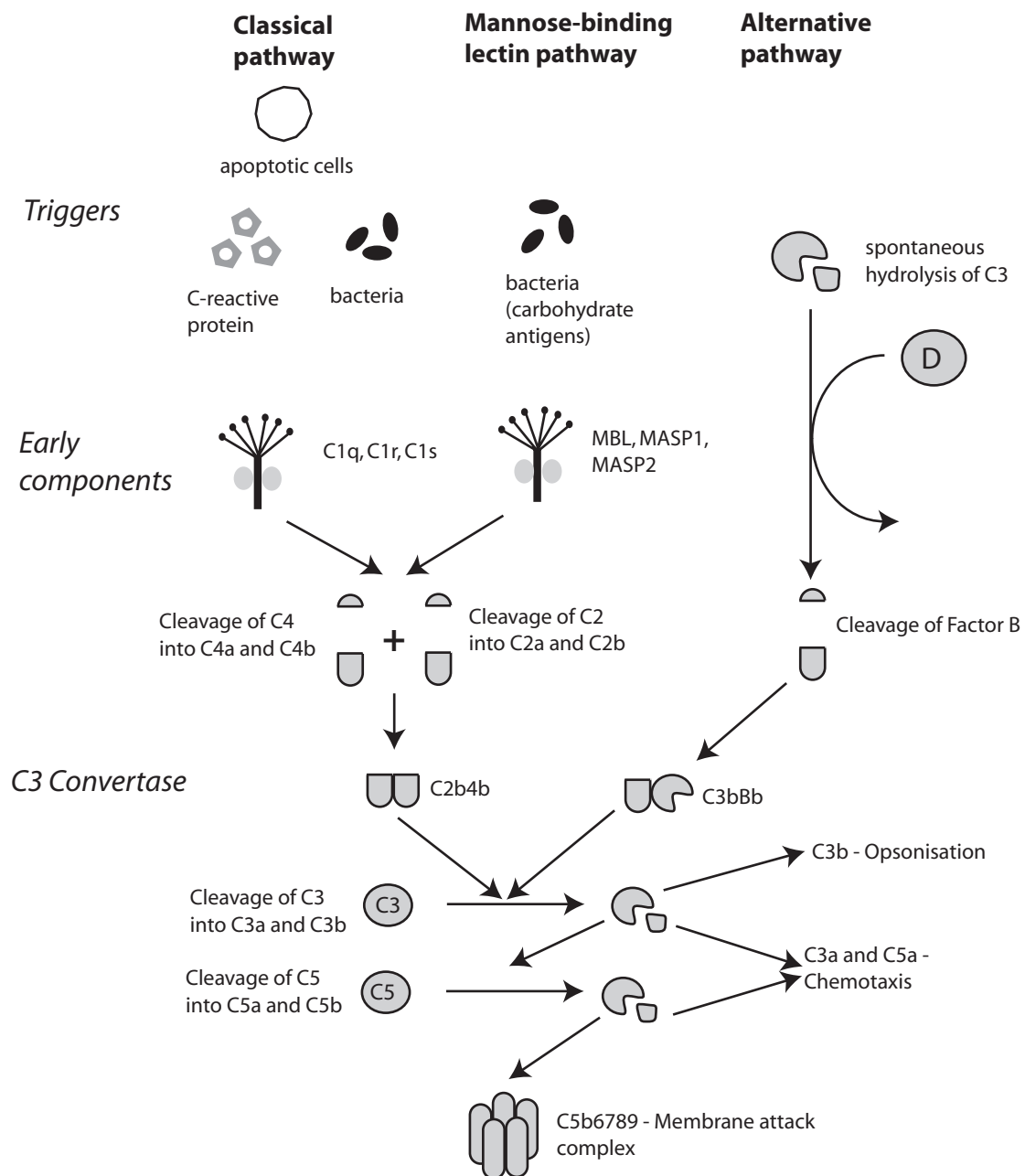
### *1.2.1: Complement – An Outline.*

An outline of the central components of the complement cascade is presented in **figure 1.1**.

The sequence of activation and functions of these central components is now well established and a number of comprehensive texts and review articles are referenced. There are three pathways of activation, each culminating in the cleavage of C3 (Frank & Fries 1991; Muller-Eberhard 1988; Murphy *et al* 2008; Walport 2001). The first component of the classical pathway is C1q, a member of the collectin family of proteins. C1q is either activated directly by binding certain antigens on pathogens or altered cells, or by binding cross-linked immunoglobulin (i.e. immunoglobulin complexed with antigen, either on cell surfaces or in immune complexes). Finally C1q can be activated by another important component of the innate immune system called C-reactive protein (CRP). CRP is a small molecule that recognises a wide range of antigens found on bacteria or altered human cells, and it will be discussed in detail in subsequent sections. Following antigen recognition, C1q binds the zymogens C1r and C1s, and the C1 complex in turn activates components C4 and C2 by proteolytic cleavage. Together the large fragments of C4 and C2 (C4b2b) form a C3 convertase which is able to activate C3, the pivotal molecule of the cascade.

The second pathway of complement activation is the Mannose-binding lectin (MBL) pathway (Endo *et al* 2006; Turner & Hamvas 2000). MBL is a collectin homologous to C1q. It binds mannose residues, and certain other sugars, which lie exposed on the surface of many bacterial cell walls. MBL forms a complex with MBL-associated serine proteases, MASP1 and MASP2, which are homologous to C1r and C1s, and the activated MBL complex is able to cleave C4 and C2 in the same way as the classical pathway.

**Figure 1.1.** The complement cascade summarising three pathways of activation



MBL – Mannose binding lectin, MASP – MBL-associated serine protease.

The final pathway of complement activation is the alternative pathway (Harboe & Mollnes 2008). Circulating C3 undergoes a low rate of spontaneous hydrolysis to C3(H<sub>2</sub>O), and hydrolysed C3 is able to bind Factor B, a protein homologous to C2. Bound factor B can undergo cleavage by factor D resulting in C3(H<sub>2</sub>O)Bb, which has C3 convertase activity. Unbound C3b is rapidly inactivated by hydrolysis but if any C3b produced through the alternative pathway does bind a cell surface it can bind factor B again and, following cleavage by factor D, form a stable C3 convertase, C3bBb. It is critical to host survival that spontaneously produced C3b does not bind to normal host cells, and a series of circulating and membrane bound regulatory proteins have evolved to inactivate C3b when deposition on normal cells occurs. A full discussion of these is beyond the scope of this thesis.

C3 will be discussed further in later sections but, in summary, activation results in the generation of biologically active fragments that form the effector arm of complement system. C3 can exist in a number of conformational states, and can be cleaved into a number of fragments with distinct biological activity. Fragment C3a acts via specific receptors to mediate local inflammatory responses including increased vascular permeability, upregulation of endothelial adhesion molecules, and recruitment of inflammatory cells. Fragment C3b remains bound to the surface on which it was activated and facilitates uptake by phagocytic cells carrying specific complement receptors, a process known as opsonisation. Proteolytic alteration of C3b results in the formation of fragments iC3b and C3dg, which also function as opsonins, and C3c. The incorporation of C3b into the C3 convertase (to form either C4b2b3b or C3bBb3b) generates a complex which cleaves C5 into C5a and C5b. C5a has potent pro-inflammatory effects similar to C3a, while C5b initiates the formation of the membrane attack complex (consisting of complement components C5b, C6-8 and multiple

copies of C9) which is capable of killing cells directly by forming a membrane destabilising pore, leading to cell lysis (Haas & van 2007; Muller-Eberhard 1986).

### *1.2.2: Complement – Physiological Functions*

Complement is highly effective at mediating bacterial killing through its three effector arms: opsonisation of bacteria, recruitment of inflammatory cells and bacterial lysis via the membrane attack complex. The importance of this role is highlighted by the increased susceptibility to infection seen in individuals with hereditary complement deficiencies. Primary homozygous C3 deficiency is exceptionally rare, but is characterised by recurrent bacterial infection from organisms such as *Haemophilus influenzae*, *Streptococcus pneumoniae* and *Neisseria meningitidis* (Reis *et al* 2006). Deficits in the membrane attack complex components are particularly associated with susceptibility to infection by *Neisseria meningitidis* (Ross & Densen 1984; Wurznner *et al* 1992).

In addition to its role in bacterial clearance, complement promotes the clearance of unwanted host cells, or cellular components in the form of circulating immune complexes. Apoptosis or “programmed cell death” is a key physiological mechanism by which unwanted host cells die, leading to rapid clearance by phagocytosis without a damaging inflammatory response. The cellular signals that control this process are incompletely understood but it is clear that complement has a crucial role (Pittoni & Valesini 2002). Cells undergoing apoptosis exhibit characteristic changes to the lipid and protein constituents of their cell walls. Both the C1q and MBL components of complement are able to bind these altered cell walls leading to complement activation and deposition of iC3b on the surface of the apoptotic cell (Hansen & Holmskov 1998; Korb & Ahearn 1997; Mevorach *et al* 1998a). The iC3b fragment of C3 is

able to mediate the uptake of these cells by phagocytes bearing complement receptor 3 (CR3), but does not form a C5 convertase, hence the pro-inflammatory and direct cytotoxic aspects of complement function are limited. Similarly CRP binds avidly to apoptotic cells and activates the classical complement pathway, but again prevents the formation of the terminal membrane attack complex (Gershov *et al* 2000).

Complement is also important in the clearance of cellular components released into the circulation from dead bacteria or host cells. Typically these antigens bind to specific immunoglobulin to form small immune complexes, either within tissues or within the circulation. Complement limits the physical size of these immune complexes, and maintains their solubility, thus preventing immune complex mediated tissue injury (Miller & Nussenzweig 1975; Naama *et al* 1984; Schifferli *et al* 1982; Schifferli *et al* 1985). Secondly, complement mediates the uptake of immune complexes by phagocytic cells, expressing CR3, which recognise iC3b deposited on the complexes (Arnaout *et al* 1983; Brown *et al* 1988). Immune complexes located in tissues are taken up directly by phagocytes, while immune complexes in the circulation are initially transported to the spleen by erythrocytes which express CR1 and bind C3b on the complex. Phagocytes within the spleen then efficiently remove and destroy immune complexes from the surface of erythrocytes by the mechanism described above (Davies *et al* 1993; Schifferli *et al* 1988).

### *1.2.3: Complement and Systemic Lupus Erythematosus*

Systemic lupus erythematosus is the archetypal systemic autoimmune disease. It is characterised by presence of pathogenic autoantibodies directed against ubiquitous cellular components, including nucleoproteins and negatively charged membrane phospholipids.

These antibodies may be directly pathogenic, or may cause damage when deposited in tissues in the form of immune complexes. Clinically the result of this immunological dysfunction is a disease characterised by inflammation in multiple organs including kidney, skin, joints and central nervous system (Lahita 2004). SLE is a heterogeneous disease, presenting differently from patient to patient, and with no single clinical or immunological feature required to make a formal diagnosis (Tan *et al* 1982). For at least 30 years we have known that there is a strong genetic basis to SLE: a disease concordance of 2-5% in dizygotic compared with 24-57% in monozygotic twins and a sibling risk ratio ( $\lambda_s$ ) of 20-29 (Alarcon-Segovia *et al* 2005; Block *et al* 1975; Deapen *et al* 1992; Hochberg 1987). Progress in understanding the immune pathology underlying SLE has been slow, and it is only over the last five years or so that we have developed a more detailed understanding of which are the key genetic variants influencing disease susceptibility (Rhodes & Vyse 2007; Rhodes & Vyse 2008).

An apparent paradox of SLE pathogenesis is how a strong autoimmune response is generated against antigens that are primarily located in the intracellular compartment. A convincing explanation for this arose from the observation that cells undergoing apoptosis display many of these intracellular antigens on their cell surface. It was also noted that mice injected with apoptotic cells develop lupus-like antibodies against ssDNA and negatively charged membrane phospholipids (Casciola-Rosen *et al* 1994; Mevorach *et al* 1998b). Apoptotic cells may therefore be the source of auto-antigen in SLE, and any impairment of apoptotic cell clearance could predispose to autoimmunity. This hypothesis has become known as the “waste-disposal hypothesis”. It is supported by a number of lines of evidence including the direct observation of impaired phagocytosis of apoptotic cells by macrophages from SLE patients, and the high prevalence of SLE-like illness in individuals deficient in C1q and other

early components of the classical complement cascade (Baumann *et al* 2002; Herrmann *et al* 1998; Pickering & Walport 2000).

The mechanism by which an impairment of complement associated apoptotic cell clearance leads to autoimmunity is unknown. Normally the uptake of iC3b opsonised apoptotic cells by dendritic cells is generally accompanied by a suppression of inflammation and a release of anti-inflammatory mediators, without dendritic cell maturation (Morelli *et al* 2003; Verbovetski *et al* 2002; Voll *et al* 1997). It is possible that a delay in clearance allows apoptotic cells to progress into more inflammatory necrotic cells, and that the resulting pro-inflammatory environment allows dendritic cell maturation and the expression of a pro-inflammatory cytokine response, triggering a break in tolerance (Manderson *et al* 2004).

There is considerable evidence, from murine models and *in vitro* experiments, to demonstrate that complement abnormalities play a central role in the impaired uptake of apoptotic cells in SLE: this literature has been reviewed elsewhere (Manderson *et al* 2004). The important role played by complement is also emphasised by the high prevalence of lupus-like illness in individuals with complement component deficiencies (Manderson *et al* 2004; Pickering & Walport 2000). Homozygous *C1q* deficiency is extremely rare, but over 90% of individuals develop moderate to severe lupus (Walport *et al* 1998). Similarly the even rarer deficiencies of C1r and C1s are associated with high risk of disease (Pickering & Walport 2000). A variety of *MBL* promoter SNPs have been associated with reduced *MBL* expression, while a structural exon-1 SNP has been associated with SLE (Garred *et al* 2001; Huang *et al* 2003).

Moderately severe SLE develops in up to 33% of patients with homozygous C2 deficiency (Pickering & Walport 2000). A *C2Q\*0* null allele is carried on a rare DRB1\*1501 containing

MHC haplotype on chromosome 6 and is usually due to a 28 base pair deletion creating a premature stop codon (Johnson *et al* 1992). Heterozygous *C2* deficiency does not seem to confer an increased risk (Hartung *et al* 1989). Although *C4* deficiency is also associated with SLE, the genetics of *C4* is particularly complex making it difficult to conclusively ascribe an effect to this gene. *C4* is transcribed from two MHC genes, *C4A* and *C4B*, which have slightly different sequences and function; *C4A* being better able to bind immune complexes (Schifferli *et al* 1986). The *C4* genes are inherited in a discrete “RCCX module”, which contains one *C4* gene (either *C4A* or *C4B*) along with three neighbouring genes. The Class III *MHC* carries between one and four copies of this module; hence each diploid genome has between two and eight *C4* genes, which may be either *C4A* or *C4B* (Yang *et al* 2007). Carrying less than 2 copies of *C4A* has been identified as a risk factor for SLE (Yang *et al* 2007). *MHC* sequencing, however, reveals that the common AH8.1 (DR3), SLE associated haplotype contains a monomodular RCCX unit with a single *C4B* gene and no *C4A* (Stewart *et al* 2004). It is therefore possible that the observed *C4A* association is present due to linkage disequilibrium between the gene copy number and functional variants in other *MHC* genes. We will only be able to answer this question when we have a comprehensive analysis incorporating both *C4A* copy number analysis and a dense *MHC* SNP map.

Interestingly, although individuals with complete *C3* deficiency are at high risk of bacterial infection they seem to develop less severe autoimmune features. Nonetheless, the most comprehensive review to date reports autoimmune manifestations (most commonly a rash) in 26% of patients with complete *C3* deficiency (Reis *et al* 2006). The majority of individuals with severe manifestations (including mesangiocapillary glomerulonephritis or full-blown SLE) are Japanese, suggesting that disease severity in this population may be determined by an interaction between *C3* deficiency and another, unknown, genetic locus (Reis *et al* 2006).

Complete deficiency of any of the complement components is extremely rare and unlikely to contribute to the genetic basis of the relatively common sporadic SLE. In the majority of cases SLE susceptibility is likely to be determined by relatively common variants that are found throughout the population, each of which only contributes modestly to disease risk; this has been termed the “common disease, common variant” hypothesis (Becker 2004). Although complete complement deficiencies may not be relevant in this context they do highlight that complement proteins, and molecules associated with complement activation may be suitable gene candidates in the search for common variants associated with SLE. The recent identification of *ITGAM* (integrin  $\alpha_M$  – the  $\alpha$ -subunit of complement receptor 3) as a major susceptibility gene in human SLE emphasises this point (Nath *et al* 2008).

In addition to a specific role for complement there is evidence that the innate immune system plays a wider role in SLE pathogenesis; much of this evidence has come from genetic studies. For example, there is biological plausibility for a role for the toll-like receptors that I introduced in section 1.1. TLRs 7 and 9 bind single stranded RNA and unmethylated CpG DNA respectively. Small sections of CpG rich, hypomethylated DNA may be released during apoptosis while endogenous ssRNA may be found in the context of small nuclear ribonucleoprotein (snRNP) immune complexes (Barrat *et al* 2005; Krystosek 1999). *In vitro* evidence suggests that ssRNA and CpG DNA are capable of stimulating autoreactive B-cells via TLR 7 and 9 respectively (Lau *et al* 2005; Viglianti *et al* 2003). In murine models *tlr9* knockout has been associated with the development of autoimmunity while the SLE enhancing effects of the *Yaa* region of the lupus-prone BXSB mouse may, in part, be due to a *tlr7* gene duplication (Christensen *et al* 2006; Pisitkun *et al* 2006). To date, however, no human studies have demonstrated an association between SLE and variation at *TLR7* or *TLR9*.

(De Jager *et al* 2006; Hur *et al* 2005; Kelley *et al* 2007; Ng *et al* 2005). There is also strong evidence for a role for type 1 interferons (IFN- $\alpha$  and IFN- $\beta$ ), cytokines released from innate immune cells as part of the early response to antigen. It has been noted that there is a general up-regulation of interferon inducible genes in SLE (Baechler *et al* 2003; Bennett *et al* 2003). Subsequently genetic studies have shown an association between variation in a number of interferon pathway genes and SLE, most notably *IRF5*, the gene encoding interferon regulatory factor-5 (Kozyrev & Alarcon-Riquelme 2007).

Before leaving this section it should also be noted that complement levels fluctuate in SLE as part of the disease process itself, presumably because the rate of complement activation by immune complexes exceed the rate of synthesis. A decrease in the levels of the complement components C4 and, particularly, C3 are often observed in patients with SLE, although the correlation with disease activity varies from patient to patient (Jonsson *et al* 1995; Valentijn *et al* 1985).

In the work presented in this thesis I will concentrate on two specific components of the innate immune system outlined above. Firstly CRP, the non-specific pattern recognition molecule which is a potent activator of the classical complement cascade and, secondly, C3 itself, the pivotal molecule at the centre of the complement cascade. In particular I will consider the possibility that common variation in the *CRP* or *C3* genes may influence gene expression, and that this, in turn, may predispose to the development of disease.

In the following sections I will therefore introduce some basic genetic concepts that are important for the understanding of genetic association studies, before outlining our current understanding of both the biological functions of CRP and C3, and the genetic basis of their expression.

### 1.3: Complex Trait Genetics – An Introduction

In this section I will summarise our current understanding of human genetics, genetic variation and the association between genetic variation and human biological traits. I will also discuss relevant methodological issues surrounding the design of genetic association studies.

#### *1.3.1: Genes, Genetic Variation and Phenotype*

The human genome consists of approximately 3.2 billion base pairs of DNA, and encodes 20-25,000 genes capable of translation to a protein product (McPherson *et al* 2001; The International Human Genome Sequencing Consortium 2004). With a few exceptions every somatic cell carries two copies of this genome, divided between 22 homologous pairs of autosomes and two sex chromosomes.

Apart from monozygotic (identical) twins, the genome varies from person to person, and this structural genetic variation takes a number of different forms. Perhaps best described is the single nucleotide polymorphism (SNP). A SNP is present when, within the population, two or more different nucleotides (alleles) are observed at the same position in the genome. The traditional definition of a “polymorphism” states that the frequency of the minor (or rare) allele in the population has to exceed 0.01. It has been estimated that there are 9-10 million SNPs with a minor allele frequency  $<0.05$ , over 3.1million of which have been characterised in four human populations by the International HapMap Project ([www.hapmap.org](http://www.hapmap.org)) (Frazer *et al* 2007). SNPs arise as a result of rare mutations due to DNA damage or defective DNA replication. They can alter the function of the gene (and hence increase disease susceptibility) in a number of ways. Non-synonymous SNPs in gene exons can alter the amino acid

sequence, and hence function, of the final protein product. Alternatively SNPs may affect gene regulatory elements, perhaps altering promoter function, gene splicing, mRNA stability or translational efficiency (Knight 2005). The majority of SNPs lie in intergenic regions and may have no function; they have arisen by chance and persist because they are not subject to the forces of natural selection.

It should be noted that there are other sources of genetic variation. In particular we are increasingly realising that the large scale duplication or deletion of DNA segments known as copy-number variation (CNV) is a major source of human genetic variation, affecting up to 12% of the genome (Redon *et al* 2006). The study of CNV is in its infancy and will not be considered further here.

All humans differ from each other; from observable phenotypes such as height or weight, to measurable biological parameters such as plasma protein levels; from behavioural characteristics to the likelihood of developing disease. To a greater or lesser extent the variation in each of these phenotypic characteristics is determined by underlying genetic variation of the kind introduced above. Some phenotypes (so called Mendelian traits) are determined by variation of a single gene, although few are entirely unmodified by background genetic effects as well (Nadeau 2001). Mendelian traits are characterised by simple patterns of inheritance within families, to the extent that it is possible to predict with a reasonable level of certainty an individual's risk of expressing that phenotype. The majority of common human diseases and all quantitative biological traits, however, are determined by variation at multiple genes, with trait expression determined by a potentially complex interaction between genes and between genes and the environment: these are termed complex genetic traits. Complex genetic traits are characterised by complex patterns of inheritance;

typically there is modest trait concordance among first degree relatives, with a rapid tailing off of concordance in more distantly related individuals (Risch 1990). Identifying the variants underlying complex genetic traits presents a major challenge in modern genetic research.

### 1.3.2: Confirming the genetic basis of a trait – Family Studies

The clustering of a disease, or phenotypic trait, within families suggests that the trait has an underlying genetic basis, even if the causal mutation is unknown (Burton *et al* 2005). A common approach to quantifying this is to compare trait concordance in monozygotic (genetically identical) twins to that observed in dizygotic (on average 50% identical) twins (Martin *et al* 1997). An alternative measure is the sibling risk ratio,  $\lambda_s$  (i.e. the incidence of disease in the siblings of affected individuals compared with the incidence of disease in the general population). As an example, **Table 1.1** outlines the twin concordance and  $\lambda_s$  for a number of diseases with a proposed genetic basis.

**Table 1.1.** Familial clustering of complex genetic diseases

| Disease | Monozygotic concordance (%) | Dizygotic concordance (%) | Sibling risk ratio ( $\lambda_s$ ) |
|---------|-----------------------------|---------------------------|------------------------------------|
| SLE     | 24-57                       | 2-5                       | 20-29                              |
| MS      | 25                          | 0-5                       | 20                                 |
| T1D     | 30-50                       | 0-13                      | 15                                 |
| RA      | 12-15                       | 3-4                       | 5-10                               |
| CHD     | 39(♂) 44(♀)                 | 26(♂) 14(♀)               | 2-3                                |

Abbreviations, and study references are as follows:

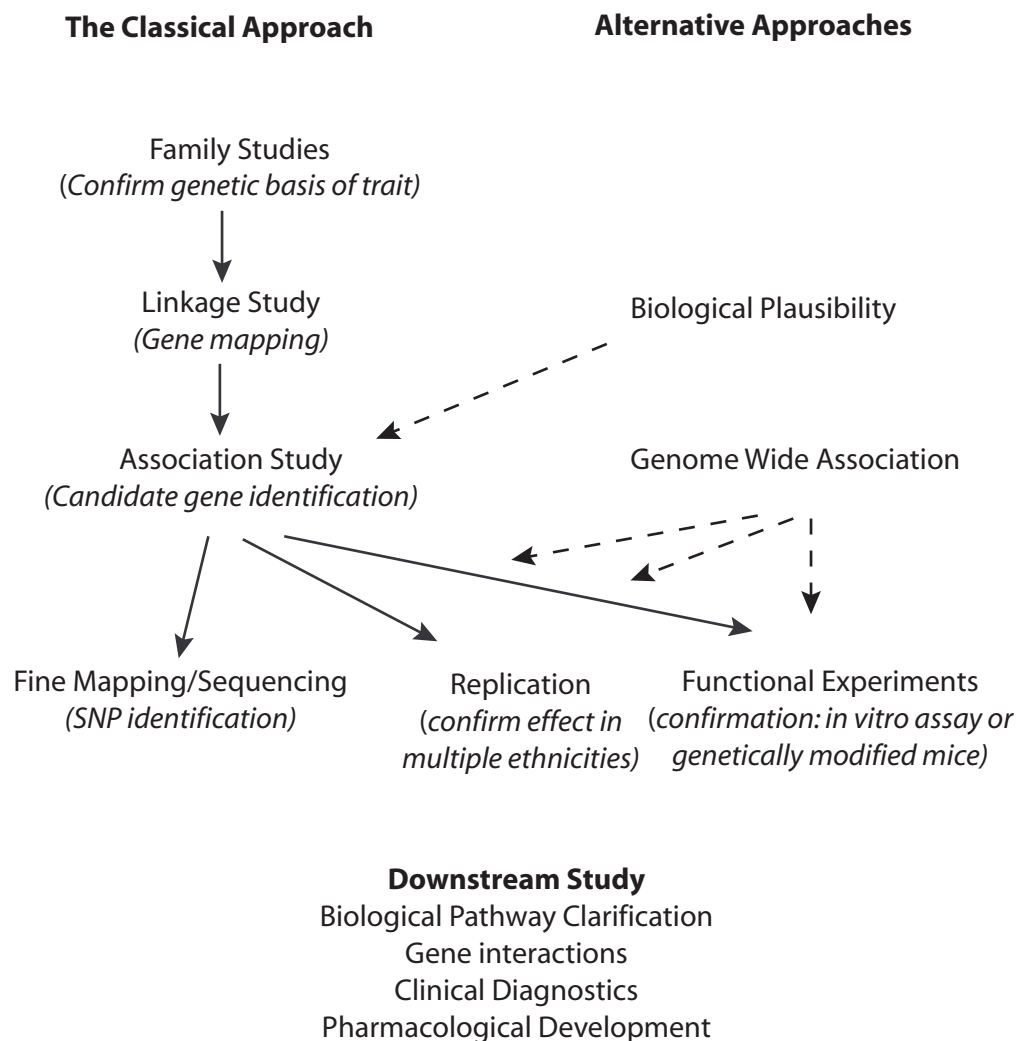
SLE – Systemic lupus erythematosus (Alarcon-Segovia *et al* 2005; Block *et al* 1975; Deapen *et al* 1992; Hochberg 1987), MS – Multiple sclerosis (Kalman & Lublin 1999), T1D – Type 1 diabetes mellitus (Hirschhorn 2003; Redondo *et al* 2001), RA – Rheumatoid arthritis (Silman *et al* 1993; Wordsworth 1995), CHD – Coronary Heart Disease (Allen *et al* 1967; Scheuner 2001).

An alternative measure of genetic effect which is often applied to quantitative traits (but can also be applied to binary data) is heritability (Visscher *et al* 2008). The observed variation in any phenotype within the population can be expressed as the sum of the variance due to environmental effects and the variance due to genetic effects, even if the nature of these actual effects is unknown. Heritability is the proportion of the variance in phenotype that can be attributed to variance in genotype. Genetic variance itself can be partitioned into additive genetic effects (where the mean effect in the offspring is determined by the sum of all the average genetic effects of the parental genes), dominance effects (where non-additive interactions occur between alleles at a locus) or epistatic effects (where non-additive interactions occur between alleles at different loci). Since non-additive genetic effects may only be estimated effectively in the setting of twin studies, heritability ( $h^2$ ) is often used in the strict sense of the proportion of phenotypic variance that can be attributed to variance in additive genetic effects. Heritability may be estimated in a number of different ways, but all depend on measuring how phenotypic effects cluster within families, in the light of what we would expect given the relatedness of the individuals and the assumed genetic and environmental causes. There has been some controversy over the use of heritability estimates, particularly when applied to human behavioural traits (Visscher *et al* 2008). Nonetheless, in the context of a quantified plasma protein in a particular population, heritability provides a useful measure of the fraction of variation in plasma protein levels due to genetic effects.

### *1.3.3: SNP discovery in Complex Genetic Disease*

**Figure 1.2** illustrates a typical approach to the identification of functional genetic variants in complex genetic disease. The “classical” approach has been successfully used in the identification of rare, high-risk, Mendelian traits; but has also been commonly applied to

**Figure 1.2.** A classical approach to the identification of functional genetic variants.



gene discovery in complex genetic traits (Botstein & Risch 2003; Glazier *et al* 2002). The initial step in investigation is the linkage study, which allows the relevant genetic variation to be localised to a particular physical region of the genome, even if the actual mutation of interest is unknown (Teare & Barrett 2005). Linkage scans are ideally conducted in large families with multiple affected individuals. A panel of genetic markers (typically a few hundred to a few thousand), spanning the entire genome, is typed, and markers that segregate with the trait of interest are used to localise the genomic region in which the functional effect lies (Weeks & Lathrop 1995).

Although non-parametric/model free linkage analysis has been applied to the study of both binary and quantitative complex genetic traits, it is notable that this has proven less successful than in the study of single gene disorders (Altmuller *et al* 2001). This may be largely due to inadequate power of current studies to identify multiple susceptibility genes each with only a modest effect on disease susceptibility. Using SLE as an example, the six whole genome linkage scans performed to date have identified multiple linkage regions, but all are large and contain many genes of potential functional significance (Sestak *et al* 2007). Furthermore the key gene candidate, *IRF5*, which has demonstrated a strong association with SLE in both candidate gene and genome-wide association studies does not lie in any of these linkage regions (Sestak *et al* 2007). In the study of complex genetic disease, the selection of candidate genes has often been based on other lines of evidence suggesting biological plausibility for a particular gene, rather than the classical approach of selecting genes from within a narrow linkage region.

The key advances in our understanding of complex trait genetics have really been made using the genetic association (candidate gene) study. In contrast to the linkage study, these studies

look for an association between a genetic variant (usually a SNP) and a particular trait in a population cohort, rather than in a pedigree of related individuals. The aim is still to identify genetic variants that are found more frequently in individuals with a particular trait of interest, but because of extensive recombination occurring during the entire genetic history of the population, the physical distance over which association operates is small and determined by population linkage disequilibrium (Cordell & Clayton 2005). Linkage disequilibrium (LD) is a property of the genome by which there is a high degree of correlation between genetic variants located in close physical proximity to each other. Since the correlation between adjacent SNPs will be destroyed if recombination occurs between them during meiosis, the presence of high LD implies that recombination has not occurred, even over the tens of thousands of generations since the founding of the population under study.

Considerable effort has been devoted to understanding and mapping the pattern of linkage disequilibrium across the human genome (Daly *et al* 2001; Frazer *et al* 2007; Gabriel *et al* 2002). It has been shown that the majority of the genome is divided into discrete haplotype blocks. Within these blocks all the SNPs are in strong LD with each other, to the extent that it is often possible to select a small number of SNPs (tagSNPs) that are representative of all the genetic variation seen within the block (Johnson *et al* 2001). Between blocks are regions termed “recombination hotspots” where there has historically been a high rate of recombination, leading to a rapid fall in LD. Without LD, genetic association could only be identified if the actual causative marker was genotyped in an experiment (direct association); because of LD, indirect associations can also be identified for a range of non-functioning neighbouring SNPs within the haplotype block.

Until recently, it has only been technologically or financially possible to type a limited number of SNPs in an association study. Even given the presence of LD, this has made it possible to look for association over a small genetic interval only, hence the selection of “candidate genes” for study. Recently, technological advances have allowed for the simultaneous genotyping of hundreds of thousands of SNPs simultaneously, and this density of markers allows us to perform hypothesis-free Genome Wide Association (GWA) analyses (Xavier & Rioux 2008). The GWA study is a major departure from the “classical” approach to the identification of functional genetic effects, and promises to revolutionise the identification of genetic effects underlying a number of traits and diseases.

The identification of a positive association signal in a single study is only the start of the process of refining and confirming the genetic effect, and elucidating the role of the functional variant. It is notable, or indeed notorious, that genetic associations, particularly from earlier studies, are not consistently reproduced. Indeed, in a review of 166 association studies only six were reproduced by >75% of subsequent published studies (Hirschhorn *et al* 2002). A single report of association should therefore be considered only suggestive of a true effect, and a key part of the SNP discovery process should be replication in at least one independent cohort.

While the existence of LD helps us to identify genetic effects arising from a particular gene, once an effect has been found LD hinders attempts to narrow the genetic signal down to a true functional variant from among a potentially large number of indirect associations. This has made it particularly difficult to identify functional variants in regions of the genome where strong LD can extend over large distances, for example the Major Histocompatibility Complex on chromosome 6 (Fernando *et al* 2008). The problem may be even more complex

if more than one SNP in a gene is functional, particularly if the LD between them is high. To some extent this problem can be overcome by the use of large cohorts, with the power to differentiate between SNPs even when the allele frequencies are similar and the LD between them is strong. Methodologically a conditional regression based analysis, combined with an appropriate model choice procedure, may allow the delineation of multiple functional SNP effects. A better approach, however, may be to search for functional effects in cohorts of different ethnicity. Haplotype mapping data suggests relatively limited haplotype diversity in European populations (mean block length 22kb, mean number of haplotypes per block 4.2), compared with Africans (mean block length 11kb, mean number of haplotypes per block 5) (Gabriel *et al* 2002). These genetic differences reflect different population histories, with the most likely explanation for the stronger LD in Europeans being a population bottleneck occurring during the founding of the modern European population (Reich *et al* 2001). These differences in haplotype structure, combined with different allele frequencies may allow genetic association studies conducted in diverse ethnic populations to dramatically narrow the genetic interval in which a functional SNP is to be found.

Even when DNA sequencing has identified all genetic variation in a region, and these have been subject to association analysis, there will always be an argument that the observed genetic association is due to correlation rather than causation. The final stage in the pathway of functional SNP identification must therefore be functional experimentation to confirm that the SNP can alter the phenotype of interest. In mice this may be feasible by genetic modification of the gene of interest; in humans obtaining this functional proof may be harder. It is perhaps notable that there is frequently a discrepancy between genetic association, and *in vitro* functional data, suggesting either there are a large number of false positive associations

reported, or alternatively our ability to perform physiologically relevant *in vitro* assays is limited (Ioannidis & Kavvoura 2006).

#### *1.3.4: Methodological Issues in Genetic Association Studies*

Although the essence of the genetic association study is straightforward there are a number of methodological issues that need to be considered if we are to avoid the problems of false positive and false negative results. Firstly it is important to ensure that we do not report associations due to confounding as a result of population stratification. In particular, subtle differences in the genetic ancestry of the cohort, between cases and controls when binary data is considered, or across the distribution of values for a quantitative trait, may lead to false-positive associations (Clayton *et al* 2005; Ewens & Spielman 1995). An extreme example of this is the study of admixed populations such as the African-American population. The majority of the African-American genome shares ancestry with current West African populations, however, a significant proportion (perhaps 10-30%) may arise from recent admixture with European populations (Wassel Fyr *et al* 2007). Even within European populations notable stratification has been described, suggesting this issue should be considered for any association study (Seldin *et al* 2006; Tian *et al* 2008c). A number of methods for circumventing this problem have been proposed. One commonly used approach is the family based study design, with related controls ensuring appropriate matching for ethnicity. The classic statistical test of family-based association is the transmission disequilibrium test (TDT) which utilises genetic data from a proband and their two parents (Spielman *et al* 1993). The null hypothesis for this test is that the transmission of alleles from parent to offspring is random. We would expect an allele associated with disease, or particular distribution of a quantitative trait, to be transmitted to an affected offspring more

commonly than expected by chance. Although providing a robust solution to the problem of population stratification, the use of a family-based design comes at the expense of reduced power, because information on allele transmission is only available from heterozygous parents. An additional problem with the classic TDT is that information is required for both parents, making the collection of large cohorts harder. A number of modifications to the TDT have been developed and I have used two of these in the analysis of data presented in this thesis. Firstly, TRANSMIT, which can infer parental genotypes using genetic data from additional siblings, thus allowing the inclusion of families with missing parental data and, secondly, qTDT, which allows for the analysis of quantitative data (Abecasis *et al* 2000; Clayton 1999).

An alternative approach to the issue of population stratification is to employ a traditional case-control design, but to ensure adequate control for the ethnic background of the population under study (Tian *et al* 2008a). With the increasing availability of population specific genotype data it has been possible to identify SNPs that are unique to one ethnic population or another, or at least have very different minor allele frequencies, leading to the concept that an individual's ethnic ancestry can be defined genetically. Either the ancestral origin of a particular gene can be determined by the selection of ancestry informative SNPs specific to that locus, or, increasingly popular in larger genetic studies, a genome wide panel of ancestry informative markers can be typed to control for admixture at a genome-wide level (Halder *et al* 2008; Kosoy *et al* 2009).

A major cause for the reporting of false positive associations is inadequate adjustment for the testing of multiple hypotheses (Elston *et al* 2002). The universally accepted measure of statistical proof is to demonstrate that the probability of observing, by chance, a difference

equal to or more extreme than the one actually observed, is less than 5% (in classical statistics  $P < 0.05$ ); in other words the false positive rate is 1 in 20. Clearly as more genetic variants are tested against the phenotypic trait of interest, the false positive rate will increase (if we tested 20 independent SNPs we would expect one false positive, if we tested 100 we would expect five false positives). In order to avoid this increased type 1 error rate it is necessary to adjust downwards the  $P$ -value we accept as “significant”. Simple tests such as the Bonferroni correction ( $P$ -value multiplied by the number of tests) are very stringent, particularly when many of the SNPs tested in an association study are not truly independent, but are correlated to some extent due to LD. This becomes a particular issue when considering genome-wide levels of significance but it is also important in association studies on the scale presented in this thesis. A number of approaches have been described. A simple but pragmatic approach that I have used in parts of this thesis is to adjust the Bonferroni correction for the degree of correlation between markers tested; no correlation between markers would result in a standard Bonferroni correction while complete correlation between SNPs would lead to an adjusted  $P = 0.05$  (Armitage & Parmar 1986; Sankoh *et al* 1997). An alternative approach that I have also used is a permutational correction; following calculation of  $P$ -values in the original data set, multiple iterations are performed on simulated datasets, to directly observe the frequency with which  $P$ -values more extreme than the one we have observed are generated by chance (Westfall & Young 1993).

## 1.4: C-reactive Protein

### 1.4.1: CRP - Structure and Physiological Function

C-reactive protein (CRP) was identified in the 1930's as a plasma protein in the blood of patients infected with *Streptococcus pneumoniae* and was named because it precipitated the pneumococcal cell wall "C" polysaccharide (Tillett & Francis 1930). It is a member of the pentraxin family of proteins and has a pentameric structure consisting of five identical 206 amino acid subunits non-covalently bound in an annular configuration around a 30Å central pore (Shrive *et al* 1996). CRP can bind a variety of ligands but has the highest affinity for phosphocholine (found in pneumococcal "C" polysaccharide) which is bound in a  $\text{Ca}^{2+}$ -dependent manner (Volanakis & Kaplan 1971). Each CRP subunit has a phosphocholine binding site consisting of two calcium ions bound to protein side chains. The binding sites are aligned so that, in the complete pentamer, they are all exhibited on one face of the CRP annulus (Thompson *et al* 1999). Other CRP ligands include ribonucleoproteins, such as the D protein of Sm and the 70kDa protein of snRNPs, chromatin-histone complexes and extrinsic ligands including glycans, phospholipids and numerous bacterial, fungal, parasitic and plant components (Du Clos *et al* 1988; Du Clos 1989; Jewell *et al* 1993; Pepys & Hirschfield 2003). CRP is also able to bind to apoptotic and necrotic cells and oxidised low density lipoproteins (LDL), although it is likely that the actual target ligand in these cases is exposed phosphocholine (Chang *et al* 2002; Gershov *et al* 2000; Pepys *et al* 1985; Volanakis & Wirtz 1979).

CRP interacts with other components of the immune system through binding sites which lie on the reverse face of the protein to the ligand binding regions. As I have already discussed,

CRP binds to the C1q component of complement resulting in potent activation of the classical complement pathway (Agrawal & Volanakis 1994; Mold *et al* 1999; Volanakis 1982).

Although initially controversial, there now a body of evidence demonstrating that CRP directly binds immunoglobulin Fc $\gamma$ -receptors, providing a bridge between innate and adaptive immune mechanisms, and allowing CRP to interact with a wide range of leucocytes (Bharadwaj *et al* 1999; Marnell *et al* 1995). Following binding to Fc $\gamma$  receptors, CRP induces a range of downstream intracellular signalling events including tyrosine phosphorylation and activation of the SYK and ERK MAPK signalling pathways (Chi *et al* 2002; Wang *et al* 2005). Ultimately CRP may result in diverse effects including an enhancement of phagocytosis and the production of cytokines, including important inflammatory mediators TNF- $\alpha$  and IL-1 $\beta$  (Ballou & Lozanski 1992; Mold *et al* 2001; Mold *et al* 2002b; Mold & Du Clos 2006; Wang *et al* 2005).

The function of CRP as a classical innate immune system receptor recognising a wide range of antigens, combined with its ability to interact with humoral and cellular components of the adaptive immune system suggests it plays a pivotal role in immune system function. The critical nature of this role is emphasised by the observation that human CRP deficiency has never been observed, suggesting its function is crucial to human viability. Despite this, however, data on what the key function of CRP actually is *in vivo* is surprisingly limited. CRP certainly provides protection against bacterial infection. By binding to bacterial cell wall components, such as phosphocholine, CRP promotes bacterial killing by activation of the complement cascade and induction of phagocytosis (Mold *et al* 2002a; Mold & Du Clos 2006). Human CRP is protective against infection by *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Salmonella enteritica* in mice and CRP gene polymorphism has been associated with increased mortality from *Streptococcus pneumoniae* bacteraemia

(Eklund *et al* 2006a; Mold *et al* 1981; Szalai *et al* 2000; Weiser & Pan 1998). However, a number of other human pattern recognition receptors would seem able to perform a similar function, and it is possible that the role of CRP is more complex.

It has been suggested that the primary role of CRP is actually anti-inflammatory or immunoregulatory. A host of *in vitro* anti-inflammatory effects have been documented for CRP. Anti-inflammatory effects include up-regulation of anti-inflammatory cytokines IL-1 receptor antagonist and IL-10, inhibition of neutrophil chemotaxis, inhibition of neutrophil adhesion due to L-selectin shedding and inhibition of the neutrophil respiratory burst (Buchta *et al* 1988; Kew *et al* 1990; Szalai *et al* 2002b; Tilg *et al* 1993; Zouki *et al* 1997). CRP also has binding sites for the complement regulatory protein Factor H (Jarva *et al* 1999). Factor H inhibits the complement C5 convertase thus limiting complement activation to the early opsonising components rather than the later inflammatory elements of the cascade (Mold *et al* 1984; Mold *et al* 1999). CRP has been shown to be protective against inflammatory conditions including a rabbit model of alveolitis, a murine experimental allergic encephalomyelitis model and a murine antigen induced arthritis model. (Heuertz *et al* 1994; Jiang *et al* 2006; Szalai *et al* 2002b; Webster *et al* 1994) Several groups have also reported a protective role for CRP in lipopolysaccharide induced inflammation, both in mice expressing a rabbit *CRP* transgene and in mice administered human CRP by injection (Mold *et al* 2002b; Xia & Samols 1997). These more complex functions for CRP are not, however, universally recognised, and some authors suggest that many of these reports are false positive results due to impurities in the CRP preparation or cross species differences in CRP structure (Hirschfield *et al* 2003). The difference in the expression and function of human CRP compared with rodent species is notable. The protein sequence itself is 70% identical between humans and mouse or rat. Mouse CRP, however, is expressed only at low level, with serum

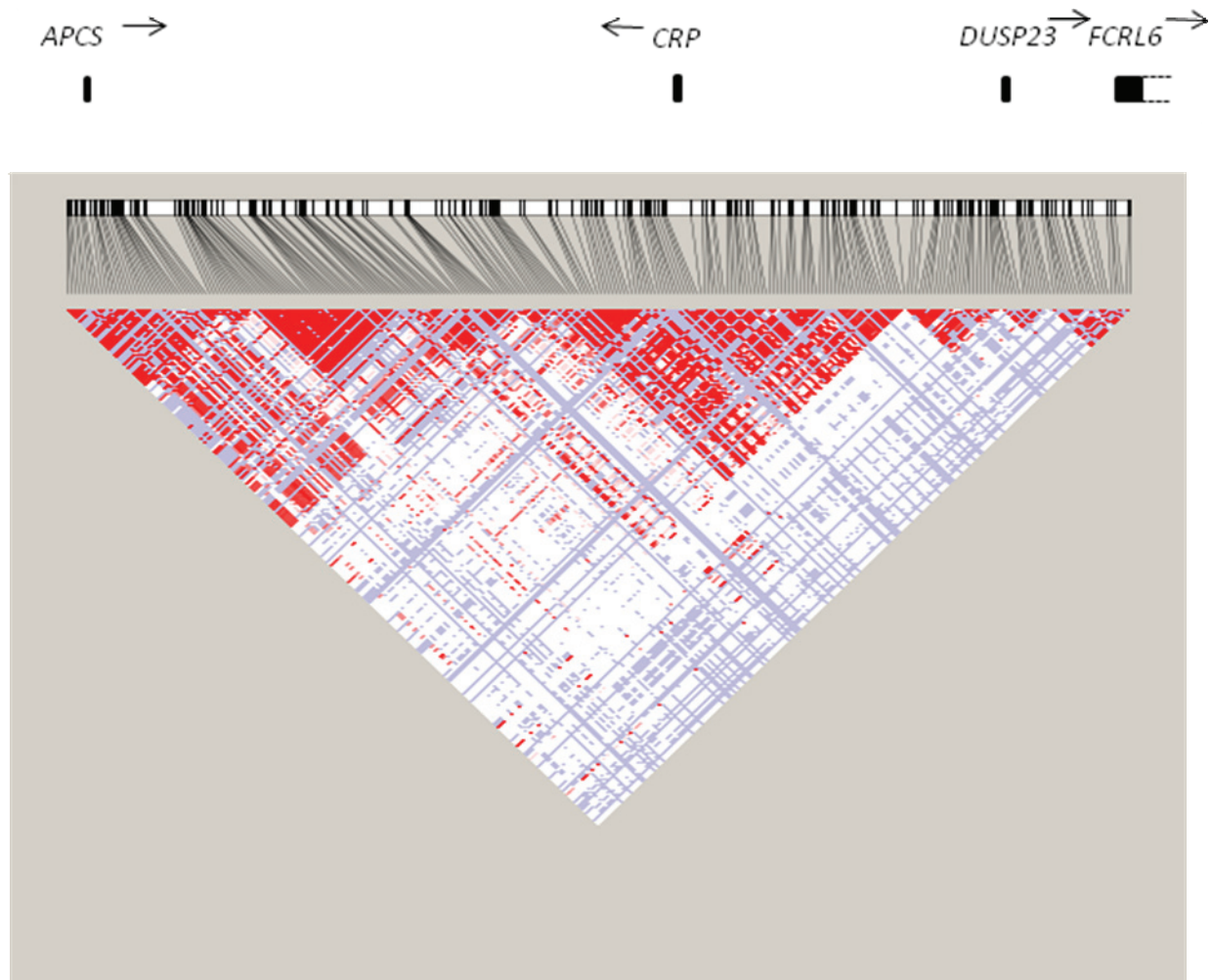
amyloid P constituting the key murine acute-phase pentraxin. In rats, CRP is an acute-phase regulated protein but differs functionally in comparison with humans, in particular it is unable to activate complement (Pepys & Hirschfield 2003).

#### *1.4.2: CRP – Gene Expression and Genetic Architecture*

CRP is expressed at low, basal, levels in apparently healthy individuals. In a study of healthy young adult blood donors, the median CRP level was 0.8mg/L and the 99th centile was 10mg/L (Shine *et al* 1981). In the presence of infection or inflammation CRP levels can increase dramatically to over 500mg/L as part of the human acute-phase response (Gabay & Kushner 1999; Pepys & Hirschfield 2003). Scintigraphic studies demonstrate rapid renal clearance of CRP, with a half-life of 19hrs, suggesting the rate of synthesis is the main determinant of serum CRP concentration (Vigushin *et al* 1993). Hepatocellular synthesis is by far the most important source of serum CRP, although synthesis of CRP by other cells has also been reported, including vascular endothelium and smooth muscle cells, neuronal and renal tissue, alveolar macrophages and peripheral blood mononuclear cells (Calabro *et al* 2003; Dong & Wright 1996; Jabs *et al* 2003; Kushner & Feldmann 1978; Murphy *et al* 1991; Venugopal *et al* 2005; Yasojima *et al* 2000). It seems unlikely that extrahepatic sources of CRP contribute significantly to circulating acute-phase CRP, but it remains unknown whether they contribute to measurable basal CRP levels.

*CRP* maps to human chromosome 1q23 where it is located adjacent to the gene encoding the related protein Serum Amyloid P (APCS) and is likely to have arisen by ancestral gene duplication. **Figure 1.3** demonstrates CRP in its genomic context in relation to neighbouring genes.

**Figure 1.3.** *CRP* in its genomic context on chromosome 1.

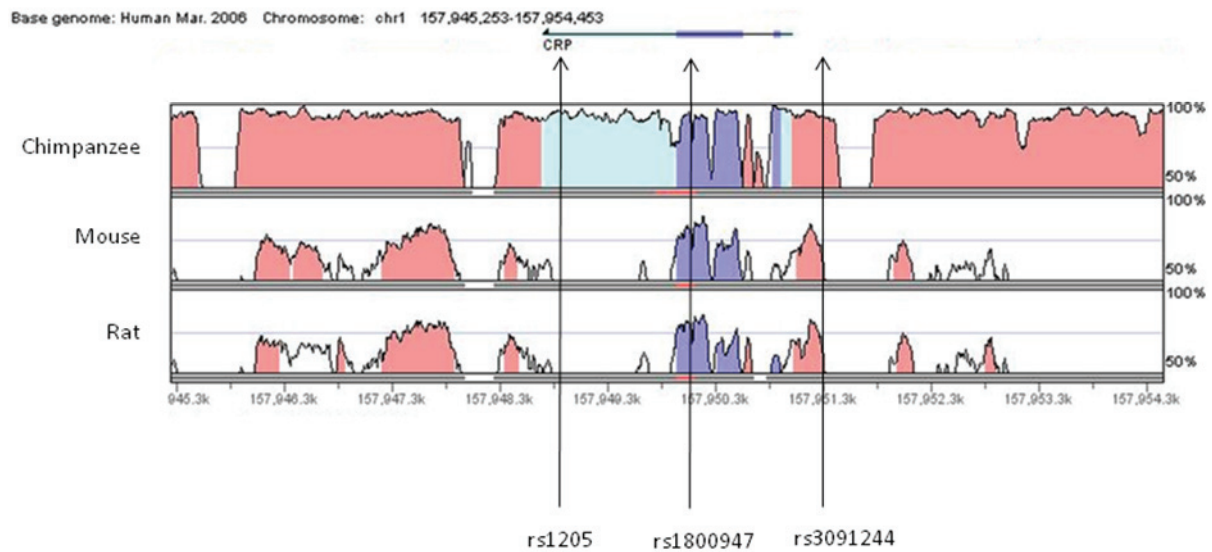


In the top section of this figure 250kb of DNA sequence is annotated to show *CRP* in its genomic location in relation to the neighbouring genes which encode serum amyloid P (*APCS*), Dual specificity phosphatase 23 (*DUSP23*) and Fc-receptor-like 6 (*FCRL6*). The lower section of the diagram is a Haploview plot demonstrating the pattern of LD between genotyped markers (data release 24, European (CEPH) population) within the same genetic interval. The position of the 414 genotyped SNPs relative to the genes is indicated by the black lines above the LD plot. The pairwise LD between any two SNPs is indicated by the colour coding of the corresponding square of the LD plot. High LD is indicated by a pink/red colour scheme (the dark red squares are linked with a  $D' = 1$  and a LOD score  $\geq 2$ )

**Figure 1.3** also demonstrates that, genetically, *CRP* is situated within a well defined haplotype block with minimal long range LD with neighbouring genes. The *CRP* gene spans 2.1kb: it has two exons, a single intron and a long 3'-untranslated region. **Figure 1.4** demonstrates the conservation of DNA sequence in humans compared with lower mammals, illustrated using a genomeVISTA plot (<http://genome.lbl.gov/vista/index.shtml>). The high level of sequence conservation between human and chimpanzee is to be expected but, outside the protein coding region the level of sequence conservation between humans and murids is relatively low. This is strikingly for the 3' UTR, a region with potentially regulatory function. It is notable that, with the exception of the coding region, no common SNPs (minor allele frequency >1%) are found in the areas of high conservation.

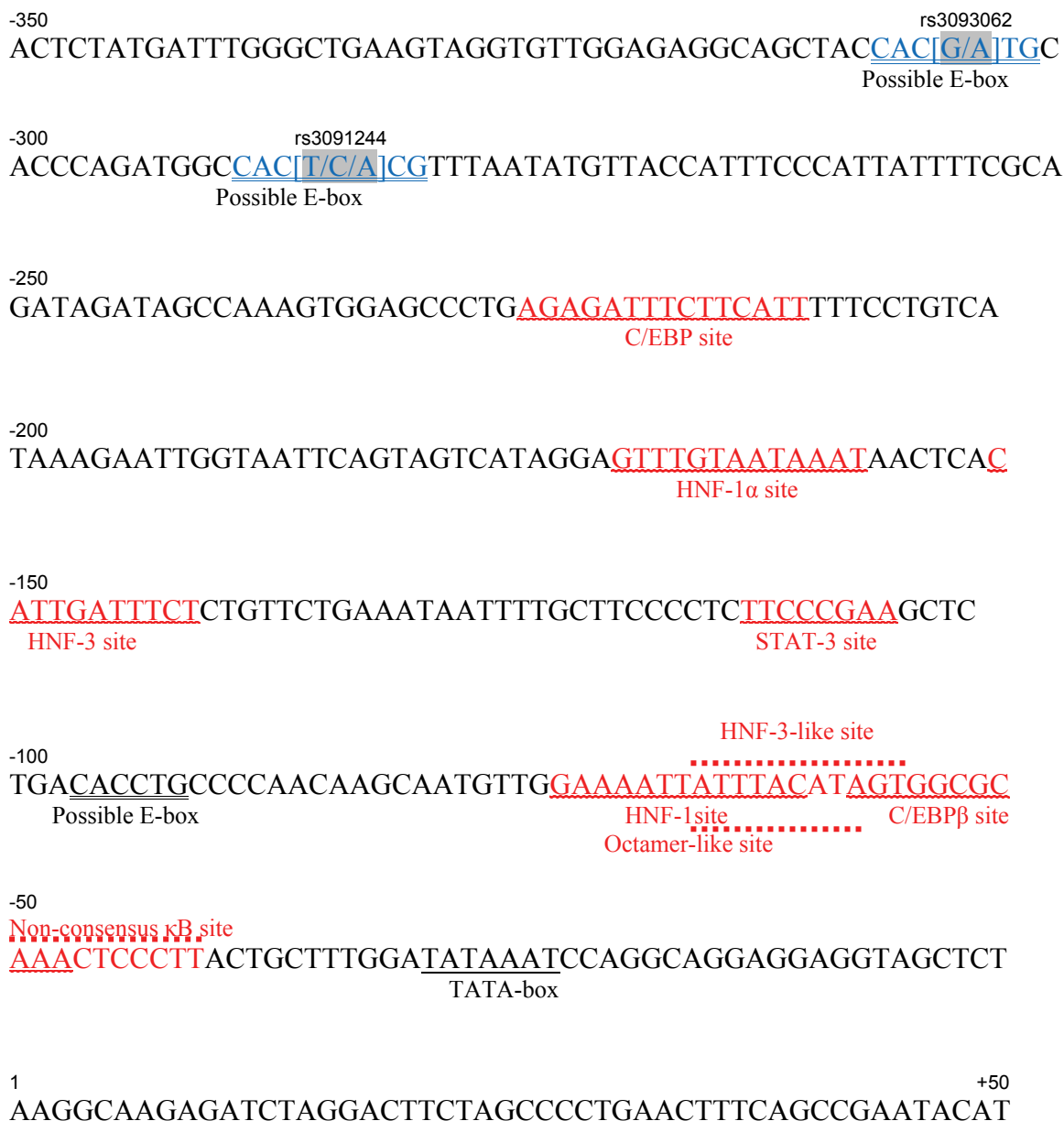
Consistent with the dramatic increase in CRP expression during active inflammation it has been identified that the main stimulus for gene expression is interleukin-6 (IL-6), with interleukin-1 $\beta$  (IL-1 $\beta$ ) acting synergistically but not in isolation (Ganter *et al* 1989). Both TNF- $\alpha$  and interleukin-17 (IL-17) also stimulate CRP expression independently of IL-6 and IL-1 $\beta$  (Choi *et al* 2007; Patel *et al* 2007). Experiments in both human hepatoma cell lines and transgenic mice have identified regions upstream and downstream to *CRP* that are essential for the regulation of gene expression. A number of transcription factors involved in gene expression have also been identified. **Figure 1.5** demonstrates the proposed binding sites for these with reference to the DNA sequence of the *CRP* proximal promoter. The 90 base-pairs upstream of the transcriptional start site are essential for any *CRP* to be expressed. This region contains the TATA box as well as sequences which bind C/EBP $\beta$  (bp -48 to -58) and HNF-1 $\alpha$  (bp -61 to -73) (Li & Goldman 1996a). Further upstream are an additional C/EBP family binding site (bp -225 to -211) and a binding site for transcription factor STAT-3 (bp -112 to -105) (Zhang *et al* 1996). There is evidence that C/EBP $\beta$  is the key transcription factor

**Figure 1.4.** Cross-species DNA sequence conservation at *CRP*



The GenomeVISTA plot provides a graphical representation of the level of DNA sequence conservation between the human genome reference sequence (Build 36) and the current genome reference sequences for the Chimpanzee, Mouse and Rat. The x-axis represents genomic location relative to *CRP* (which is shown at the top). The y-axis represents the level of cross-species conservation, with each data point plotted according to the percentage conservation of the 100 base pairs centred on the plotted point. Conservation levels >50% are plotted, and the graph is coloured when the levels of conservation exceeds 75% (coded DNA sequence in purple, 3' UTR sequence in blue and non-coded sequence in pink). For reference, the location of three key *CRP* SNPs has been indicated. It should be noted that, with the exception of synonymous exonic SNP rs1800947, no common human SNP (MAF>1%) is located in a region with >75% conservation compared with rat or mouse.

**Figure 1.5.** Transcription factor binding sites at the *CRP* promoter



The sequence of the 350bp of DNA upstream of the *CRP* transcription start site has been annotated to demonstrate the position of all proposed and confirmed transcription factor binding sites. Sequence has been highlighted in red where there is good functional evidence for transcription factor binding (see text). The two E-boxes highlighted in blue contain known polymorphic bases (indicated by brackets and shading). Some functional evidence exists for transcription factor binding to these sites (Szalai *et al* 2005) but, as I discuss in section 8.4.6, the evidence put forward in relation to the E-box containing rs3091244 is weak. An E-box centred at base -95 has been proposed on the basis of sequence, but there is no functional data.

responsible for cytokine induced acute-phase CRP production, and we now have a reasonably detailed understanding of the proteins involved in this process (Young *et al* 2008). The binding of C/EBP $\beta$  to promoter DNA requires the presence of members of the NF- $\kappa$ B/Rel family of proteins, in particular c-Rel which binds to C/EBP $\beta$  directly and Rel p50 which is capable of binding to a non-consensus  $\kappa$ B sequence adjacent to the C/EBP $\beta$  site (Agrawal *et al* 2003). Furthermore, Rel p50 interacts with the protein  $\beta$ -catenin, which itself interacts with TCF/LEF proteins bound to distal regulatory elements located about 700 bp downstream of CRP exon 2 (Choi *et al* 2007). STAT-3 and the upstream C/EBP $\beta$  site may act synergistically with the proximal promoter to produce maximal induced *CRP* expression. The expression of cytokine induced CRP therefore depends on a complex of transcriptional proteins interacting with both proximal and distal DNA regulatory elements.

Less well characterised are the presence regulatory proteins similar to HNF-3 and Octamer that may also bind the proximal promoter region (bp -56 to -66) (Li & Goldman 1996b). Recently 4 potential E-box elements have been identified in the *CRP* promoter. E-boxes are CAnnTG consensus sequences that can bind a variety of transcription factors including Max, Myc and USF (Upstream stimulatory Factor) and transcription may be inhibited or stimulated according to the combination of factors bound (Baudino & Cleveland 2001; Sirito *et al* 1998). USF-1 has been proposed to bind to an E-box centred at bp -497 and modulate *CRP* expression (Szalai *et al* 2005b). In transgenic mice further regulatory elements may be located at greater distances from the *CRP* coding region. In particular the intervals between bp -1270 and -541 and +7510 and +10050 appear to contain elements that suppress the production of CRP in resting cells (Murphy *et al* 1995). These experiments have also shown the 3' poly-A tail sequence is essential for gene expression.

### 1.4.3: CRP – Genetic Variation and its Relation to Gene Expression

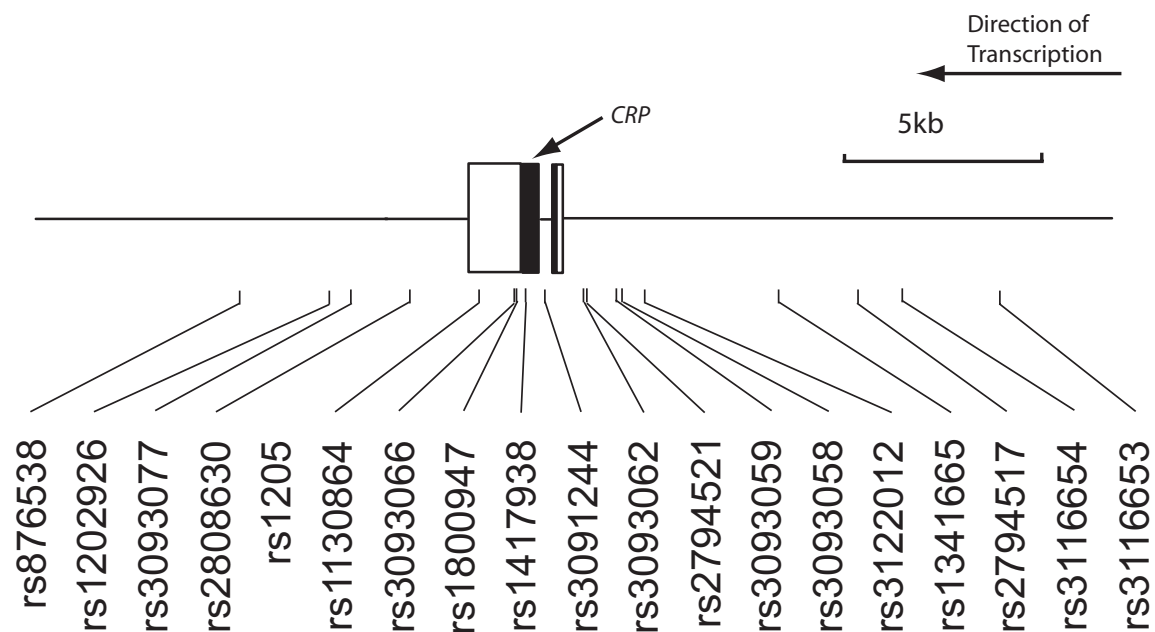
A number of studies have now estimated the heritability of basal CRP levels. Figures vary across studies, in part a consequence of the different methodologies employed and the use of either twin or family cohorts, but are all within the range of 10-56% (Austin *et al* 2004; Fan *et al* 1996; Fox *et al* 2008; Lange *et al* 2006a; MacGregor *et al* 2004; Pankow *et al* 2001; Saunders & Gulliford 2006; Vickers *et al* 2002; Wessel *et al* 2007). It is also notable that basal CRP in any one person is remarkably consistent over a long period of time, certainly many years (Danesh *et al* 1998; Kivimaki *et al* 2007; Ridker *et al* 1999). Given these observations it would certainly appear feasible to identify genetic variants which account for this heritability, and naturally efforts have focussed on *CRP* itself.

The first two reports of *CRP* polymorphisms affecting basal CRP level appeared in 2002. These were a GT dinucleotide repeat polymorphism in the *CRP* intron, and a SNP, *rs1800947* in the second exon (Szalai *et al* 2002a; Zee & Ridker 2002). Subsequently other reports of SNPs and haplotypes associated with basal CRP have been published. **Tables 1.2** and **1.3** summarise the data from studies that were published at the time these doctoral studies were commenced. These studies vary in their size (ranging from less than 100 patients to over 1000) and in their methodology. In particular the larger studies tended to model the alleles at each SNP as additive traits, while the smaller studies tended to pool genotypes from heterozygous individuals and those homozygous for the minor allele. The majority of these studies were conducted in cohorts of white European ancestry; studies in other ethnic populations are much more limited. **Figure 1.6** demonstrates the position of these published SNPs in relation to the *CRP*. It should be noted with reference to **figure 1.3**, that all these SNPs fall well within the single LD block, predicted from the HapMap European (CEPH)

population, that encompasses CRP and the surrounding upstream and downstream sequence. Those studies (**table 1.3**) utilising a haplotype analysis on their own data also confirm that these SNPs fall within a single block of high LD.

There are some notable points about the data presented in **tables 1.1** and **1.2**. Firstly multiple SNPs are often associated with serum basal CRP, often within the same cohort. Secondly, no single SNP is associated with serum CRP in every study where it has been tested, indeed for the commonly tested SNPs results range from a very strong association to no association at all. The only consistent effect, in fact, is the lack of association observed with SNP *rs2794521*. The number of studies inferring *CRP* haplotypes and analysing the haplotypic association with serum CRP are much more limited, and given the different selection of SNPs used by these authors it is difficult to directly compare the results of these studies.

**Figure 1.6.** The position of *CRP* SNPs reported in the published literature.



The location of *CRP* SNPs that appear in publications up to 2006 when these doctoral studies were planned

**Table 1.2.** Summary of reported *P*-values for the association of *CRP* SNPs with basal *CRP* levels in publications to 2006.

| Reference                              | CRP Single Nucleotide Polymorphism |            |           |           |        |           |           |           |           |           |           |           |           |           |           |           |           |           |           |
|----------------------------------------|------------------------------------|------------|-----------|-----------|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| (Zee & Ridker 2002) (U)                | rs876538                           | rs12029262 | rs3093077 | rs2808630 | rs1205 | rs1130864 | rs3093066 | rs1800947 | rs1417938 | rs3091244 | rs3093062 | rs2794521 | rs3093059 | rs3093058 | rs1341665 | rs3122012 | rs2794517 | rs3116654 | rs3116653 |
| (Brull <i>et al</i> 2003) (E)          |                                    |            |           |           |        | 0.014     |           | NS        |           |           |           | NS        |           |           |           |           |           |           |           |
| (Russell <i>et al</i> 2004) (E)        |                                    |            |           |           | 0.003  | NS        |           | 0.01      |           | NS        |           |           |           |           |           |           |           |           |           |
| (Obisesan <i>et al</i> 2004) (M)       |                                    |            |           |           |        |           | 0.03      |           | NS        |           |           | NS        |           |           |           |           |           |           |           |
| (Chen <i>et al</i> 2005) (H)           |                                    |            |           |           | NS     |           |           |           |           |           |           | NS        |           |           |           |           |           |           |           |
| (Suk <i>et al</i> 2005) (M)            |                                    |            |           |           |        |           |           | <0.001    | 0.003     |           |           |           |           |           |           |           |           |           |           |
| (Kovacs <i>et al</i> 2005) (E)         |                                    |            |           |           |        | NS        |           | NS        |           | 0.005     |           | NS        |           |           |           |           |           |           |           |
| (Miller <i>et al</i> 2005) WHS (E)     |                                    |            |           |           | 0.038  | <0.001    |           | 0.024     | <0.001    | <0.001    |           | NS        |           | 0.01      |           |           |           |           |           |
| (Miller <i>et al</i> 2005) PRINCE (E)  |                                    |            |           |           | <0.001 | 0.013     |           | 0.002     | 0.01      | <0.001    |           | NS        |           | <0.001    |           |           |           |           |           |
| (Miller <i>et al</i> 2005) PHS (E)     |                                    |            |           |           | <0.001 | 0.013     |           | NS        | NS        | <0.001    |           | NS        |           | 0.001     |           |           |           |           |           |
| (Davey Smith G. <i>et al</i> 2005) (U) |                                    |            |           |           |        |           |           | <0.001    |           |           |           |           |           |           |           |           |           |           |           |
| (Szalai <i>et al</i> 2005c) (E)        |                                    |            |           |           |        |           |           |           |           | NS        | NS        |           |           |           |           |           |           |           |           |
| (Szalai <i>et al</i> 2005c) (AA)       |                                    |            |           |           |        |           |           |           |           | NS        | NS        |           |           |           |           |           |           |           |           |
| (Kathiresan <i>et al</i> 2006) (E)     | NS                                 | 0.002      | 0.002     | NS        | <0.001 | 0.001     |           | NS        | 0.001     |           |           |           |           |           |           | 0.001     | NS        | 0.04      | 0.001     |
| (Crawford <i>et al</i> 2006) (E)       |                                    |            |           | NS        | NS     |           | NS        | <0.001    | NS        | NS        |           |           |           | NS        |           |           |           |           |           |
| (Crawford <i>et al</i> 2006) (AA)      |                                    |            |           | 0.024     | 0.003  |           | NS        | NS        | NS        | <0.001    |           |           |           | <0.001    |           |           |           |           |           |
| (Crawford <i>et al</i> 2006) (MA)      |                                    |            |           | NS        | 0.002  |           | <0.001    | NS        | NS        | 0.01      |           |           |           | NS        |           |           |           |           |           |
| (Marsik <i>et al</i> 2006) (E)         |                                    |            |           |           |        | 0.04      |           |           |           |           |           |           |           |           |           |           |           |           |           |
| (Balistreri <i>et al</i> 2006) (E)     |                                    |            |           |           |        |           |           | 0.008     |           |           |           |           |           |           |           |           |           |           |           |
| (Lange <i>et al</i> 2006b) (E)         |                                    |            |           |           |        | <0.001    | <0.001    |           |           | <0.001    |           |           |           |           |           |           |           |           |           |
| (Lange <i>et al</i> 2006b) (AA)        |                                    |            |           |           | 0.04   | 0.005     | NS        |           |           | NS        |           |           |           | <0.001    |           |           |           |           |           |
| (Wang <i>et al</i> 2006) (E)           |                                    |            |           |           |        | NS        |           |           |           | 0.007     |           | NS        | <0.001    |           | NS        |           |           | NS        | NS        |
| (Morita <i>et al</i> 2006) (J)         |                                    |            |           |           | NS     | 0.002     |           | NS        |           | 0.002     |           |           |           |           |           |           |           |           |           |

These are the studies that were published on, or shortly after the date when my doctoral studies commenced. All SNPs are located within a single *CRP* haplotype block, and those studies directly inferring haplotypes are outlined in **table 1.3**.

Studies have been annotated according to the ethnicity of their participants: E – White European, H – Han Chinese, AA – African American, MA – Mexican American, J – Japanese, M – Mixed ethnicity cohort, U – Unreported ethnicity

**Table 1.3.** Summary of published *CRP* haplotype association data

| Reference                                        | CRP Single Nucleotide Polymorphism   |                                      |                                      |                                      |                                      |                                 |                                 |                                      |                                 |                                 |                  |                            |                            |                                 | Freq                            | CRP level                            |                                                              |                                                                                |                                                              |                                         |
|--------------------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|---------------------------------|---------------------------------|--------------------------------------|---------------------------------|---------------------------------|------------------|----------------------------|----------------------------|---------------------------------|---------------------------------|--------------------------------------|--------------------------------------------------------------|--------------------------------------------------------------------------------|--------------------------------------------------------------|-----------------------------------------|
|                                                  | rs876538                             | rs12029262                           | rs3093077                            | rs2808630                            | rs1205                               | rs1130864                       | rs3093066                       | rs1800947                            | rs1417938                       | rs3091244                       | rs3093062        | rs2794521                  | rs3093059                  | rs3093058                       | rs3122012                       | rs2794517                            | rs3116654                                                    | rs3116653                                                                      |                                                              |                                         |
| (Szalai <i>et al</i> 2005c)<br>(U)               |                                      |                                      |                                      |                                      |                                      |                                 |                                 |                                      |                                 | C<br>T<br>A<br>T                | G<br>G<br>G<br>A |                            |                            |                                 |                                 |                                      | a<br>a<br>a<br>a                                             | Low <sup>b</sup><br>High <sup>b</sup><br>High <sup>b</sup><br>Low <sup>b</sup> |                                                              |                                         |
| (Russell <i>et al</i> 2004)<br>(E)               |                                      |                                      |                                      | G<br>A<br>G<br>G<br>A                | T<br>C<br>C<br>C<br>C                |                                 | G<br>G<br>G<br>G<br>C           |                                      |                                 | T<br>C<br>C<br>A<br>C           |                  |                            |                            |                                 |                                 |                                      | 0.28<br>0.24<br>0.23<br>0.06<br>0.05                         | High<br>Low<br><br><br>Low                                                     |                                                              |                                         |
| (Miller <i>et al</i> 2005)<br>(E)                |                                      |                                      |                                      | G<br>A<br>G<br>A<br>G<br>G           | T<br>C<br>C<br>C<br>C<br>C           |                                 | G<br>G<br>G<br>C<br>G<br>G      | T<br>A<br>A<br>A<br>A<br>A           |                                 | T<br>C<br>C<br>A<br>C           |                  | A<br>A<br>G<br>A<br>A<br>A | T<br>T<br>T<br>T<br>C<br>T |                                 |                                 |                                      | 0.30<br>0.29<br>0.26<br>0.06<br>0.06<br>0.02                 | High <sup>c</sup><br>Low<br><br><br>High                                       |                                                              |                                         |
| (Carlson <i>et al</i> 2005) <sup>d</sup><br>(E)  |                                      |                                      |                                      | A<br>G<br>A<br>A<br>A<br>A           | G<br>G<br>A<br>G<br>A<br>G           |                                 | C<br>C<br>C<br>C<br>C<br>C      | G<br>G<br>G<br>G<br>A<br>C           | T<br>A<br>A<br>A<br>A<br>C      | T<br>C<br>C<br>A<br>C<br>C      |                  |                            |                            | A<br>A<br>A<br>A<br>A<br>A      |                                 |                                      | 0.29<br>0.28<br>0.28<br>0.06<br>0.06<br>0.02                 | High<br><br>Ref <sup>e</sup><br>High                                           |                                                              |                                         |
| (Carlson <i>et al</i> 2005) <sup>d</sup><br>(AA) |                                      |                                      |                                      | A<br>A<br>G<br>A<br>A<br>A<br>A<br>A | G<br>A<br>G<br>G<br>G<br>G<br>A      |                                 | A<br>C<br>C<br>C<br>C<br>C<br>C | G<br>G<br>G<br>G<br>A<br>A<br>A      | A<br>A<br>A<br>A<br>A<br>A<br>C | A<br>C<br>A<br>C<br>A<br>A<br>C |                  |                            |                            | A<br>A<br>A<br>T<br>A<br>A<br>A |                                 |                                      | 0.22<br>0.18<br>0.17<br>0.17<br>0.12<br>0.09<br>0.03<br>0.01 | Ref <sup>e</sup><br><br><br>High<br>High<br><br>High                           |                                                              |                                         |
| (Kathiresan <i>et al</i> 2006)<br>(E)            | G<br>A<br>G<br>G<br>G<br>G<br>G<br>G | C<br>C<br>C<br>C<br>C<br>C<br>C<br>C | T<br>T<br>T<br>T<br>T<br>T<br>T<br>T | A<br>G<br>A<br>A<br>A<br>A<br>A<br>A | G<br>G<br>A<br>A<br>A<br>A<br>A<br>A | T<br>C<br>C<br>C<br>C<br>C<br>C |                                 | G<br>G<br>G<br>G<br>G<br>G<br>G<br>G | T<br>A<br>A<br>A<br>A<br>A<br>A | T<br>C<br>C<br>C<br>C<br>C<br>C |                  |                            |                            |                                 | G<br>A<br>A<br>A<br>A<br>A<br>A | C<br>T<br>C<br>C<br>C<br>C<br>C<br>C | T<br>T<br>T<br>T<br>T<br>T<br>T<br>T                         | G<br>C<br>C<br>C<br>C<br>C<br>C                                                | 0.29<br>0.19<br>0.12<br>0.09<br>0.08<br>0.05<br>0.05<br>0.05 | High<br><br><br><br><br>High<br><br>Low |

The five studies to analyse associations between *CRP* haplotype and serum CRP level have been annotated according to the ethnicity of their participants: E – White European, AA – African-American, U – Unspecified (although likely to be predominantly African-American)

<sup>a</sup> Study only analysed homozygous individuals. Population frequencies not given.

<sup>b</sup> Only an overall haplotype *P*-value presented in paper. High and Low haplotypes inferred from graphical data.

<sup>c</sup> Data from three cohorts presented in this paper. The significance of the CRP association with this haplotype varied from very significant ( $P < 0.0001$ ) to just significant (0.01)

<sup>d</sup> Although Carlson *et al* presents haplotype frequency data for each ethnicity the CRP association analysis was performed on a combined cohort.

<sup>e</sup> These studies have modelled their haplotype associations in comparison with a reference haplotype, annotated here as “ref”

In addition to the candidate gene studies described above, four genome-wide association (GWA) studies and a single meta-analysis of existing publications have also recently been published. Since these were published after the majority of work in this thesis was performed, a full discussion of these papers, and how they fit in with my results will be presented in Chapter 8.

All the studies discussed so far have aimed to address the question of association between genetic variants and basal CRP, a potentially different question is which genetic variants are associated with CRP levels in the context of an acute-phase response. I will argue later that this question is equally important and it may, in fact, turn out to be of greater clinical relevance. For reasons that I will discuss later in section 6.4.3, this is a difficult question to answer and work in this field is therefore limited. An early study looking at the role of *CRP* genetics in determining CRP levels in patients with periodontitis suggested the rare allele of *rs1130864* was associated with higher CRP, however, the magnitude of the acute-phase stimulus was fairly modest (D'Aiuto *et al* 2005). CRP rises acutely following myocardial infarction and the magnitude of rise predicts overall outcome, so there has been interest in whether the magnitude of this rise is influenced by *CRP* genetics (Pietila *et al* 1996; Ueda *et al* 1996). The only current study to address this question found higher post-myocardial infarction CRP levels in carriers of the A allele of *rs3091244* and the rare allele of *rs3093059*, with dramatically higher levels in individuals homozygous for these variants. Furthermore the haplotype containing these two alleles was also associated with higher CRP. Conversely lowest post MI CRP was seen in carriers of the rare alleles of *rs1800947* and *rs1205* and a haplotype containing both of these (Suk *et al* 2006). A similar pattern of association was seen in another study looking at the magnitude of acute-phase CRP following coronary artery bypass surgery, in particular confirming the lower level of acute-phase CRP

rise in carriers of the minor allele of *rs1800947* (Perry *et al* 2009). An association between *rs3091244* and higher CRP was also observed (but with only borderline significance following a conservative Bonferroni correction), although interestingly these authors, using a mixed ethnicity population, observed their strongest association with the T allele. Neither of these papers addressed the question of whether these variants were also associated with clinical outcome.

No studies to date have considered whether *CRP* variants influence the magnitude of acute-phase CRP in the context of chronic inflammatory disease. Potentially this is a very important question since serum CRP is measured extensively in clinical practice to diagnose and monitor infection and inflammatory diseases across a whole range of clinical specialities. Surprisingly little published data exists on the inter-individual variability of acute-phase CRP. This is particularly important if CRP is used to make prognostic decisions, or to make comparisons between individuals, say as part of a clinical scoring system. Rheumatoid arthritis is a case in question. The original DAS28 scoring system was based on erythrocyte sedimentation rate (ESR), a tender and swollen joint counts and a global assessment by patient and physician (Prevoo *et al* 1995). Recently it has been suggested that CRP could replace ESR as the objective measure of inflammation (Inoue *et al* 2007; Paulus *et al* 1999). Some authors have suggested the ESR DAS28 and CRP DAS28 could be used interchangeably, however if there is a significant *CRP* genetic effect then an adjustment for genotype may need to be considered. In addition it has been suggested that a prognostic algorithm, incorporating CRP level among its parameters can be used to predict the course of disease in patients with early, undifferentiated arthritis (van der Helm-van Mil *et al* 2007; van der Helm-van Mil *et al* 2008). The introduction of a genetic adjustment to the acute-phase

CRP level observed in these patients may well increase the prognostic power of this algorithm.

Finally, in addition to the *cis*-acting effects described above there are a number of *trans*-acting genetic variants reported to influence basal CRP levels. At the time the work presented here was planned, published data focussed on three genes, *IL6*, *IL1B* and *APOE*. Of these, the most studied are the genes encoding the acute-phase cytokines IL-1 $\beta$  and IL-6. The number of studies incorporating genotyping at *IL6* and *IL1B* are extensive, and the literature is particularly complicated by the heterogeneity of the cohorts in which studies are undertaken. Cohorts vary from healthy individuals to those with metabolic or potentially pro-inflammatory conditions such as diabetes mellitus or CHD. A detailed discussion of this literature is beyond the scope of this thesis, although a summary is given below, and further discussion may be found in section 8.5.1.

For *IL6* most work has focussed on promoter SNP *rs1800795*, located 174 bp upstream of the transcriptional start site. Numerous authors have demonstrated an association between *rs1800795* and levels of serum CRP (Eklund *et al* 2006b; Libra *et al* 2006; Vickers *et al* 2002; Walston *et al* 2007). There are also, however, also a number of negative reports (Latkovskis *et al* 2004; Potaczek *et al* 2007). A second variant that also appears frequently in the literature is *rs1800796*, also found in the *IL6* promoter, 572 bp upstream of the transcriptional start site. Again there is inconsistency over the reported association between this variant and serum CRP. There are a number of positive reports, notably from populations of East Asian ancestry, where the minor allele frequency is much more common than that observed in European studies (Jang *et al* 2008; Paik *et al* 2007; Wong *et al* 2007). There are also several negative studies (Latkovskis *et al* 2004; Walston *et al* 2007). Furthermore, it has

also been proposed that basal CRP is influenced by an interaction between these two variants (Ferrari *et al* 2003).

There is less published literature concerning the role of *IL1B* variation in determining serum CRP levels. Earlier papers tended to focus on synonymous exonic SNP *rs1143634* with a number of reported associations with serum CRP (Berger *et al* 2002; Eklund *et al* 2005; Latkovskis *et al* 2004). Subsequently there has been interest in promoter SNP *rs16944*, and some suggestion that both SNPs influence mediate the effects of *IL1B* variation on CRP (Eklund *et al* 2003; Sainz *et al* 2008).

Several authors report that apolipoprotein E type is associated with basal CRP (Apo  $\epsilon$  type is determined by a haplotype of two coding SNPs in *APOE*, *rs429358* and *rs7412*). The  $\epsilon 2$  variant appears to be associated with higher CRP than the  $\epsilon 4$  variant, but the effect of the  $\epsilon 3$  variant has not been consistently reported (Austin *et al* 2004; Chasman *et al* 2006; Judson *et al* 2004; Lange *et al* 2006a; Manttari *et al* 2001; Marz *et al* 2004; Rontu *et al* 2006).

Other reported CRP associations from candidate gene studies include the genes encoding TNF- $\alpha$ , LAMIN A/C, haptoglobin, CD14, the glucocorticoid receptor, the  $\beta$ -adrenergic receptor, tyrosine hydroxylase and cyclooxygenase 2 (Cipollone *et al* 2004; Hegele *et al* 2003; Hubacek *et al* 2003; Lakka *et al* 2006a; van Rossum & Lamberts 2004; Wessel *et al* 2007; Wuyts *et al* 2002). Finally, linkage has been demonstrated between CRP and a quantitative trait locus on chromosome 20q13 (Lakka *et al* 2006b). The recent Genome-wide Association studies have also identified novel *trans*-acting effects and these will also be discussed further in chapter 9.

#### 1.4.4: *CRP – Functional and Genetic Role in Systemic Lupus Erythematosus*

CRP has been implicated in the pathogenesis of two important diseases: systemic lupus erythematosus and coronary heart disease (along with other forms of arterial disease sharing atherosclerosis as the primary pathological lesion). We have already discussed a possible role for CRP in protecting against autoimmunity, by promoting clearance of apoptotic cells and autoantigen containing immune complexes. A failure of these clearance mechanisms may contribute to the pathogenesis of SLE so, potentially, deleterious *CRP* genetic variation could increase susceptibility to SLE, perhaps by resulting in lower levels of CRP expression. There are several lines of evidence that support this hypothesis. Firstly there is a long standing observation that patients with SLE have acute-phase CRP levels that are somewhat less than one would expect for any given level of active inflammation; certainly less than that induced by infection (Pereira Da Silva *et al* 1980). Secondly there is recent experimental evidence from mouse models that CRP, either expressed by a transgene or administered parenterally, can ameliorate or delay the onset of immune mediated glomerulonephritis (Rodriguez *et al* 2005; Rodriguez *et al* 2006; Szalai *et al* 2003). Interestingly this effect appears not to be mediated by complement but rather by CRP-induced production of anti-inflammatory IL-10.

There is also some genetic evidence to support a role for *CRP* in determining susceptibility to SLE. The *CRP* genetic interval on chromosome 1q23 has been identified as an SLE linkage region in a number of genome wide studies, although it should be noted that this region also contains a number of other immunologically relevant genes (Cantor *et al* 2004; Edberg *et al* 2002; Moser *et al* 1998). Four studies (three of which were published during or after the course of the work contributing to this thesis) have now reported an association between SLE and *CRP* genetic variants. The first, a result of earlier work in my supervisor's laboratory,

reported the association of SNP *rs1800947* with SLE in a large family based study (Russell *et al* 2004). An association between *rs1800947* and lupus nephritis was also demonstrated in a second study, but a third study suggested the association was with a different SNP, *rs3091244* (Edberg *et al* 2008a; Jonsen *et al* 2007). A more recent study identified no disease association with five commonly studied *CRP* SNPs but did report an association between *CRP* haplotypes, and SLE (Shih *et al* 2008). The genetic data is therefore inconsistent and requires clarification through further study.

#### *1.4.5: CRP – Functional and Genetic Role in Cardiovascular Disease*

In the mid-1990's evidence arose to suggest that basal *CRP* in otherwise healthy individuals was a predictor of future coronary events (Thompson *et al* 1995). This was confirmed both in individuals with unstable angina (but no myocardial necrosis), and in the general population (Haverkate *et al* 1997; Kuller *et al* 1996; Liuzzo *et al* 1994). A meta-analysis of 22 prospective studies showed an odds ratio of 1.45 (1.25-1.68) for an acute coronary event between individuals with basal *CRP* in the upper third compared with an individual with *CRP* in the lower third (Danesh *et al* 2004). These observations led some groups to recommend that serum *CRP* could be used as a clinical measure of coronary risk (Danesh *et al* 2004; Pearson *et al* 2003). There is now a large amount of evidence demonstrating that elevated basal *CRP* is associated with all forms of atherosclerotic vascular disease, including stroke and peripheral vascular disease (Bassuk *et al* 2004).

Various theories have been proposed to explain the association between *CRP* and coronary disease. *CRP* may be a marker of inflammation occurring within atheromatous plaques themselves, or alternatively it may reflect low grade systemic inflammation, perhaps caused

by pathogens such as *Chlamydia pneumoniae* or *Helicobacter pylori* (Danesh *et al* 1997; Ross 1999). Alternatively CRP may simply be a marker of other known cardiovascular risk factors such as smoking and obesity (Lagrand *et al* 1999). Finally there has been much interest in whether CRP is proatherogenic and proinflammatory itself.

It is clear that a number of acute-phase markers are associated with cardiovascular disease, including serum amyloid A, fibrinogen, albumin and leukocyte count, and this is suggestive of a general inflammatory effect rather than a special role for CRP (Danesh *et al* 1998).

Certainly many studies have found that classical cardiovascular risk factors, including smoking, obesity, and diabetes are associated with CRP levels, and can explain much of the CRP effect (Danesh *et al* 1999; Ford 1999). Some authors have suggested that CRP may actually have a rather poor predictive value for coronary risk, but despite this a consensus opinion would be that a residual association between CRP and coronary events does persist even after adjustment for multiple covariates in large studies (Levinson *et al* 2004).

There is circumstantial evidence to suggest that CRP may have a role in causing direct vascular damage. Firstly, in addition to the immunological targets discussed so far, CRP can bind to oxidised or otherwise modified LDL, a key constituent of atheromatous plaques (Bhakdi *et al* 1999; Chang *et al* 2002; Taskinen *et al* 2002; Taskinen *et al* 2005). Once bound, CRP is capable of activating complement (de Beer *et al* 1982). It is also capable of stimulating macrophages into producing procoagulant tissue factor, and forming foam cells (Cermak *et al* 1993; Zwaka *et al* 2001). Finally, CRP has been demonstrated to stimulate vascular smooth muscle proliferation, another characteristic of atheromatous plaques (Wang *et al* 2003).

Although CRP is frequently found in atheromatous plaques, its presence does not prove it has a role in plaque pathogenesis; indeed, some authors suggest CRP may actually be playing a protective role (Zhang *et al* 1999). It has been suggested that CRP may actually prevent foam cell formation, that it may limit complement activation and induce apoptosis of vascular smooth muscle (Bhakdi *et al* 2004; Blaschke *et al* 2004; Singh *et al* 2008a). In addition there have been suggestions that the pro-inflammatory effects observed in many early studies may be due to contamination of the CRP preparations, or degradation of the intact CRP molecules (Singh *et al* 2008b; Taylor *et al* 2005). The contradictory results from *in vitro* studies have been matched by contradictory results from mouse models of atherosclerosis. CRP was initially suggested to be proatherogenic in the *apoE*<sup>-/-</sup> knockout mouse model of atherosclerosis (Paul *et al* 2004; Schwedler *et al* 2005). Subsequent authors, however, have been unable to demonstrate any such effect (Hirschfield *et al* 2005; Tennent *et al* 2008; Trion *et al* 2005).

Perhaps some of the most compelling evidence for an important role for CRP comes from studies relating to the use of HMG-CoA reductase inhibitors (Statins). Statins are cholesterol lowering drugs known to improve long-term cardiovascular morbidity. They are also known to lower CRP, although the exact mechanism for this is unknown (Ridker *et al* 1999). A study has suggested that the beneficial effect of statins is similar between people with elevated LDL cholesterol and people with elevated CRP but normal LDL cholesterol (Ridker *et al* 2001).

Even this brief review of the current literature highlights ongoing controversy over the true role of CRP in cardiovascular disease: genetic studies have the potential to clarify this situation. The finding of a consistent association between *CRP* genotype and cardiovascular disease would provide further evidence for a functional role for *CRP*; current data however

remains far from convincing. The only replicated association is for *rs2794521*, with a fairly strong effect reported in a Han Chinese population, and a rather weak effect in men from the Physicians Health Study; interestingly this SNP has never been associated with CRP levels (Chen *et al* 2005; Miller *et al* 2005). The AA genotype of *rs3091244* has also been associated with CHD, but only in the non-Hispanic white subset of the large NHANES III cohort. (Crawford *et al* 2006). Numerous other studies, however, have only reported negative findings (Hage & Szalai 2007; Kardys *et al* 2006; Kathiresan *et al* 2006; Wang *et al* 2006; Zee & Ridker 2002). A more convincing demonstration that CRP plays a direct causal role in atherosclerosis would be to demonstrate a *CRP* genetic variant that influences CRP levels and also increases cardiovascular risk by a magnitude that could be predicted given the known association between serum CRP and cardiovascular risk; the so-called “Mendelian Randomisation” experiment (Davey & Ebrahim 2003; Hingorani & Humphries 2005). An initial attempt to demonstrate this three way relationship was unsuccessful, although only a single *CRP* SNP was included for analysis (Casas *et al* 2006). Recently, a more extensive Mendelian Randomisation study has been published (Zacho *et al* 2008). An extensive array of *CRP* SNPs was genotyped in four cohorts consisting of nearly 50,000 participants. The usual association between genotype and serum CRP was seen, but no association with cardiovascular disease was observed, even though it could be predicted that the relevant genetic variants should increase the CHD risk by 32% given the effect on CRP level.

## 1.5: The C3 component of complement

An outline of the components and function of the human complement cascade has already been presented in section 1.1, here I will consider in more detail the central component of this cascade, C3.

### 1.5.1: C3 - Structure and Physiological Function

Complement component C3 is a member of the  $\alpha_2$ -macroglobulin family of proteins and, like CRP, it is a phylogenetically ancient molecule. Homologues have been described in primitive chordates (*urochordata* (tunicates)), *echinodermata* (sea urchins/starfish) and, more recently, C3-like genes have been reported in the *cnidaria* (coral/jellyfish) suggesting C3 evolved at least one billion years ago (Al-Sharif *et al* 1998; Azumi *et al* 2003; Dishaw *et al* 2005; Nonaka *et al* 1999). It is synthesised as a 1663 amino acid precursor molecule, prepro-C3. Post-translational modifications include removal of a 22-residue signal peptide from the N-terminal end and a quadruplet of arginine molecules from the centre of the pro-C3 amino acid chain, resulting in a secreted molecule consisting of 2 distinct chains: a 645 amino acid, 70kDa  $\alpha$ -chain and a 992 amino acid, 115kDa  $\beta$ -chain (de Bruijn & Fey 1985). The crystal structure of C3 has now been resolved and each chain has been shown to contribute 6 domains to the final molecule, with a 13th domain made up of residues from both chains (Janssen *et al* 2005). The core of the molecule is made up of eight homologous macroglobulin (MG) domains. A key functional feature of C3, and other related molecules including C4, is the presence of an internal thioester bond which mediates the covalent attachment of C3 to target surfaces (Law *et al* 1979; Law & Levine 1977). This bond is formed by the side chains of residues Cys988 and Gln991, both of which are within the

thioester domain, and access to this thioester site is limited in the intact C3 molecule by its location tucked against domain MG8 (Fredslund *et al* 2006; Janssen *et al* 2005). Activation of C3 by the C3 convertases occurs by cleavage between Arg727 and Ser727 which yields C3a from the N-terminal region of the  $\alpha$ -chain (de Bruijn & Fey 1985; Janssen *et al* 2005). This cleavage also results in marked conformational changes in C3b, exposing the reactive thioester site (Nishida *et al* 2006). Further rounds of proteolytic cleavage, catalysed by complement factor I (but dependent on other co-factors) generate the bioactive fragments iC3b and C3dg and, finally, fragment C3c (Gros *et al* 2008; Harrison & Lachmann 1980).

I have already introduced some of the biological functions of these C3 fragments in section 1.2.1.; these will be discussed in further detail here. The initial activation of C3 yields a small 77 amino acid fragment, C3a. C3a and C5a (and possibly C4a) share many biological functions and are known collectively as anaphylotoxins (Kohl 2001). Their effects are mediated via a seven-transmembrane, G-protein coupled cell-surface receptor. C3a acts as a chemoattractant, recruiting inflammatory cells to the site of complement activation and can triggers an oxidative burst in phagocytic cells, thus enhancing bacterial killing. Finally it can influence the local microcirculation by increasing vascular permeability and inducing vascular smooth muscle contraction.

C3 cleavage fragments C3b, iC3b, C3dg and C3c differ in their ability to bind the complement receptors CR1-4 and CRIg, and this, in turn, leads to a range of physiological consequences. The key receptors for C3b are CR1 and CRIg. We have already discussed the role of CR1 in transporting circulating C3b opsonised immune complexes to the spleen for removal. CR1 is also borne on neutrophils where it induces phagocytosis of the opsonised particle (Sengelov *et al* 1994; van Lookeren *et al* 2007). CRIg is expressed on a limited range

of tissue macrophages and receptor binding triggers phagocytosis (Helmy *et al* 2006). iC3b can also bind CR1g, but its key receptors are CR2-4. CR2 is predominantly a B-cell receptor and when co-ligated with the B-cell receptor, B-cell immunity is enhanced (Fearon & Carroll 2000; Weis *et al* 1984). CR3 and CR4 are integrin receptors which share a common  $\beta$ -chain (CD18) and differ in their  $\alpha$ -chain (CD11b and CD11c respectively). CR3 is expressed predominantly on macrophages and neutrophils, and plays a role in mediating phagocytosis; CR4 is poorly characterised but is expressed on the same cells and is likely to have a similar function (Ross & Lambris 1982; Ross & Vetvicka 1993; van der Vieren *et al* 1995). C3dg binds CR2 only and therefore triggers adaptive, but not innate effector responses (Weis *et al* 1984). C3c is generally considered biologically inactive, although, theoretically at least, it can bind to CR1g, so this opinion may need to be revised (Wiesmann *et al* 2006).

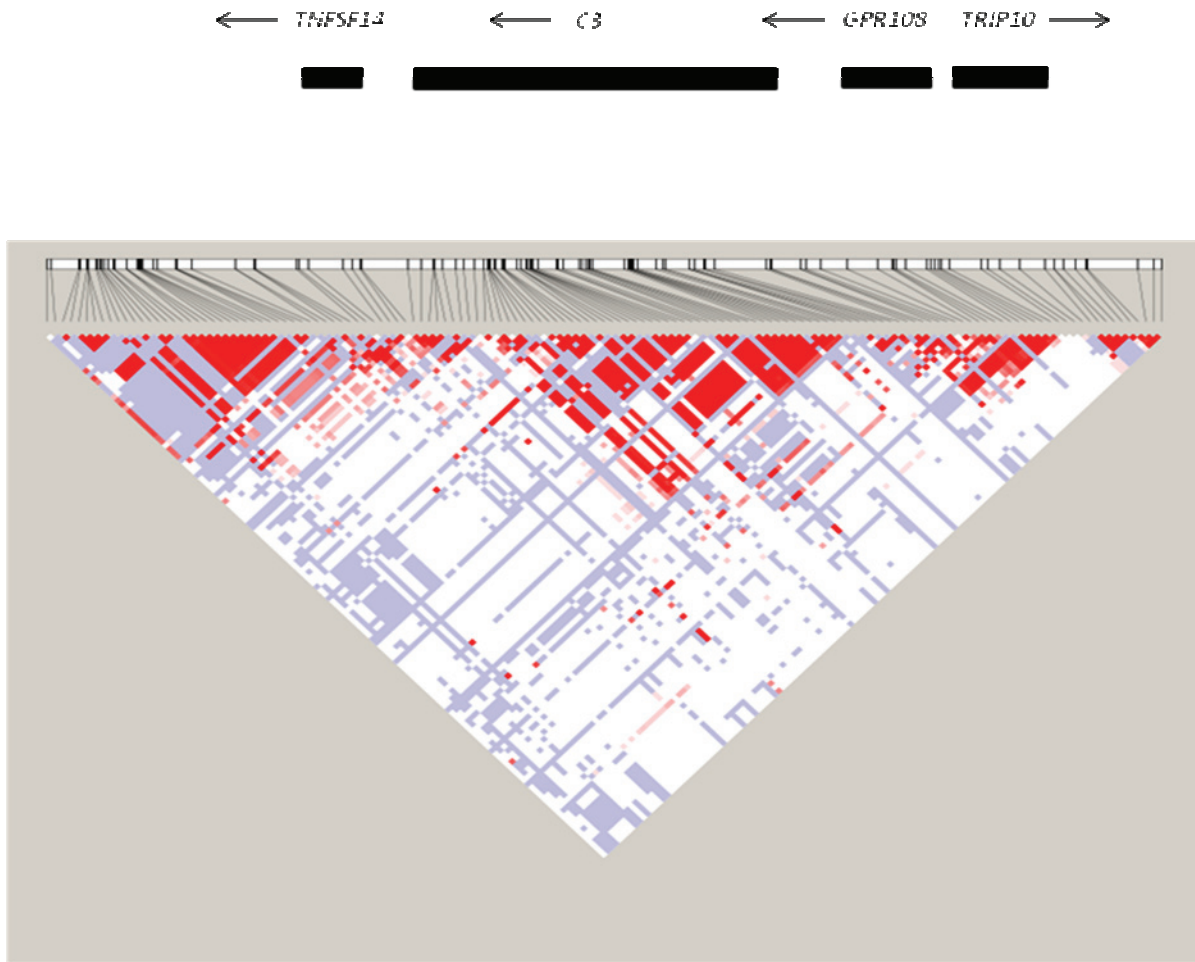
It should be emphasised that the summary of events presented here is simplistic and the reader is referred to the referenced articles for further detail. In particular it should be noted that complement receptors often work in co-operation with other families of receptors in determining the overall cellular response. For example binding to both CR3 and associated immunoglobulin Fc $\gamma$ Rs is required to trigger migration and ingestion by phagocytic cells (Krauss *et al* 1994; Xiong *et al* 2006). It should also be noted that activation of the various complement receptors may also lead to immunomodulatory events, including feedback on the complement pathway itself. For example, activation of CR1 inhibits both the classical and alternative pathway through convertase inhibition, and also acts as a co-factor for the cleavage of C3b to iC3b and C3dg (Krych-Goldberg & Atkinson 2001).

### 1.5.2: C3 – Gene Expression and Genetic Architecture

C3 is expressed in relatively large amounts in the serum of healthy individuals; only albumin and immunoglobulin are present in greater quantity. A suggested reference range of 0.8-1.9mg/mL has been proposed, although it should be noted that there is no internationally recognised standard assay or normal range (Ritchie *et al* 2004). Like CRP, the expression of C3 is regulated as part of the acute-phase response, and levels increase in the presence of infection or inflammation (Gabay & Kushner 1999). In contrast to CRP, the rise in C3 tends to be more modest, with only an approximate doubling of serum concentration (Baumann 1989). The main source of C3 synthesis is the hepatocyte (Alper *et al* 1969). A number of extrahepatic sources of C3 have been described including monocytes/macrophages, renal tubular epithelial and interstitial cells, fibroblasts, endothelial cells, alveolar and intestinal epithelial cells and glial cells (Barnum *et al* 1993; Beatty *et al* 1981; Brooimans *et al* 1991; Johnson & Hetland 1988; Katz & Strunk 1988; Moon *et al* 1997; Strunk *et al* 1988; Ueki *et al* 1987; Warren *et al* 1987; Welch *et al* 1993; Zhao *et al* 2000).

The C3 gene is located on chromosome 19p13.2-3. It is a large gene spanning nearly 43kb, and consisting of 41 exons. Exons 1-16 encode the C3  $\beta$ -chain and exons 17-41 encode the  $\alpha$ -chain (Fong *et al* 1990). **Figure 1.7** demonstrates C3 in its genomic context in relation to neighbouring genes. The majority of studies attempting to delineate the role of cytokines in influencing C3 expression have been conducted *in vitro*, using hepatocyte derived cell lines. A number of cytokines have been shown to increase C3 expression, including IL1 $\alpha$ , tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) and transforming growth factor- $\beta$  (TGF- $\beta$ ) (Andrews *et al* 2003; Mackiewicz *et al* 1990; Perissutti & Tedesco 1994). However, the key cytokines that, individually, trigger C3 expression in these cell lines are IL-6, IL-1 $\beta$  and interferon- $\gamma$  (IFN- $\gamma$ ),

**Figure 1.7.** *C3* in its genomic context on chromosome 19.



In the top section 150kb of DNA sequence is annotated to show *C3* in its genomic location in relation to the neighbouring genes which encode the tumour necrosis factor (ligand) superfamily, member 14 (*TNFSF14*), the G-protein coupled receptor 108 (*GPR108*) and the thyroid hormone receptor interactor 10 (*TRIP10*). The lower section of the diagram is a Haploview plot demonstrating the pattern of LD between genotyped markers (data release 24, CEPH population) within the same genetic interval. The position of the 220 genotyped SNPs relative to the genes is indicated by the black lines above the LD plot. The pairwise LD between any two SNPs is indicated by the colour coding of the corresponding square of the LD plot. High LD is indicated by a pink/red colour scheme (the dark red squares are linked with a  $D' = 1$  and a LOD score  $\geq 2$ )

although studies disagree whether IL-1 $\beta$  can induce C3 expression alone, or only in the presence of IL-6 or IFN- $\gamma$ . Nonetheless, in a pattern that mirrors that seen with CRP, IL-6 and IL-1 $\beta$  appear to act synergistically to induce maximal C3 expression (Andrews *et al* 2003; Darlington *et al* 1986; Falus *et al* 1990).

Compared with CRP, our knowledge of the specific transcription factors involved in C3 expression, or the important C3 promoter elements, is limited. A dissection of the C3 promoter using luciferase constructs suggests that the key region mediating the synergistic response to IL-6 and IL-1 $\beta$  lies between base pairs -127 to -70 (Wilson *et al* 1990). This region contains one, or possibly two, CCAAT/enhancer binding protein regions. The key transcription factor binding this region in IL-1 $\beta$  stimulated cells is C/EBP $\delta$  (in contrast to C/EBP $\beta$  which is the key transcription factor stimulating CRP production). C/EBP $\alpha$  also binds this region and potentially mediates the expression of basal level C3 (Juan *et al* 1993). Additional promoter elements have also been described. Located approximately 150 base pairs upstream of the transcriptional start site is a consensus sequence that forms a farnesoid X receptor (FXR) target. FXR acts as an intracellular bile acid receptor and binding to the C3 promoter has been shown to increase C3 expression. It has been proposed this pathway may be important in regulating gene expression in intestinal epithelial cells (Li *et al* 2005). The C3 promoter also contains three oestrogen response elements (positions -235 to -223, -212 to -180 and -180 to -152) that can bind intracellular oestrogen receptors and stimulate gene expression (Fan *et al* 1996).

An important question, analogous to the problem with CRP, is whether extrahepatic expression of C3 is regulated in the same way as hepatic production. There is limited data on which to draw any conclusions. The key IL-6/IL-1 $\beta$  responsive promoter element would

appear to be functioning in glial cells, suggesting regulation similar to the hepatocellular models (Maranto *et al* 2008). However, a family has been described with C3 deficiency in whom hepatic C3 production was almost completely abrogated, but macrophage C3 expression was maintained; similarly murine models with mutated C3 have been developed in which C3 mRNA was undetectable in hepatic tissue but was measurable in a variety of other tissues (Circolo *et al* 1999; Fey *et al* 1983). Both these lines of evidence suggest there may be tissue specific differences in the genetic regulation of C3 expression.

### *1.5.3: C3 – Genetic Variation, Expression and Disease Susceptibility*

Even prior to the delineation of specific genetic mutations a number C3 structural variants were identified within the population. The best characterised of these variants was termed C3 fast (C3F) and C3 slow (C3S) according to their rate of migration through an agarose gel (Alper & Propp 1968; Wieme & Demeulenaere 1967). Subsequently the molecular basis of this variation was identified as a non-synonymous SNP (now labelled *rs2230199*) in exon three leading to an arginine to glycine change at amino acid position 102 (R102G) (Botto *et al* 1990). On the basis of *in silico* protein modelling it has been suggested that this mutation may have functional consequences. The R102G residue lies in the C3 MG1 domain, which, after C3 activation to C3b, lies in close proximity to the thioester domain. It has therefore been proposed that this mutation could affect thioester activity, or influence the binding of other molecules to the thioester domain, in particular the regulatory protein complement factor H (Jokiranta *et al* 2006; Yates *et al* 2007). Direct function studies, however, have given rather inconsistent results. Although an initial report suggested differences in the ability of the two variants to bind mononuclear cells, subsequent studies showed no difference between the variants in their ability to solubilise immune complexes or bind complement receptors

(Arvilommi 1974; Bartok & Walport 1995; Welch *et al* 1990). Similarly, there is no evidence that carrying either of these variants results in a quantitative difference in C3 level (Bronnestam 1973). Despite this lack of functional evidence, there is now considerable evidence that the C3F/S variant is associated with disease. The strongest evidence is for an association with age-related macular degeneration (AMD) (Maller *et al* 2007; Yates *et al* 2007). An association has also been reported with IgA nephropathy, systemic vasculitis, PR3 ANCA-positive vasculitis, partial lipodystrophy and membranoproliferative glomerulonephritis type II (MPGNII) and long-term graft survival following renal transplantation (Brown *et al* 2006; Finn *et al* 1994; Finn & Mathieson 1993; McLean & Winkelstein 1984; Persson *et al* 1999). Although a number of other structural variants of C3 have been identified, none have been shown to be functionally different or to be associated with disease.

In addition to these non-synonymous C3 SNPs there are nearly 1000 additional SNPs that are either intronic or found in close proximity to the gene, and hence have the potential to alter regulatory sequences. Compared with the non-synonymous polymorphisms, these SNPs are less well studied, and it is only recently that authors have begun to employ a more comprehensive screen of all C3 variants in their studies. Only four studies to date have proposed an association between disease and any of these non-coding SNPs. Three papers have shown an association between C3 variation and bronchial asthma: an initial report and a larger follow-up from a Japanese cohort and a study in an African-American cohort (Barnes *et al* 2006; Hasegawa *et al* 2004; Inoue *et al* 2008). In both cohorts haplotypes were identified that were associated with asthma, and this effect did not appear to be accounted for by the presence of known coding SNPs. The identified regions were different in the two cohorts (spanning intron 19-23 in the African-Americans and a large block incorporating all

of the 5' end of the gene downstream of exon 23 in the Japanese), and clearly further work is required to delineate these effects in more detail.

The final C3 disease association that has been reported is with SLE (Miyagawa *et al* 2008). Given the strong biological plausibility for a role for C3 in the pathogenesis of SLE and the, albeit mild, autoimmune phenotype observed in individuals with C3 deficiency, it is perhaps surprising that this association has only been reported in a single study. The key association was seen for SNP *rs7951*, and haplotypes containing the minor allele of this SNP. This SNP results in a synonymous base pair substitution in exon 37 although, of course, this may not be a functional variant but simply in strong linkage disequilibrium with a true functional variant elsewhere. Importantly, however, these authors also find SNP *rs7951* to be associated with reduced C3 levels in 87 SLE patients on whom serum C3 readings were collected. This not only provides a potential mechanism for the association with SLE, but also represents the first direct association of C3 variation with serum C3 levels. Prior to this it has not even been demonstrated that serum C3 is a genetically determined trait.

## **1.6: Summary and Aims of Thesis**

Although they are structurally unrelated molecules, CRP and C3 share many functional characteristics. Both are phylogenetically ancient plasma proteins that are constitutively expressed in healthy humans. The expression of both molecules is regulated so that serum levels increase in the presence infection or inflammation; this is most notable for CRP, which forms a key component of the human acute-phase response. Both molecules are considered components of the innate immune system, being able to recognise ubiquitous pathogen-expressed molecules. In the case of CRP this recognition occurs directly while in the case of

C3 it is indirect, through activation of the classical, alternative, or mannan-binding pathways of complement. Both molecules also have typical innate immune effector functions, most notably through the opsonisation of potentially pathogenic or immunogenic material, for clearance by phagocytes. In higher animals, both molecules also play sensory and effector roles within the evolved adaptive immune system. This is most notable for C3, which can be potentially activated through the classical complement pathway by cross-linked immunoglobulin. There is also growing evidence that CRP can interact directly with immunoglobulin receptors, and hence it may have complex adaptive immune functions which are not yet fully elucidated.

Serum levels of both proteins are measured in routine clinical practise for diagnostic purposes, while there are several lines of evidence to implicate both these molecules in specific disease processes, most notably systemic lupus erythematosus (a clinical syndrome reflecting the development of systemic autoimmunity) and, for CRP, ischaemic heart disease (a clinical syndrome reflecting the development of atherosclerosis).

Numerous authors have reported an association between *CRP* genotype and serum CRP expression, but there is little consistency in the literature, with no single SNP associated in each study where it has been tested. *CRP* variation has also been associated with both SLE and CHD but again data is limited. For *C3* the genetic data is much more limited with only a single study reporting an association between *C3* genotype and serum C3 levels.

The aims of the work presented in my thesis are therefore as follows:

- To explore the association between *CRP* variation and serum basal CRP expression.
- To replicate findings in cohorts of differing ethnicity and to employ a variety of methodological strategies to ensure the robustness of the observed association.
- To use ethnic diversity at the *CRP* locus to clarify the complex pattern of association reported in the literature.
- To consider the effect of *trans*-acting genetic effects on basal *CRP* expression.
- To perform a pilot study analysing the role of *CRP* variation in determining acute-phase CRP expression in patients with chronic inflammatory disease.
- To re-evaluate the association between *CRP* variation and SLE.
- To demonstrate serum complement C3 level is a genetically determined trait by an analysis of familial clustering and heritability.
- To determine the association between *C3* genotype, serum C3 levels and SLE.

## **CHAPTER 2**

# **MATERIALS AND METHODS**

## 2.1: Study Cohorts

Four cohorts were used for the analyses presented in this thesis: these are described below.

### 2.1.1: UK white European SLE families (Cohort 1).

This cohort constituted a growing collection of predominantly simplex SLE families recruited through United Kingdom rheumatology clinics or by direct patient contact following media publicity. Each proband met ACR 1982 revised criteria for diagnosis (Tan *et al* 1982).

Clinical data and biological samples were collected at a single time-point at study enrolment. Disease phenotype varied, but data collection took place in an outpatient setting from patients who were not “acutely unwell”. Ethical approval was obtained (MREC/98/2/6) and all participants gave appropriate written informed consent. A copy of the patient questionnaire may be found in Appendix C, from which the range of clinical information collected for each subject can be seen.

All biological specimens from the cohort participants were collected prior to the start of the work presented in this thesis, and had been processed by laboratory technical staff. In brief, 40mL whole blood was obtained from each participant for the extraction of DNA by the phenol-chloroform method. After extraction, DNA was quantified and diluted to 50ng/μL for storage at 4°C. An additional 10mL whole blood was obtained for centrifugation and serum collection. Serum was stored frozen at -80°C.

Previous genotyping projects using this cohort led to the identification of a number of ex-paternity families who were excluded from analysis at the outset. In addition only participants

of self-reported white European ethnicity were included. The final cohort used for the analyses presented in chapter 3 therefore consisted of 496 SLE probands of white European ethnicity along with their family members. The availability of data on family members varied such that the final cohort broke down into 323 parent/proband trios, 173 single-parent families (consisting of proband, parent and sibling) and 370 additional unaffected siblings. By the time the analyses presented in chapter 8 were performed, recruitment of additional cases and family members had resulted in an expansion of the cohort to a total of 2211 individuals from 585 families (418 parent/proband trios and 167 additional families). Subsets of this cohort were employed for different components of the analysis presented in this thesis and full details can be found in chapters 3 and 8.

#### *2.1.2: Filipino SLE families (Cohort 2).*

The Filipino cohort is a growing resource collected in collaboration with Dr Sandra Navarra (University of Santo Tomas Hospital, Manila, Philippines). At the time this work was carried out it consisted of 328 ethnic Filipinos with SLE. Each proband met ACR 1982 revised criteria for diagnosis (Tan *et al* 1982). Additional data was also available on 308 unaffected family members, leading to a complex family structure as follows:

- 27 Proband/two parent trios
- 36 Proband/one parent/sib (or sibs) families
- 57 Proband/single parent (no sibs) families
- 62 Proband/sib or sibs (no parents) families
- 146 Probands with no other relatives

Data collection followed the model from cohort 1, with DNA, serum and clinical information collected at a single timepoint at patient recruitment and clinical information obtained via the same questionnaire (Appendix C). Written informed consent was obtained from each subject, and the study protocol was approved by the Institutional Review Board and Ethics Committee of the University of Santo Tomas (UST) Faculty of Medicine and Surgery and UST Hospital. This cohort was employed for the analyses presented in chapter 4.

### *2.1.3: African-American population cohort (Cohort 3).*

This cohort consisted of 594 unrelated individuals of African-American origin. These individuals were from two sources, provided by collaboration with Prof. Philip de Jager (Harvard Medical School/Partners Healthcare Center for Genetics and Genomics, Center for Neurologic Diseases, Brigham & Women's Hospital and Harvard Medical School and Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, USA), Prof. Carlos Mendes de Leon (Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, USA) and Prof. Jorge Oksenberg (Department of Neurology, University of California, San Francisco, USA). Cohort 3a consisted of 294 subjects from the Chicago Health and Aging Project (CHAP), which is a population based study of community-dwelling adults aged 65 years old and over, from the Chicago area, IL, USA. Details on the recruitment and data collection for these individuals may be found in the published literature (Bienias *et al* 2003). Cohort 3b consisted of 150 unrelated multiple sclerosis (MS) cases and 150 controls from the UCSF Multiple Sclerosis Genetics Group. Ascertainment protocols, clinical and demographic characteristics have also been published (Cree *et al* 2004; Oksenberg *et al* 2004). Recruitment

to both cohorts received institutional review board approval and written informed consent was obtained. These cohorts were employed for the analyses presented in chapter 5.

#### *2.1.4: UK Rheumatoid Arthritis cases (Cohort 4).*

This cohort consisted of 281 UK white European rheumatoid arthritis cases, provided in collaboration with Dr Sophia Steer (Kings College London School of Medicine, University of London, UK). Details on the collection of this cohort have been published elsewhere (Steer *et al* 2005). In brief, patients were recruited consecutively from rheumatology clinics across South London; each met 1987 ACR criteria for diagnosis, although disease activity and duration varied between patients (Arnett *et al* 1988). Data was collected for rheumatological parameters including the DAS 28 disease activity score broken down into its constituent components of a tender and swollen joint count for 28 defined joints, physician and patient global assessments of disease activity (on a 10 point visual analogue scale) and an erythrocyte sedimentation rate (ESR). Each patient had a single CRP measurement performed on a sample collected simultaneously to the ESR sample (Prevoo *et al* 1995). This cohort was employed for the analysis presented in chapter 6.

## 2.2: Genotyping

### 2.2.1: Selection of SNPs for Genotyping

An extensive array of SNPs has already been identified for the genes investigated in this thesis, and these data are available on public databases. I therefore decided not to devote resources to DNA sequencing in the search for novel polymorphisms. For the primary genes under investigation, namely *CRP* and *C3*, the aim of the work conducted in Cohorts 1-3 was to build a comprehensive map of common genetic variation at the loci. The selection of SNPs for typing followed a similar pattern for each cohort. Firstly, the Tagger facility of Haploview (using an  $r^2=0.8$  threshold) was used to select tagSNPs that defined the haplotype block or blocks for the relevant gene from the most ethnically appropriate HapMap panel (<http://www.hapmap.org/>) (Barrett *et al* 2005; Gabriel *et al* 2002). For cohort 1 this was taken to be the HapMap European (CEPH) panel, for cohort 2 the HapMap Han Chinese (CHB) panel and for cohort 3 the HapMap Yoruban (YRI) panel. Since Filipinos are not specifically represented in the HapMap datasets the ethnicity considered evolutionarily most closely related to this population, namely Han Chinese, was selected (however see discussion point in section 4.5.4) (Nei & Roychoudhury 1993).

In addition to these tagSNPs, I also aimed to genotype all SNPs on which there was published association data at the time of project planning. While in some cases this was predicted to provide redundant genetic information I felt it was important to perform this typing to facilitate a direct comparison with existing literature. For the *CRP* analysis in cohort 3 (African-Americans) the additional step was also taken to select SNPs to address the question of genetic admixture (the presence in the population of genetic elements of recent European

origin on a background ancestral West African genome). Three SNPs for typing were therefore selected from the HapMap datasets as being monomorphic in the YRI population but polymorphic in CEPH population.

For the study of cohort 4 a different approach was taken. In this study I aimed to assess the association between *CRP* variation and acute phase *CRP* expression, in a population whose *CRP* architecture had already been established from the work in cohort 1. For cohort 4, I therefore only genotyped the minimal selection of tagSNPs required to define the common *CRP* haplotypes

Finally, for some analyses, additional genotyping was undertaken at other loci (namely *IL1B*, *IL6* and *APOE*), with the intention of using these data as genetic covariates in the primary association analyses. Resources precluded a full tagSNP approach for these genes (the LD across *IL1B* in particular is rather weak and would require a considerable number of SNPs to map fully). For *IL1B* and *IL6* we therefore selected a number of SNPs which have been suggested in the literature to have a potential functional role, while for *APOE* we selected the two non-synonymous SNPs which determine Apo  $\epsilon$  serotype.

### *2.2.2: Genotyping by mass spectrometry – theoretical aspects*

The majority of genotyping for the work described in this thesis was performed by me using Matrix-Assisted Laser Desorption and Ionization-Time Of Flight (MALDI-TOF) mass spectrometry (Sequenom, Hamburg, Germany), available at the Imperial College Genome Centre, Hammersmith Campus. The protocol I used varied during the course of this research as the technology for high multiplex assays developed from the Homogeneous

massEXTEND™ (HME) assay, to single base extension assays (iPLEX™ and iPLEX Gold™ protocols).

MALDI-TOF mass spectrometry allows allele specific DNA products to be separated on the basis of their molecular weight (Oeth *et al* 2005; Ross *et al* 1998; Storm *et al* 2003; Tang *et al* 2003). The initial step for all protocols is the amplification of an approximately 100 base-pair fragment containing the SNP of interest, followed by treatment with Shrimp Alkaline Phosphatase (SAP), to dephosphorylate unincorporated dNTPs that would interfere with subsequent steps. An allele specific fragment is then synthesized in an extension reaction, using an extension primer binding to bases just adjacent to the polymorphic site. The mass separation of two products differing by a single base is 9-40Da, depending on the base substitution, and the mass spectrometry genotyping protocols are designed to optimise this allele-specific mass difference, and hence the accuracy of allele discrimination. In the earlier Homogeneous MassEXTEND (hME) protocol the reaction extended from the primer, through the polymorphic site, terminating with a further 2-3 base pair fragment: the additional mass of this fragment allowing larger mass separations between multiplexed products. The later iPLEX and iPLEX Gold protocols involve a single base extension only but using mass modified terminators. By ensuring good mass separation between the extension reaction products (as a consequence of the extension primer design), multiplexing of assays is possible (theoretically up to 36-plex reactions for the iPLEX Gold protocol).

### *2.2.3: Genotyping by mass spectrometry – practical aspects*

Genotyping using Sequenom MALDI-TOF mass spectrometry is a highly automated procedure. All stages were carried out according to Sequenom recommended protocols, with

the exception of the initial PCR step where I used a modified touchdown protocol, previously established in our laboratory to improve genotyping efficiency.

Assay pooling and the design of all primers was performed using Sequenom Assay Design Software v.3.1, with default settings applied. The software automatically optimises primer design to ensuring maximum assay plexing and to minimize primer-primer interactions. All primers were ordered from Metabion (Martinsreid, Germany). High purity HPLC purified extension primers were used, since contamination of the extension reaction with oligomers of similar molecular weight to the genuine extended primers can severely compromise the reliability of the assay.

DNA at the required concentration was diluted from the 50ng/μl stock plates and aliquoted into 384well PCR plates (Abgene, Cambridge, UK) using a Biomek FX robot (Beckman Coulter, Luton, UK). The three laboratory steps are as follows (a full list of reagents and reaction conditions is given in Appendix A):

1. Amplification of genomic DNA by Polymerase Chain Reaction.
2. Dephosphorylation of unincorporated dNTPs using Shrimp Alkaline Phosphatase (SAP).
3. Extension reaction: Either hME or iPLEX reaction

Excess salts were then removed from the extension reaction mixture by the application of Clean Resin (Sequenom, Hamburg, Germany), mixing and centrifugation. Using a robotic nanodispenser, minute aliquots of the clear supernatant were dispensed onto a SpectroCHIP®

(Sequenom, Hamburg, Germany), which was inserted into the mass spectrometer for data acquisition.

Following data acquisition, analysis was performed using Sequenom TYPER v.3.4 software. In essence this software is able to identify peaks on the mass spectrometer output at positions corresponding to the predicted mass of the allele specific extension product for each of the multiplexed assays. By assessing whether one (homozygotes) or two (heterozygotes) extension peaks are present the software can automatically assign a sample's genotype. The algorithm is also able to base a decision to call a genotype on quality control parameters such as the level of noise in the background spectrum (often indicative of salt contamination), or the relative heights of the extended and unextended primer peaks (indicative of reaction efficiency). The default genotype calling settings were used throughout and in all cases genotypes were reviewed manually and considered to be sound calls. In a very limited number of cases additional manual calls were made where peaks could be clearly distinguished which were symmetrically centred on the line of predicted mass, but where automated calling may have failed on the basis of reaction efficiency.

Genotyping for cohort 3a was performed using the Sequenom platform by colleagues at the Broad Institute Center for Genotyping and Analysis (Cambridge, MA, USA). The genotyping method employed for cohort 3a was therefore identical to that used for the in-house genotyping, and is a robust and recognised technology. Unfortunately, due to financial and organisational constraints, an exchange of DNA samples between centres to definitively ensure reproducibility of genotyping was not possible.

#### *2.2.4: Genotyping – Quality Control*

To ensure genotyping was of high quality I adopted a series of pre-determined quality control criteria, below which data were considered unreliable for further analysis. These were:

- Genotyping for any SNP successful in >90% of individuals
- Genotyping in any individual successful for >80% of markers
- Genotypes in Hardy-Weinberg equilibrium (Hardy-Weinburg Probability (HWP)

$P < 0.01$  as assessed by  $\chi^2$ ). For family cohorts this was assessed in the subset of unrelated population founders.

For the family cohorts data were also checked for Mendelian inconsistencies using PedCheck (O'Connell & Weeks 1998). Individuals were excluded from analysis if >10% of markers demonstrated Mendelian inconsistency. If this individual was a disease proband, or if it was ambiguous which individual in the family showed inconsistency then the whole family was excluded from analysis.

### **2.3: Immunological assays**

#### *2.3.1: CRP quantification – Immunoturbimetry*

Serum CRP was quantified by latex-enhanced immunoturbimetry using an Olympus AU2700 analyser (Olympus Diagnostics, Southall, UK) in the Department of Clinical Chemistry, Hammersmith Hospital, London, UK. To summarise this fully automated procedure, CRP is quantified by mixing a 2 $\mu$ L serum aliquot with a suspension of anti-human CRP antibody-

coated latex particles, resulting in the formation of insoluble aggregates. The absorbance of light passed through these aggregates is proportional to the CRP concentration in the solution (Whicher *et al* 1983). Using standard settings the analytical range (over which the relationship between absorbance and concentration is linear) is 0.2-480mg/L. The Olympus OSR1699 assay is well validated and used in routine clinical use. Local quality control was ensured by re-calibration of the immunoturbidimeter every six hours.

CRP quantification on the 294 individuals from cohort 3a was quantified by a particle-enhanced immunonephelometric assay using a Dade-Behring BNII nephelometer at University of Vermont Laboratory for Clinical Biochemistry Research (Dade Behring, Deerfield, IL, USA). Although it was not ideal to use this additional method of CRP quantification, it should be noted that there are internationally recognised CRP reference samples which should ensure direct comparability between assays.

### *2.3.2: C3 quantification – Radial Immunodiffusion Assay*

Serum C3 levels were quantified by Radial Immunodiffusion Assay (RID) using the method described by Mancini (Mancini *et al* 1965). 64µL sheep anti-human C3 antibody (The binding site, Birmingham, UK) were added to 10mL 1.2% Complement fixation test agarose (CFT dilutant tablets from Oxoid Ltd, Cambridge, UK). The molten agarose was allowed to solidify on a 64cm<sup>2</sup> glass plate and 30 x 2mm wells were punched. Sample sera were thawed from -80°C storage and kept on ice. 5µL was added to each well and then the plates were incubated overnight in a wet box at room temperature. After washing with phosphate buffered saline, the gels were dried onto Gelbond film (Cambrex Bio Science, New Jersey,

USA) and stained with Coomassie brilliant blue. The diameter of the stained ring of precipitation was measured to the nearest 1µm using a graduated eye piece.

For each plate of assays the C3 concentration was determined by reference to a standard curve of dilutions of pooled human serum (whose concentration had previously been determined by validated assay in an external laboratory). Plate to plate consistency was ensured by the consistent use of control samples from two individuals with low and high C3 respectively. To confirm that none of the C3 plates gave unexpected, outlying, results the Kolmogorov-Smirnov test was used to assess whether the distribution of control samples departed from the normal distribution expected due to standard measurement error. No significant deviation was seen ( $P=0.20$ ).

## **2.4: Statistical analysis**

In this section I have outlined the statistical methods used for the analysis of data presented in this thesis. A brief discussion of the theoretical background to some of these methods may be found in section 1.3.

### *2.4.1: Data storage and management*

Data was stored and managed using BC/GENE v.2.2.5 software (Biocomputing platforms Ltd, Espoo, Finland). This facilitated the formatting of genotype and phenotype data, and the running of dedicated family-based genetic analysis packages directly through a customised interface. Except in cases specified below, all other statistical analyses were performed using the SPSS v. 14.0 (Statistical Package for the Social Sciences; SPSS inc., Chicago, IL, USA)

#### 2.4.2: Heritability estimates

The concept of heritability has already been introduced in section 1.3.2. In analyses presented in this thesis we have used a number of different methods to estimate the heritability of CRP and C3 in the family-based cohorts. The simplest method, applied in section 8.2.1, was a regression based approach of estimating heritability from the slope of the line of the regression of offspring C3 on mean parental C3 (Khoury *et al* 1993; Visscher *et al* 2008). In this case an unaffected sibling was selected from each family at random for analysis. The alternative approach that was used was to fit a variance components model to estimate the variance attributable to additive genetic effects. In the CRP analysis presented in section 3.3 this was done using a maximum likelihood approach within the qTDT package, modelling for polygenic additive genetic variance and common environmental variance (Abecasis *et al* 2000; Hopper & Mathews 1982).

For the revised C3 heritability estimate, presented in section 7.2.2, a Bayesian approach to variance components analysis was adopted, and modelled using WinBUGS (Burton *et al* 1999; Lunn *et al* 2000). Variances were modelled for of additive polygenic effects, shared nuclear family effects and common sibling effects (i.e. not shared with parents). A weakly informative prior distribution was specified for the within family correlations (Half normal (0,10) distribution). This was felt to be more appropriate than a non-informative prior since family members are obviously correlated to some extent, while it would be inappropriate to give equal weight to all levels of correlation. Nonetheless, sensitivity to different prior distributions for the variance components was also tested by running the same models with a  $\gamma$  (2, 2) distribution, and different shapes of half normal distribution (0, 0.25) and (0, 0.4). The results did not change to any appreciable extent.

### 2.4.3: Data Imputation

The level of missing genotype and phenotype data was low across all of the cohorts and, except where specified, analyses were conducted on “raw data” with no imputation of missing values. Nonetheless, even small amounts of missing data can lead to a considerable reduction in the size of the analysable cohort if multiple phenotypic or genotypic covariates are considered together in the same model. Therefore, for the analysis of cohort 3 a Bayesian approach was adopted for imputing missing data, by inputting a prior distribution for missing variables, with information from existing data updating these priors. For example, existing lnBMI data had a mean 3.32 kg/m<sup>2</sup>, s.d. 0.22, so a normal prior distribution with these parameters was used for missing values. A Bernoulli prior distribution was used for binary phenotype data. Missing SNP data was estimated using FastPhase (Scheet & Stephens 2006). In practise the levels of missing genotype data were low (see **table 5.1**) as were the levels of missing data for phenotypes collected in both halves of the cohort (0.18% for age and none for sex or multiple sclerosis affection). Data on BMI and other disease states were not collected for cohort 3b and hence were missing for 49.2% of the combined data set. It was only for these variables that the use of imputed data was likely to impact upon the observed results. To ensure that the use of these data did not dramatically alter our findings all significant models were also re-evaluated without the use of inferred values. This is discussed further in chapter 5.

In the analysis of association between *C3* genotype and SLE, presented in section 7.4.2., data imputation was used to estimate genotype at an extensive array of *untyped* SNPs. Using the pattern of LD predicted from the HapMAP CEPH panel data, and informed by existing direct genotype data, the programme BEAGLE 3.0 was used to impute 287 untyped genotypes

spanning the entire *C3* gene. BEAGLE 3.0 uses a haplotype clustering algorithm and applies a Hidden Markov Model approach to inferring missing data: it is also capable of inferring data in family cohorts (Browning & Browning 2009; Browning & Browning 2007b).

#### *2.4.4: Haplotype imputation*

Population haplotype structure was inferred for each cohort using Haploview v.4.1 and the algorithm of Gabriel *et al.* with default settings applied (Barrett *et al* 2005; Gabriel *et al* 2002). In the family cohorts haplotype inference was performed using the subset of unaffected cohort founders only. Some genetic association packages, including TRANSMIT and qTDT incorporate haplotype inference algorithms in their analysis, but do not provide an output of individual haplotypes on a case by case basis. Where explicit inference of individual haplotypes was required this was performed using PHASE v.2.1 (Stephens *et al* 2001; Stephens & Donnelly 2003). For family cohorts, where particular genotypes or haplotypes may cluster in families, PHASE can potentially overestimate the occurrence of rare haplotypes. For family datasets I therefore re-checked haplotype assignment using MERLIN, which infers haplotypes on the basis of patterns of gene flow within families (Abecasis *et al* 2002). Haplotype cladograms were inferred using TCS: Phylogenetic network association using statistical parsimony v1.21 (Clement *et al* 2000).

#### *2.4.5: Association analyses*

Transformation of quantitative data was performed where relevant. The distribution of serum CRP in each of the study cohorts was positively skewed, so a logarithmic transformation was employed to approximate a normal distribution and facilitate analysis of data on the

continuous scale. Other than the presentation of basic CRP distributions within the cohorts (i.e. median and range) all analyses are performed on log-transformed CRP, unless specified. Serum C3 approximated a normal distribution *in vivo* so no transformation was required.

For the population-based cohorts (Cohorts 3 and 4) the association between genotype and CRP or C3 as a quantitative outcome was assessed using linear regression, incorporating both continuous and categorical covariates. The clade-based haplotype association presented in section 5.4.3 was modelled by ANOVA using Treescan v1.0: a bioinformatic application to search for genotype/phenotype associations using haplotype trees (available at <http://Darwin.uvigo.es>) (Templeton *et al* 2005). A different approach was required to analyse association in the family cohorts (Cohorts 1 and 2), to account for the clustering of effects within groups of related individuals. Analysis of the association between genotype and serum CRP concentration was performed either by introducing a random effects factor into the linear regression models described above, or alternatively by using the dedicated family based analysis, qTDT, specifying the orthogonal model of association (Abecasis *et al* 2000). For the association of genotype with serum C3 a similar random effects factor was introduced for the maximum likelihood analysis. Bayesian association analysis was performed using WinBUGs with the familial clustering of data analysed through the modelling of variance components as for the heritability estimate (section 2.5.2). A standardised normal (0, 1) distribution was specified as a prior distribution for all model parameters.

The analysis of genetic association with a binary (disease status) outcome was performed by logistic regression in the population based cohorts. In the family-based cohorts it was performed by transmission disequilibrium test using TRANSMIT (Clayton 1999).

Association analysis of the imputed C3 genotype data was carried out by an allelic test

(Fisher's exact test on the 2 x 2 table of transmitted/untransmitted allele) using BEAGLE (Browning & Browning 2007a).

As outlined in section 1.3.4, I adopted one of two approaches to the adjustment of significance levels to account for multiple hypothesis testing. Either I used a modified Bonferroni correction to account for the degree of correlation between markers tested, or alternatively I used a permutation test-based approach (Armitage & Parmar 1986; Sankoh *et al* 1997; Westfall & Young 1993).

#### 2.4.6: Admixture Analysis

As I outlined in section 1.3.4, a significant proportion of the African American genome (perhaps 10-30%) may arise from recent admixture with European populations (Wassel Fyr *et al* 2007). At the level of a single locus this may complicate analyses by distorting the inferred haplotype structure, if the pattern of LD differs between the background and the admixed genetic elements. It may also affect our ability to detect an indirect genetic association. For the analysis of Cohort 3 I therefore identified three SNPs (*rs1255606*, *rs11585798* and *rs17459580*) predicted to be polymorphic in European populations but monomorphic in populations of an African origin. I used PHASE v.2.1 to explicitly infer a haplotype pair for each individual and I hypothesised that haplotypes containing these SNPs would have a European origin.

#### 2.4.7: Model Choice Analyses

The Bayesian model choice analyses presented in chapter 5 were performed using the association studies toolkit for WinBUGS, employing a reverse jump algorithm on the model space, in the Markov Chain Monte Carlo (MCMC) framework (Lunn *et al* 2000; Lunn *et al* 2006). This Bayesian approach allowed me to make some prior assumptions about the nature of the genetic model, on the basis of existing literature. Firstly I assumed that the magnitude of genetic effect would be relatively small, and secondly that the genetic model would be most likely to incorporate one or two genetic effects but was much less likely to have more than three effects (regression parameters  $\beta$  and  $\gamma$ , normal distribution, mean=0, variance=0.25; number of genetic effects specified by a Poisson (2) distribution). In a duplicate analysis I re-ran the Bayesian model selection using vague prior assumptions; in other words making no assumption about the magnitude of genetic effect or number of markers in the model (regression slope parameters  $\alpha$ ,  $\beta$  and  $\gamma$  - normal distribution, mean=0, variance=100,000; residual variance ( $\sigma^2$ ) - inverse gamma distribution, shape 0.01, scale 0001). By using vague priors similar results to those obtained using a classical approach would be expected.

The weight of evidence to support a particular model was the assessed by the Bayes factor. Harold Jeffries suggested a scale for the interpretation of Bayes Factors and this is reproduced in **table 2.1**. (Jeffreys 1961).

**Table 2.1.** Jeffreys' interpretation of Bayes Factors

| Bayes Factor | Strength of Evidence favouring model           |
|--------------|------------------------------------------------|
| <1           | Negative Evidence (supports alternative model) |
| 1 to 3.2     | Barely worth a mention                         |
| 3.2 to 10    | Substantial                                    |
| 10 to 32     | Strong                                         |
| 30 to 100    | Very Strong                                    |
| >100         | Decisive                                       |

## **CHAPTER 3**

# ***CRP* ASSOCIATIONS IN UK SLE FAMILIES**

### 3.1: *CRP* Associations in UK SLE Families – Study Aims

The work presented in this chapter was planned as a family-based *CRP* association analysis in a cohort of white European ancestry. I aimed for comprehensive coverage of genetic variation through the use of *CRP* tagSNPs and, to ensure comparability with existing studies, I aimed to genotype all SNPs on which there was existing published *CRP* association data. I also planned to consider the effect of relevant *trans*-acting genetic variants, and to include these as genetic covariates in the *CRP* analysis. By using a family-based cohort I hoped to demonstrate robust associations between *CRP* genotype and serum CRP. In addition, using a variance components approach with this family data, I planned to estimate CRP heritability and to quantify the proportion of the genetic component of CRP variation accounted for by *CRP* genotype. Finally I planned to refine and extend the association between *CRP* genotype and SLE that was initially reported by Russell *et al* in a cohort that overlapped with the one presented here (Russell *et al* 2004).

An initial power calculation, assuming 500 informative families, an allelic odds ratio of 1.2 and SNP a minor allele frequency of 33% (based on the frequency of rs1205), suggested the power to detect an association between genotype and SLE at the  $P=0.05$  level by TDT was 45% (power calculator at <http://pengu.mgh.harvard.edu/~purcell/gpc/dtdt.html>). An estimate of power to detect a genetic association with serum CRP as a quantitative trait is more complex because it requires an estimate of effect of genetics on both the mean and the variance of the log-transformed CRP distribution, and in addition needs to incorporate an estimate of the within family correlation of CRP levels in this cohort of related individuals. This calculation would perhaps best performed by data simulation, however, since estimates for the variance of log-transformed CRP, and the within family CRP correlation are not

widely reported the estimates of these parameter would be so speculative as to render the calculation almost meaningless. A formal *ad hoc* power calculation was not therefore performed. Despite the relatively low estimated power of the disease-based analysis and the unknown power of the CRP-based study I felt it worthwhile continuing these studies because both associations had already been demonstrated by Russell *et al* in a smaller cohort which overlapped with the one presented here. The data presented in this chapter may be found in published form (Rhodes *et al* 2008a).

### **3.2: UK SLE Cohort - Demographics, CRP Distribution and Heritability.**

Basal CRP was quantified in all family members from the first 294 families (numbered in order of recruitment and not systematically with respect to any phenotypic manifestation). Serum CRP was quantified in additional individuals from later families only if they carried a rarer *CRP* haplotype, whether or not they had SLE. Therefore, from the cohort of 1858 individuals with genetic data, serum CRP was available for 1265. Median CRP was 2.01mg/L (0.19 to 161.51mg/L), mean age was 54.0, 66.1% were female and 24.8% had SLE. A linear regression model, presented in **table 3.1**, showed that CRP increased with age and had a trend towards higher levels in women. CRP was also higher in SLE probands.

**Table 3.1.** Effect of age, sex and SLE affection on CRP

| Variable               | $\beta$ -coefficient | standard error | t     | P-value |
|------------------------|----------------------|----------------|-------|---------|
| Total cohort           |                      |                |       |         |
| Age                    | 0.018                | 0.003          | 6.450 | <0.001  |
| Sex (Female)           | 0.148                | 0.077          | 1.920 | 0.055   |
| SLE (Affected)         | 0.233                | 0.089          | 2.623 | 0.009   |
| Cohort with CRP<10mg/L |                      |                |       |         |
| Age                    | 0.017                | 0.002          | 6.992 | <0.001  |
| Sex (Female)           | 0.171                | 0.068          | 2.510 | 0.012   |
| SLE (Affected)         | 0.024                | 0.08           | 0.301 | 0.764   |

It is unknown whether basal and acute-phase *CRP* expression is regulated by the same genetic factors. Since epidemiological studies have suggested the 99th centile for CRP is 10mg/L in a young adult population, we therefore excluded all individuals with a CRP >10mg/L from further analysis on the basis that their CRP may represent an acute-phase response (Shine *et al* 1981). Interestingly, however, the unaffected siblings in our cohort (generally healthy young adults) had CRP levels higher than this, with the 99th centile 13.01mg/L. The cohort of individuals with CRP<10mg/L contained 1152 individuals with a mean age of 54.0. 65.2% were female and 23.4% had SLE. A linear regression model in this basal CRP cohort, (also presented in **table 3.1**) demonstrated, that basal CRP increased with age and in females. SLE itself showed no association with basal CRP once these individuals with acute-phase CRP levels were excluded ( $P=0.764$ ).

CRP heritability was estimated by variance components modelling in qTDT, which uses the cohort family structure to estimate environmental and genetic components of variance. This gave a maximum likelihood estimate of 27.7% heritability ( $h^2=0.277$ ;  $P=1 \times 10^{-7}$ ) after adjustment for age, sex and SLE affection. In an analysis restricted to unaffected family

members we estimated 21.3% heritability ( $h^2=0.213$ ;  $P=0.001$ ), the reduced level of significance presumably being a consequence of the smaller number of siblings in this subset.

### 3.3: CRP Genetic Architecture

Twenty-eight *CRP* SNPs and an [AT/-] insertion/deletion polymorphism were selected for genotyping, using the approach outlined above and in chapter 2. In addition five *IL6* markers and four at *IL1B* were typed to encompass previous reports and provide additional coverage across these genes. Finally the two non-synonymous *APOE* polymorphisms required to infer Apo ε serotype were genotyped. Detailed marker information is presented in **table 3.2**.

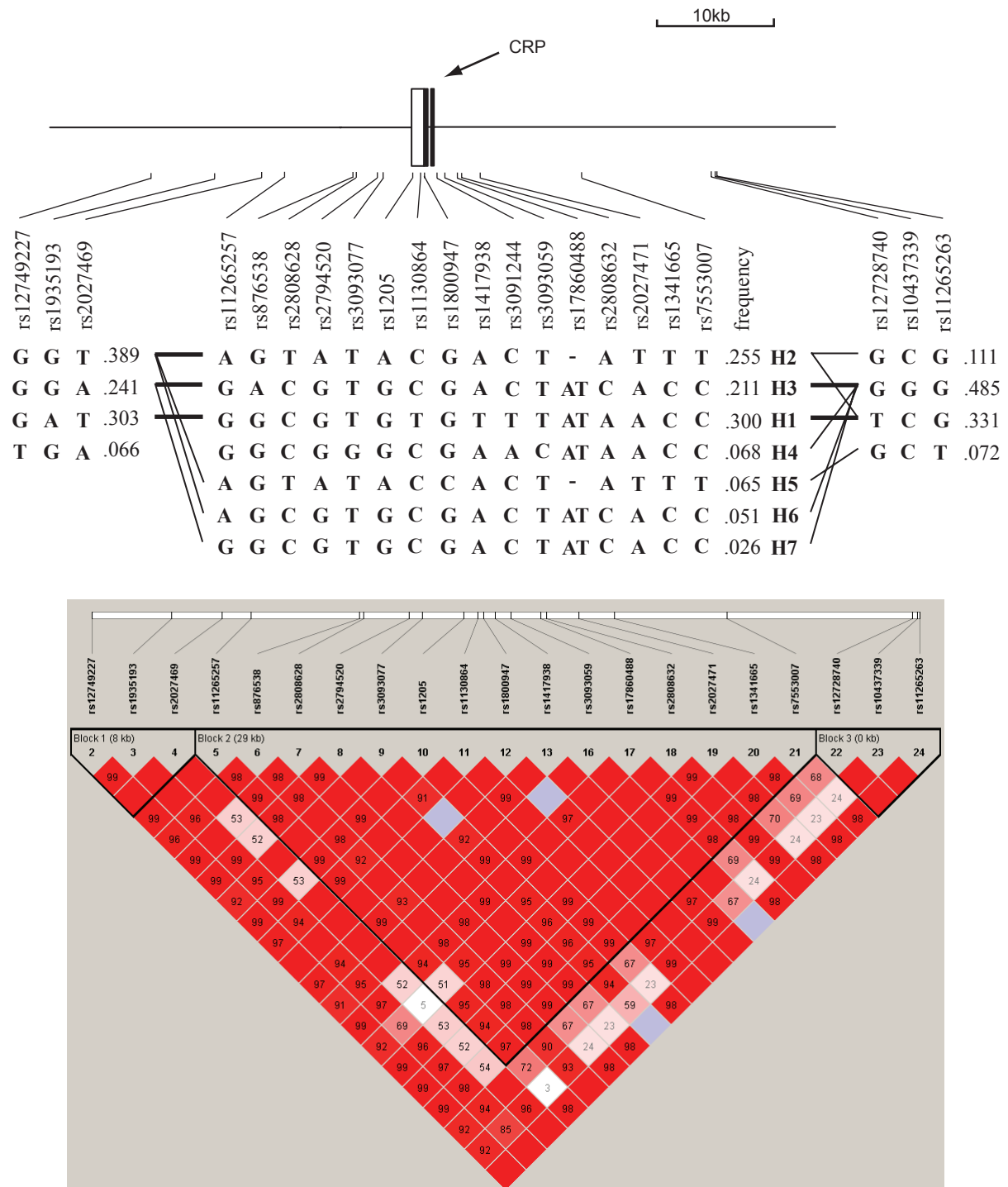
SNPs *rs2794521* and *rs2808630* were excluded from further analysis as they typed in <90% of individuals, while *rs3095058*, *rs3093062*, *rs3093063* and *rs16842502* were excluded as they had very low minor allele frequency in this population. The remaining SNPs were in Hardy-Weinberg equilibrium ( $\chi^2$   $P>0.05$  in the subset of unaffected, unrelated population founders). **Figure 3.1** demonstrates the haplotype structure in the 799 unrelated cohort parents on whom genetic information was available. Seven haplotypes, H1 to H7, occurred with a frequency >1% between *rs11265257* and *rs7553007*. A full pairwise comparison between markers is presented in **table 3.3**. With the scale of fine mapping adopted in this study I was able to demonstrate a falling off of LD both upstream and downstream of the central *CRP* haplotype block. It should be noted, however, that this entire region would fall within the region of strong LD surrounding CRP on a genomic scale (as illustrated in **figure 1.3**.)

**Table 3.2.** Typed SNPs demonstrating alleles and rare allele frequencies

| rs number                | SNP   | Location            | Rare allele frequency | HWP <sup>§</sup> | missing data (%) |
|--------------------------|-------|---------------------|-----------------------|------------------|------------------|
| <i>CRP</i> SNPs          |       |                     |                       |                  |                  |
| <i>rs11265263</i>        | G>T   | 5' flank            | 0.07                  | 0.76             | 2.0              |
| <i>rs10437339</i>        | C>G   | 5' flank            | 0.49                  | 0.59             | 3.6              |
| <i>rs12728740</i>        | G>T   | 5' flank            | 0.32                  | 0.66             | 2.4              |
| <i>rs7553007</i>         | C>T   | 5' flank            | 0.33                  | 0.55             | 3.3              |
| <i>rs1341665*</i>        | C>T   | 5' flank            | 0.33                  | 0.32             | 3.7              |
| <i>rs2027471</i>         | A>T   | 5' flank            | 0.34                  | 0.26             | 3.0              |
| <i>rs2808632</i>         | A>C   | 5' flank            | 0.30                  | 0.08             | 1.9              |
| <i>rs17860488</i>        | AT>-  | 5' flank            | 0.34                  | 0.69             | 5.7              |
| <i>rs3093058</i>         | A>T   | 5' flank            | 0.00                  | N/A              | 2.2              |
| <i>rs3093059*</i>        | T>C   | 5' flank            | 0.07                  | 0.06             | 3.1              |
| <i>rs2794521</i>         | A>G   | 5' flank            | 0.30                  | 0.24             | 14.8             |
| <i>rs3093062</i>         | G>A   | upstream promoter   | 0.00                  | N/A              | 2.8              |
| <i>rs3091244*</i>        | C>T>A | upstream promoter   | 0.29(T)/ 0.07(A)      | 0.07             | 1.9              |
| <i>rs3093063</i>         | C>T   | upstream promoter   | 0.00                  | N/A              | 5.8              |
| <b><i>rs1417938*</i></b> | A>T   | intron 1            | 0.30                  | 0.81             | 2.2              |
| <b><i>rs1800947*</i></b> | G>C   | exon 2 (synonymous) | 0.07                  | 0.59             | 5.5              |
| <i>rs1130864*</i>        | C>T   | 3'-UTR              | 0.30                  | 0.98             | 3.6              |
| <i>rs3093066</i>         | C>A   | 3'-UTR              | 0.00                  | 0.97             | 2.4              |
| <b><i>rs1205*</i></b>    | G>A   | 3'-UTR              | 0.33                  | 0.34             | 2.4              |
| <i>rs2808630</i>         | A>G   | 3' flank            | 0.30                  | 0.23             | 14.0             |
| <b><i>rs3093077*</i></b> | T>G   | 3' flank            | 0.08                  | 0.07             | 1.9              |
| <i>rs2794520</i>         | G>A   | 3' flank            | 0.33                  | 0.41             | 2.8              |
| <i>rs2808628</i>         | C>T   | 3' flank            | 0.33                  | 0.40             | 2.3              |
| <b><i>rs876538</i></b>   | A>G   | 3' flank            | 0.22                  | 0.21             | 3.2              |
| <b><i>rs11265257</i></b> | G>A   | 3' flank            | 0.38                  | 0.67             | 3.6              |
| <i>rs2027469</i>         | C>T   | 3' flank            | 0.18                  | 0.94             | 2.5              |
| <i>rs1935193</i>         | T>A   | 3' flank            | 0.32                  | 0.26             | 5.0              |
| <i>rs12749227</i>        | G>A   | 3' flank            | 0.29                  | 0.71             | 2.6              |
| <i>IL1B</i> SNPs         |       |                     |                       |                  |                  |
| <i>rs16944</i>           | C>T   | 5' flank            | 0.34                  | 0.82             | 5.1              |
| <i>rs1143634</i>         | C>T   | exon 5 (synonymous) | 0.23                  | 0.88             | 2.5              |
| <i>rs7596461</i>         | G>A   | 3' flank            | 0.08                  | 0.38             | 2.7              |
| <i>rs4849124</i>         | C>T   | 3' flank            | 0.30                  | 0.26             | 5.2              |
| <i>IL6</i> SNPs          |       |                     |                       |                  |                  |
| <i>rs1800796</i>         | G>C   | 5' flank            | 0.07                  | 0.56             | 2.2              |
| <i>rs1800795</i>         | G>C   | 5' flank            | 0.39                  | 0.25             | 6.0              |
| <i>rs2069837</i>         | A>G   | intron 2            | 0.07                  | 0.31             | 3.3              |
| <i>rs2069840</i>         | C>G   | intron 3            | 0.34                  | 0.43             | 2.7              |
| <i>rs1554606</i>         | G>T   | intron 3            | 0.43                  | 0.70             | 3.6              |
| <i>APOE</i> SNPs         |       |                     |                       |                  |                  |
| <i>rs429358</i>          | T>C   | exon 4              | 0.13                  | 0.61             | 2.7              |
| <i>rs7412</i>            | C>T   | exon 4              | 0.08                  | 0.31             | 3.3              |

SNPs highlighted in bold have been selected to tag the key *CRP* haplotype block. SNPs marked with a \* have been previously associated with serum CRP levels (see **table 1.2**) <sup>§</sup> Hardy-Weinburg probability *P*-value calculated by  $\chi^2$

**Figure 3.1.** *CRP* haplotypes and LD structure.



The position of SNPs relative to *CRP* is shown in the top section of this diagram. The middle section demonstrates inferred haplotype structure with haplotype frequencies are shown to the right of each block. Thick line connect blocks occurring together in >10% of chromosomes and thin lines if in >1% of chromosomes. The lower section demonstrates the pairwise LD between markers (a graphical representation of the data in **table 3.3**) with haplotype blocks defined by the default algorithm of Gabriel *et al* (see methods section 2.4.4). The pairwise LD between any two SNPs is indicated by the colour coding of the corresponding square of the LD plot. High LD is indicated by a pink/red colour scheme (the dark red squares are linked with a  $D' = 1$  and a LOD score  $\geq 2$ ). Numbers within the squares correspond to the  $D'$  score, with 96 representing a  $D' = 0.96$  and 44 a  $D' = 0.44$  etc. Where squares have no number the  $D' = 1$ .

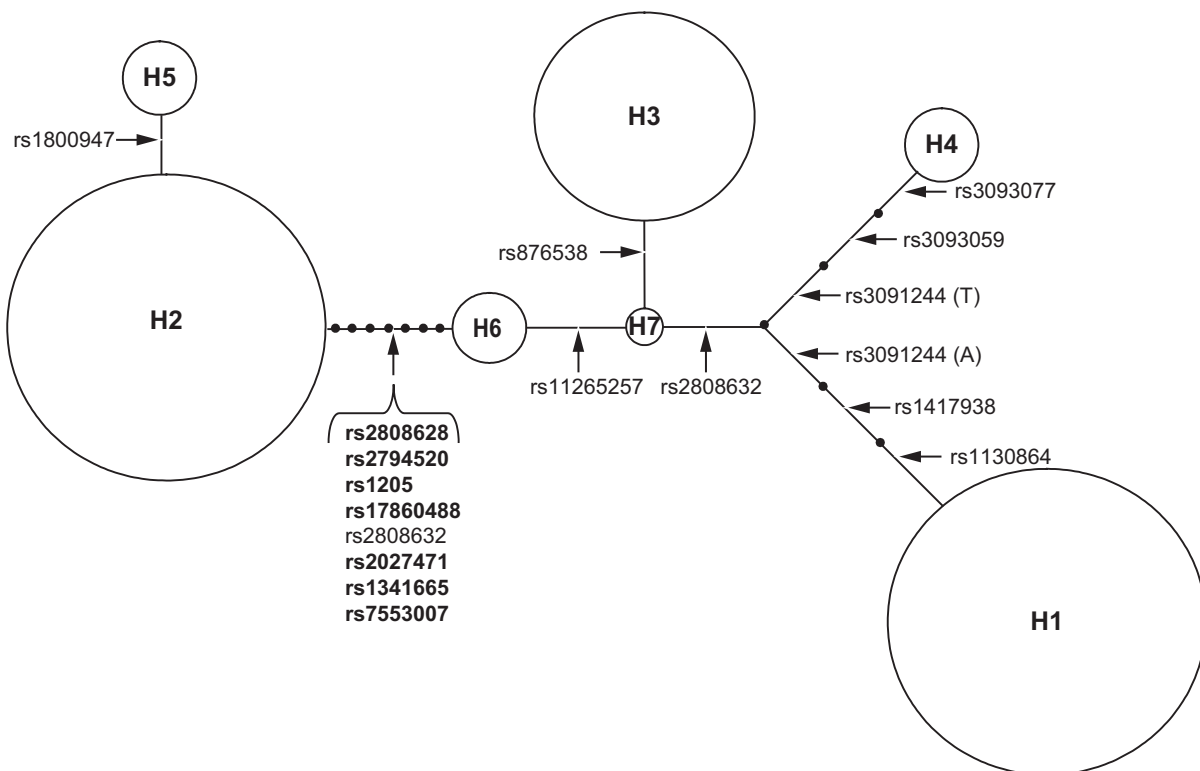
**Table 3.3.** Pairwise linkage disequilibrium between *CRP* SNPs, calculated in the unrelated parents\*

|              |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |              |            |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------------|------------|
| rs12749227   | X     | 0.192 | 0.096 | 0.265 | 0.113 | 0.210 | 0.212 | 0.026 | 0.212 | 0.932 | 0.030 | 0.935 | 0.001 | 0.931 | 0.024 | 0.221 | 0.157 | 0.219 | 0.218 | 0.212 | 0.815 | 0.356 | 0.032 | rs11265263   |            |
| rs1935193    | 0.989 | X     | 0.099 | 0.271 | 0.593 | 0.214 | 0.213 | 0.162 | 0.215 | 0.175 | 0.031 | 0.177 | 0.113 | 0.149 | 0.157 | 0.217 | 0.463 | 0.211 | 0.212 | 0.210 | 0.194 | 0.400 | 0.035 | rs10437339   |            |
| rs2027469    | 1.000 | 1.000 | X     | 0.354 | 0.058 | 0.122 | 0.122 | 0.016 | 0.121 | 0.093 | 0.015 | 0.094 | 0.033 | 0.109 | 0.016 | 0.115 | 0.000 | 0.121 | 0.119 | 0.128 | 0.106 | 0.222 | 0.017 | rs12728740   |            |
| rs11265257   | 0.992 | 1.000 | 1.000 | X     | 0.167 | 0.786 | 0.788 | 0.044 | 0.789 | 0.261 | 0.116 | 0.264 | 0.111 | 0.313 | 0.044 | 0.763 | 0.075 | 0.751 | 0.749 | 0.795 | 0.169 | 0.002 | 0.124 | rs1341665    |            |
| rs876538     | 0.966 | 0.980 | 1.000 | 1.000 | X     | 0.132 | 0.131 | 0.021 | 0.132 | 0.103 | 0.020 | 0.104 | 0.045 | 0.124 | 0.020 | 0.141 | 0.625 | 0.139 | 0.137 | 0.132 | 0.114 | 0.258 | 0.021 | rs2027471    |            |
| rs2808628    | 0.990 | 1.000 | 0.524 | 0.993 | 1.000 | X     | 0.991 | 0.036 | 0.991 | 0.212 | 0.141 | 0.214 | 0.086 | 0.261 | 0.036 | 0.951 | 0.195 | 0.953 | 0.950 | 0.988 | 0.120 | 0.155 | 0.036 | rs2088632    |            |
| rs2794520    | 0.990 | 1.000 | 0.522 | 0.993 | 1.000 | 0.997 | X     | 0.036 | 1.000 | 0.214 | 0.137 | 0.216 | 0.085 | 0.256 | 0.036 | 0.961 | 0.165 | 0.953 | 0.957 | 0.988 | 0.119 | 0.028 | 0.152 | rs17860488   |            |
| rs3093077    | 0.877 | 0.982 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.036 | 0.033 | 0.005 | 0.033 | 0.952 | 0.001 | 0.970 | 0.037 | 0.030 | 0.034 | 0.034 | 0.032 | 0.038 | 0.062 | 0.005 | rs3093059    |            |
| rs1205       | 0.990 | 1.000 | 0.523 | 0.993 | 1.000 | 0.997 | 1.000 | 1.000 | X     | 0.214 | 0.142 | 0.216 | 0.085 | 0.257 | 0.036 | 0.955 | 0.196 | 0.954 | 0.954 | 0.988 | 0.120 | 0.028 | 0.155 | rs3091244C/T |            |
| rs1130864    | 0.974 | 0.954 | 1.000 | 0.992 | 0.928 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.030 | 0.997 | 0.000 | 0.986 | 0.033 | 0.223 | 0.179 | 0.218 | 0.221 | 0.215 | 0.867 | 0.400 | 0.034 | rs3091244C/A |            |
| rs1800947    | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.030 | 0.011 | 0.035 | 0.005 | 0.131 | 0.029 | 0.138 | 0.135 | 0.147 | 0.034 | 0.066 | 0.862 | rs1417938    |            |
| rs1417938    | 0.975 | 0.954 | 1.000 | 0.992 | 0.929 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.000 | 0.990 | 0.033 | 0.224 | 0.180 | 0.219 | 0.221 | 0.217 | 0.863 | 0.402 | 0.035 | rs3091244C/T |            |
| rs3091244C/A | 0.092 | 0.960 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.984 | 1.000 | 1.000 | 1.000 | X     | X     | 1.000 | 0.088 | 0.073 | 0.090 | 0.090 | 0.084 | 0.004 | 0.036 | 0.013 | rs3093059    |            |
| rs3091244C/T | 0.972 | 0.955 | 1.000 | 0.992 | 0.931 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 1.000 | 1.000 | X     | X     | 0.000 | 0.267 | 0.215 | 0.263 | 0.266 | 0.257 | 0.845 | 0.369 | 0.039        | rs17860488 |
| rs3093059    | 0.853 | 0.964 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.990 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.037 | 0.038 | 0.038 | 0.035 | 0.038 | 0.066 | 0.006 | 0.206 | rs2088632    |            |
| rs17860488   | 0.989 | 0.979 | 0.522 | 0.952 | 1.000 | 0.994 | 1.000 | 1.000 | 0.994 | 1.000 | 0.978 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.206 | 0.991 | 0.994 | 0.954 | 0.128 | 0.033 | 0.151 | rs2088632    |            |
| rs2088632    | 0.931 | 0.710 | 0.025 | 0.553 | 0.959 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.204 | 0.204 | 0.190 | 0.200 | 0.430 | 0.032 | rs2027471    |            |
| rs2027471    | 0.990 | 0.970 | 0.531 | 0.953 | 1.000 | 0.997 | 0.997 | 0.997 | 0.950 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.989 | X     | 0.997 | 0.956 | 0.132 | 0.032 | 0.151 | rs1341665    |            |
| rs1341665    | 0.990 | 0.970 | 0.526 | 0.950 | 1.000 | 0.994 | 0.997 | 0.997 | 0.950 | 0.994 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.950 | 0.125 | 0.030 | 0.148 | rs7443007    |            |
| rs7443007    | 0.990 | 0.990 | 0.539 | 0.997 | 1.000 | 0.997 | 0.997 | 0.997 | 0.946 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.989 | 0.994 | 0.991 | X     | 0.119 | 0.028 | 0.156 | rs12728740   |            |
| rs12728740   | 0.953 | 0.940 | 1.000 | 0.748 | 0.910 | 0.703 | 0.696 | 1.000 | 0.696 | 1.000 | 0.993 | 1.000 | 0.990 | 1.000 | 0.989 | 1.000 | 0.714 | 0.990 | 0.734 | 0.707 | 0.696 | X     | 0.462 | 0.039        | rs10437339 |
| rs10437339   | 0.931 | 0.927 | 0.979 | 0.057 | 0.936 | 0.249 | 0.247 | 0.876 | 0.876 | 0.246 | 0.994 | 1.000 | 0.994 | 0.906 | 0.986 | 0.906 | 0.259 | 0.995 | 0.256 | 0.253 | 0.248 | 1.000 | X     | 0.075        | rs11265263 |
| rs11265263   | 1.000 | 1.000 | 1.000 | 0.980 | 1.000 | 0.982 | 0.982 | 0.876 | 0.876 | 0.983 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.989 | 1.000 | 0.982 | 0.982 | 0.982 | 1.000 | 1.000 | X            |            |

\*  $r^2$  values presented at the top right and D' values at the bottom left.

In comparison with the Hapmap CEPH panel, the dense array of SNPs I have typed, in this large cohort, has determined that strong linkage disequilibrium extends 6.7 to 7.7 kb further downstream than previously estimated. SNP *rs11265257* incorporates into the main *CRP* block and delineates a seventh haplotype. A proposed phylogenetic relationship between these haplotypes is presented in **Figure 3.2**.

**Figure 3.2.** Cladogram of *CRP* haplotypes



Haplotypes are represented by circles corresponding to their population frequency. SNPs differing between haplotypes are listed next to the connecting lines, with the SNPs associated with basal CRP level marked in bold (the line length is arbitrary and does not represent evolutionary distance). Small dark circles on the connecting lines indicate ancestral haplotypes which were not observed but can be inferred to have existed in the history of the population.

### 3.4. – Genetic Associations with basal CRP

#### 3.4.1: Association of CRP SNPs with basal CRP

Each SNP within the *CRP* haplotype block was evaluated for association with basal CRP in a multiple linear regression model incorporating age, sex and SLE affection as covariates. Although my initial analysis suggested SLE had no significant effect on basal CRP (see above) and hence I felt justified in including the full cohort in an analysis, I also considered it prudent to retain SLE affection as a covariate. **Table 3.4** shows the regression parameters for each SNP, analysed assuming alleles act in an additive manner. For triallelic *rs3091244* models were constructed assessing each allele versus the remaining two: the most significant model (C vs. not-C) is shown. Seven polymorphisms, *rs7553007*, *rs1341665*, *rs2027471*, *rs17860488*, *rs1205*, *rs2794520*, and *rs2808628* were significantly associated with basal CRP even after applying a partial Bonferroni correction adjusted for correlation between SNPs (16 tests, mean correlation  $r=0.480$ ). For each SNP the rare allele was associated with lower basal CRP. Weak effects were also seen for *rs11265257* and *rs1800947*. I checked for departure from an additive model for each of the *CRP* markers, but none was found.

The *CRP* SNP association was re-analysed using the orthogonal model of qTDT. **Table 3.4** also shows the significance of the associations determined using this method. Each of the seven key polymorphisms associated in the regression model were also significantly associated using the orthogonal test. It was noted that for the majority of SNPs the results obtained by linear regression and by TDT were in agreement. For SNPs *rs1417938* and *rs1130864*, however, there was a minor discrepancy, with no evidence of CRP association seen by regression, but a result approaching significance seen with the orthogonal model of

qTDT. It is likely that this observation was a “chance” observation due to the differing methodologies used by the two tests. In particular the requirement of the TDT based tests to have informative data on parental genotype transmission means that the final set of families incorporated into the analysis may differ from the complete dataset employed in the regression based test. In any case it should be noted that the results from either methodology do not approach significance when the appropriate significance threshold is considered with regards multiple testing ( $P < 0.012$ ).

**Table 3.4.** Single SNP associations with CRP

| Marker                        | $\beta$ -coefficient | standard error | t      | Regression*<br><i>P</i> -value | Orthogonal model<br><i>P</i> -value |
|-------------------------------|----------------------|----------------|--------|--------------------------------|-------------------------------------|
| <i>rs7553007</i>              | -0.210               | 0.067          | -3.110 | 0.002                          | 0.004                               |
| <i>rs1341665</i>              | -0.191               | 0.066          | -2.898 | 0.004                          | 0.002                               |
| <i>rs2027471</i>              | -0.165               | 0.066          | -2.506 | 0.012                          | 0.005                               |
| <i>rs2808632</i>              | 0.090                | 0.067          | 1.337  | 0.182                          | 0.225                               |
| <i>rs17860488</i>             | -0.189               | 0.068          | -2.766 | 0.006                          | 0.011                               |
| <i>rs3093059</i>              | 0.053                | 0.122          | 0.433  | 0.665                          | 0.419                               |
| <i>rs3091244</i> <sup>†</sup> | 0.082                | 0.066          | 1.240  | 0.215                          | 0.131                               |
| <i>rs1417938</i>              | 0.075                | 0.074          | 1.014  | 0.311                          | 0.035                               |
| <i>rs1800947</i>              | -0.233               | 0.120          | -1.952 | 0.051                          | 0.048                               |
| <i>rs1130864</i>              | 0.084                | 0.073          | 1.145  | 0.252                          | 0.044                               |
| <i>rs1205</i>                 | -0.186               | 0.067          | -2.755 | 0.006                          | 0.002                               |
| <i>rs3093077</i>              | 0.043                | 0.119          | 0.358  | 0.721                          | 0.483                               |
| <i>rs2794520</i>              | -0.181               | 0.068          | -2.668 | 0.008                          | 0.003                               |
| <i>rs2808628</i>              | -0.183               | 0.067          | -2.727 | 0.007                          | 0.004                               |
| <i>rs976538</i>               | 0.020                | 0.080          | 0.247  | 0.805                          | 0.983                               |
| <i>rs11265257</i>             | -0.152               | 0.065          | -2.330 | 0.020                          | 0.021                               |

Covariate (age, sex and SLE affection) adjusted regression parameters for each *CRP* SNP. Significance determined by the orthogonal model of qTDT shown on the right.

\*Unadjusted *P*-values. An individual *P*-value  $< 0.012$  is required to achieve overall significance at the 0.05 level after applying a Bonferroni correction adjusted for correlation between SNPs (these *P*-values are highlighted in bold).

<sup>†</sup>Modelled C vs. not-C alleles.

### 3.4.2: *Trans-acting Genetic Effects*

A preliminary analysis was performed to determine which, if any, of the *trans*-acting SNPs were important to include as covariates in the primary model. This determined that *IL1B* SNPs *rs7596461* was associated with basal CRP (P=0.012). The rare allele of *rs7596461* was associated with lower CRP levels and. One *IL6* SNP, *rs1800796* was also associated with basal CRP, although this effect was restricted to the SLE probands (*rs1800796*\*SLE P=0.18). Both *rs7596461* and *rs1800796*\*SLE were therefore also included as a covariate term in the CRP regression models presented below. For illustrative purposes the effect of these *trans*-acting markers is shown in the *CRP* haplotype model (**table 3.7**). Genotype at *CRP*, *IL1B* and *IL6* had an additive effect on basal CRP, with no evidence for higher order interaction. I did not find an association between *APOE* SNPs and basal CRP, either individually or as a haplotype (as would determine Apo ε serotype).

### 3.4.3: *Mixed genetic models*

The association of *CRP* SNPs with basal CRP was then reassessed in the multiple linear regression model, now incorporating the key *trans*-acting genetic effects (*IL1B* SNP *rs7596461* and *IL6* SNP *rs1800796*) as genetic covariate. This additional covariate model is presented in **table 3.5**. The level of significance attached to each of the *CRP* associations increased marginally with the addition of this additional genetic data.

**Table 3.5.** Single SNP associations with serum CRP with genetic covariates

| Marker                        | $\beta$ -coefficient | standard error | t      | Regression*<br>P-value |
|-------------------------------|----------------------|----------------|--------|------------------------|
| <i>rs7553007</i>              | -0.230               | 0.068          | -3.365 | 0.001                  |
| <i>rs1341665</i>              | -0.237               | 0.068          | -3.462 | 0.001                  |
| <i>rs2027471</i>              | -0.197               | 0.068          | -2.911 | 0.004                  |
| <i>rs2808632</i>              | 0.107                | 0.069          | 1.542  | 0.123                  |
| <i>rs17860488</i>             | -0.231               | 0.070          | -3.301 | 0.001                  |
| <i>rs3093059</i>              | 0.106                | 0.124          | 0.850  | 0.396                  |
| <i>rs3091244</i> <sup>†</sup> | 0.103                | 0.067          | 1.535  | 0.125                  |
| <i>rs1417938</i>              | 0.077                | 0.075          | 1.022  | 0.307                  |
| <i>rs1800947</i>              | -0.230               | 0.122          | -1.885 | 0.060                  |
| <i>rs1130864</i>              | 0.079                | 0.074          | 1.062  | 0.289                  |
| <i>rs1205</i>                 | -0.226               | 0.069          | -3.254 | 0.001                  |
| <i>rs3093077</i>              | 0.096                | 0.123          | 0.784  | 0.434                  |
| <i>rs2794520</i>              | -0.220               | 0.070          | -3.159 | 0.002                  |
| <i>rs2808628</i>              | -0.222               | 0.069          | -3.209 | 0.001                  |
| <i>rs976538</i>               | 0.011                | 0.081          | 0.140  | 0.889                  |
| <i>rs11265257</i>             | -0.153               | 0.066          | -2.306 | 0.021                  |

Regression parameters for each *CRP* SNP adjusted for phenotypic covariates (age, sex and SLE affection) and trans-acting genetic covariates (genotype at *IL1B* SNP rs7596461 and *IL6* SNP rs1800796). \*Unadjusted *P*-values. An individual *P*-value <0.012 is required to achieve overall significance at the 0.05 level after applying a Bonferroni correction adjusted for correlation between SNPs (these *P*-values are highlighted in bold).

<sup>†</sup>Modelled C vs. not-C alleles.

#### 3.4.4: Influence of SLE affection on CRP associations

Because it has been reported that the magnitude of the acute-phase rise in CRP (although not necessarily basal CRP levels) may be blunted in patients with SLE, I performed a parallel analysis in a subset of the cohort (n=558) consisting of all the unrelated, unaffected founders on whom serum CRP was available (i.e. all the parents and, if no parental data was available, an unaffected sibling selected at random). This avoided the need to adjust for familial clustering, while restricting analysis to individuals without SLE. The results of this analysis are presented in **table 3.6**.

**Table 3.6.** Single SNP associations in the unaffected, unrelated founders

| Marker            | $\beta$ -coefficient | standard error | t      | P-value |
|-------------------|----------------------|----------------|--------|---------|
| <i>rs7553007</i>  | -0.179               | 0.061          | -2.945 | 0.003   |
| <i>rs1341665</i>  | -0.179               | 0.060          | -3.013 | 0.003   |
| <i>rs2027471</i>  | -0.190               | 0.060          | -3.191 | 0.002   |
| <i>rs2808632</i>  | 0.022                | 0.064          | -0.348 | 0.728   |
| <i>rs17860488</i> | -0.181               | 0.060          | -3.040 | 0.003   |
| <i>rs3093059</i>  | 0.276                | 0.107          | 2.589  | 0.010   |
| <i>rs3091244</i>  | 0.165                | 0.055          | 3.006  | 0.003   |
| <i>rs1417938</i>  | 0.109                | 0.061          | 1.787  | 0.075   |
| <i>rs1800947</i>  | -0.295               | 0.107          | -2.766 | 0.006   |
| <i>rs1130864</i>  | 0.116                | 0.060          | 1.927  | 0.055   |
| <i>rs1205</i>     | -0.172               | 0.060          | -2.843 | 0.005   |
| <i>rs3093077</i>  | 0.243                | 0.107          | 2.273  | 0.024   |
| <i>rs2794520</i>  | -0.177               | 0.061          | -2.908 | 0.004   |
| <i>rs2808628</i>  | -0.186               | 0.061          | -3.062 | 0.002   |
| <i>rs976538</i>   | 0.097                | 0.072          | 1.339  | 0.181   |
| <i>rs11265257</i> | -0.169               | 0.060          | 2.830  | 0.005   |

The same seven SNPs were significant in this subset analysis but, in addition, significant associations were observed with *rs11265257* ( $P=0.005$ ), *rs1800947* ( $P=0.006$ ), *rs3091244* (C vs. not C) ( $P=0.003$ ) and *rs3093059* ( $P=0.010$ ). SNPs *rs11265257* and *rs1800947* were both weakly associated in the full cohort and the stronger association in this subset is likely due to the increased power attained by requiring no adjustment for familial clustering. For *rs3093059* and *rs3091244* the increase in significance appeared greater, and the estimated magnitude of effect was larger in this subset ( $\beta = 0.276$  and  $0.165$  respectively). I specifically looked for *rs3093059*\*SLE and *rs3091244*\*SLE interactions in the full cohort but none was observed ( $P=0.581$  and  $P=0.116$  respectively). Again it is possible that the observed association may be due to increased power, rather than differential genetic regulation of basal CRP between people with and without SLE.

### 3.4.5: Haplotypic Association with basal CRP

CRP haplotypes were evaluated with reference to commonest haplotype H1 in a multiple linear regression model and these results are presented in **table 3.7**. A pairwise comparison between haplotypes is presented in **table 3.8**.

**Table 3.7.** The association of CRP with CRP haplotypes

| Parameter               | $\beta$ -coefficient | standard error | t      | P-value |
|-------------------------|----------------------|----------------|--------|---------|
| rs7596461 (IL1B)        | -0.311               | 0.123          | -2.535 | 0.012   |
| rs1800796*SLE (IL6)     | 0.844                | 0.252          | 3.342  | 0.001   |
| rs1800796*Healthy (IL6) | 0.195                | 0.158          | 1.232  | 0.218   |
| H7                      | 0.206                | 0.210          | 0.983  | 0.326   |
| H6                      | 0.138                | 0.159          | 0.868  | 0.386   |
| H4                      | 0.000                | 0.139          | -0.002 | 0.998   |
| H1                      | Ref                  |                |        |         |
| H3                      | -0.100               | 0.099          | -1.007 | 0.314   |
| H2                      | -0.235               | 0.093          | -2.533 | 0.012   |
| H5                      | -0.305               | 0.135          | -2.270 | 0.024   |
| Hx                      | 0.121                | 0.247          | 0.490  | 0.624   |

Haplotypes modelled with reference to haplotype H1. Hx represents pooled haplotypes with frequency <1%.

**Table 3.8.** Pairwise comparison of CRP levels by CRP haplotype.

| Haplotype | P-value |       |       |       |       |       |    |
|-----------|---------|-------|-------|-------|-------|-------|----|
| H7        | X       |       |       |       |       |       |    |
| H6        | 0.788   | X     |       |       |       |       |    |
| H4        | 0.384   | 0.455 | X     |       |       |       |    |
| H1        | 0.326   | 0.386 | 0.998 | X     |       |       |    |
| H3        | 0.155   | 0.158 | 0.480 | 0.314 | X     |       |    |
| H2        | 0.035   | 0.023 | 0.085 | 0.012 | 0.167 | X     |    |
| H5        | 0.026   | 0.020 | 0.072 | 0.024 | 0.137 | 0.602 | X  |
| Haplotype | H7      | H6    | H4    | H1    | H3    | H2    | H5 |

Haplotypes H2 and H5 appeared to form a group associated with lower CRP levels. Both carry the rare allele of all the strongly associated SNPs which are in tight linkage disequilibrium (mean pairwise  $r^2=0.969$ ). By reference to **figure 3.2** we can see that H2 and H5 form a clade which is distinguished from the remaining haplotypes by eight of the typed SNPs. Seven of these are associated with CRP expression, while the remainder, *rs2808632*, also occurs at another position in the tree, explaining the lack of association with this SNP. The CRP levels associated with the H2/H5 clade are significantly lower than that associated with all other haplotypes combined ( $\beta$  -0.240;  $t$  -3.417;  $P=0.001$ ). Using the regression parameters it could therefore be predicted that a healthy woman, aged 30 and homozygous for the common alleles of *rs7596461* and *rs1800796*, would have a CRP of 2.99mg/L if she was homozygous for H2/H5 compared with 4.37mg/L if she carried no H2/H5 haplotypes. Haplotype analysis in unaffected founders only showed a similar pattern of association (**tables 3.9** and **3.10**), including the association of the H2/H5 clade with lower basal CRP ( $\beta$  -0.183; S.E. 0.061;  $t$  -3.016;  $P=0.003$ ).

**Table 3.9.** The association of CRP with *CRP* haplotypes (unaffected founders)

| Parameter | $\beta$ -coefficient | Standard error | t      | P-value |
|-----------|----------------------|----------------|--------|---------|
| H7        | 0.228                | 0.177          | 1.288  | 0.198   |
| H4        | 0.175                | 0.119          | 1.475  | 0.141   |
| H1        | Ref                  |                |        |         |
| H6        | -0.058               | 0.140          | -0.418 | 0.676   |
| H3        | -0.128               | 0.081          | -1.566 | 0.118   |
| H2        | -0.138               | 0.077          | -1.785 | 0.075   |
| H5        | -0.300               | 0.114          | -2.638 | 0.009   |
| Hx        | 0.124                | 0.212          | 0.583  | 0.560   |

*CRP* haplotypes modelled with reference to haplotype H1. Hx represents pooled haplotypes with frequency <1%.

**Table 3.10.** Pairwise comparison of CRP levels with *CRP* haplotype (unaffected founders).

| Haplotype | <i>P</i> -value |       |       |       |       |       |    |
|-----------|-----------------|-------|-------|-------|-------|-------|----|
| H7        | X               |       |       |       |       |       |    |
| H4        | 0.788           | X     |       |       |       |       |    |
| H1        | 0.198           | 0.141 | X     |       |       |       |    |
| H6        | 0.184           | 0.158 | 0.676 | X     |       |       |    |
| H3        | 0.52            | 0.014 | 0.118 | 0.643 | X     |       |    |
| H2        | 0.044           | 0.007 | 0.075 | 0.576 | 0.906 | X     |    |
| H5        | 0.009           | 0.001 | 0.009 | 0.137 | 0.170 | 0.182 | X  |
| Haplotype | H7              | H4    | H1    | H6    | H3    | H2    | H5 |

### 3.5: Genetic contribution to total CRP variability

Using the variance components framework within qTDT we can provide a measure of SNP effect on population CRP variability by estimating the polygenic component of basal CRP variation before and after addition of genotype. Of the seven associated *CRP* polymorphisms the proportion of CRP variance attributable to that SNP ranged from 0.51% (*rs976538*) to 2.54% (*rs2808628*). Basal CRP variance explained by *CRP* haplotype as a whole was 3.77%. Given the estimated 28% CRP heritability, *CRP* haplotype therefore accounted for approximately 15% of the total genetic effect. SNPs *rs7596461* and *rs1800796* both have low rare allele frequency and hence a negligible effect on total basal CRP variation (<0.5%). In an analysis of healthy family members only we estimated *CRP* haplotype to account for 5.51% of basal CRP variance. Given the lower CRP heritability estimate for this healthy cohort *CRP* haplotype therefore accounted for approximately 26% of the total genetic effect.

### 3.6: CRP Genetic Association with SLE

The association between SLE and *CRP* SNPs or haplotypes was assessed by transmission disequilibrium testing using TRANSMIT. No difference in allele or haplotype frequency was observed between probands and their relatives and hence no association with SLE was identified in this cohort (“best” haplotype association:  $P=0.22$  for haplotype H4).

### 3.7: CRP Genetics in UK SLE Families – Discussion

#### 3.7.1: Summary of Findings

The data presented in this chapter confirm the genetic basis of basal CRP expression by demonstrating 28.7% heritability. Seven genotyped *CRP* polymorphisms were found to be associated with lower basal CRP. These span 30kb of *CRP* and lie upstream (*rs7553007*, *rs1341665*, *rs2027471*, *rs17860488*), downstream (*rs2794520*, *rs2808628*) and within the 3'-UTR (*rs1205*). I have clarified the haplotype structure of *CRP* and defined the limits of the haplotype block in this population. The rare alleles of all the consistently associated polymorphisms are carried solely on haplotypes H2 and H5 which are themselves associated with significantly lower basal CRP. These haplotypes form a distinct phylogenetic clade, thus strengthening the argument that a major functional SNP lies on this branch of the haplotype tree. My use of a variance components approach to estimate the magnitude of effect of associated SNPs on basal CRP variation is novel. Despite confirming that variation at *CRP* is associated with basal CRP, the total effect appears rather small, explaining only 3.77% of basal CRP variability. I was unable to demonstrate any association between *CRP* variation and SLE.

### 3.7.2: A Comparison with Existing Studies – *cis*-effects

Of the SNPs associated with basal CRP level in this study, only *rs1205* has also been associated with CRP in existing published studies (**table 1.2**) although, as discussed in section 1.3.3, it has not been associated in every study where it has been tested. Those studies which do show an association show the rare allele to be associated with lower CRP expression, as shown here. Studies which analyse *CRP* haplotypes (**table 1.3**) also generally confirm that haplotypes containing the rare allele of *rs1205* are associated with low basal CRP. I have therefore extended this finding to a group of linked SNPs delineating “low-producing” haplotypes. Recently published work suggests H2 may be further broken down by SNPs *rs12029262* and *rs3116654*, although these additional haplotypes are also associated with low basal CRP and may have little additional functional consequence (Kathiresan *et al* 2006). There is strong linkage disequilibrium between all of the associated SNPs in my study, making it difficult to determine a likely functional candidate on the basis of this genetic study alone. A detailed discussion of potential functional variants at *CRP* in the light of all the studies presented in this thesis can be found in section 8.4.5.

The variance components estimate of the proportion of basal CRP variability accounted for by the typed genetic variants is of particular interest as this approach has not been used before, although other authors have made a similar sort of estimate in regression-based analyses using the partial  $r^2$  for associated *CRP* polymorphisms (see **table 8.2**). Both my results, and existing data, suggest that the effect of these typed *cis*-acting variants accounts for only a small proportion of the genetic component to CRP variability. In this study *CRP* haplotypes accounted for 3.77% of basal CRP variability. Given the estimated 28% CRP heritability this therefore represented only 15% of the total genetic effect.

### 3.7.3: A Comparison with Existing Studies – *trans-effects*

I identified associations between *IL1B* SNP *rs7596461* and *IL6* SNP *rs1800796* and basal CRP, although the latter was only significant in the SLE probands. Individually these SNPs have previously been associated with basal CRP, but they have not been included as covariates in large *CRP* haplotype analyses (Berger *et al* 2002; Eklund *et al* 2005; Latkovskis *et al* 2004; Vickers *et al* 2002). Perhaps the key point about these associations is that even within a so-called basal range, CRP expression may be influenced by classical inflammatory stimuli. I chose to confine the analysis to individuals with serum CRP <10mg/L, primarily to avoid the complicating factor of including patients with active SLE. Some other authors have taken a different approach and analysed complete cohorts including individuals with higher CRP. A more detailed discussion of the boundaries between basal and acute-phase CRP, and how this may affect the interpretation of CRP expression studies may be found in section 8.3.1. At this point it suffices note that the inclusion of *IL1B* or *IL6* genotyping could have been relevant in many of the large existing *CRP* studies, particularly, given the proposed inflammatory aetiology of atherosclerosis (section 1.3.5), those using CHD cohorts.

### 3.7.4: A Comparison with Existing Studies – *SLE association*

In this study I was unable to demonstrate an association between *CRP* genotype and SLE. While other authors have shown an association, it is perhaps notable that the reported effects have been modest and there is inconsistency over the key associated variants (section 1.3.4) (Edberg *et al* 2008b; Jonsen *et al* 2007; Russell *et al* 2004; Shih *et al* 2008). The negative findings were initially surprising since the cohort used in this study overlapped with that used for the study of Russell *et al*, (the difference being 200 individuals from 50 families) where a

positive association was identified. A detailed comparison of genotyping obtained in this study compared with that obtained by Russell (genotyping using RFLP) revealed no systematic errors in genotyping to account for this. A combination of cohort differences (Russell included a small number of non-European families, while a slightly different selection of siblings was genotyped due to plate re-organisations), and a small number of discrepant genotypes accounted for the loss of association. However it was clear that even with a cohort of the size presented in this study the power to detect a disease association was limited. As outlined in section 3.1, given the magnitude of effect reported by Russell *et al* the current study had only an estimated 45% power to detect an effect at the  $P=0.05$  level.

### *3.7.5: Key Study Features and Potential Limitations*

The primary aim of the work presented in this chapter was to fine-map the *CRP* locus, to enable me to draw together existing published work, and to re-evaluate the associations between *CRP* genetic variants, serum CRP and SLE. This is the first large scale *CRP* study to utilise a family-based approach.

It is important, nonetheless to consider potential sources of error in this study. In any study, measurement error will occur, resulting in an analysis of data values that deviate from the “true” values for the individuals in the study. It is important to appreciate that measurement error can be divided into two components; systematic error and random error. Systematic errors are constant, theoretically predictable, factors leading to a systematic deviation in the mean of the measured values away from the true values. Random errors, on the other hand are due to unpredictable “random” factors influencing each measurement in a different way. Since these errors result in a measured value being randomly either higher or lower than the

true value they will increase the variance of the measured values in relation to the true values but will not influence the mean. The critical difference between the two sorts of measurement error is the effect they may have on the interpretation of study results. Systematic errors may result in the generation of false associations. On the other hand random errors may make it harder to demonstrate a true association (as it introduces so much “noise” into the data), however, it should not affect the validity of a significant association, should one still be demonstrated.

It should be acknowledged that the work presented in this chapter forms a *post-hoc* analysis of CRP expression in a cohort collected for the purpose of the genetic analysis of SLE disease expression. The analysed cohort is therefore a mixture of related individuals, some of whom are healthy (although presumably enriched for SLE susceptibility genes) and some of whom have SLE. At first glance attempting to unpick the association between genetic variation and serum CRP in the context of SLE presents an insurmountable problem. I would argue, however, that this is not the case.

Firstly, SLE itself is known to have a complex relationship with CRP expression (see section 1.3.4) so the validity of including affected individuals in an analysis needs to be justified. In particular if SLE is exerting a systematic effect on serum CRP then this may introduce a systematic error, which would become important if alleles influencing CRP expression were also segregating in individuals with disease. Some existing studies suggest, within the basal CRP range, the presence of SLE does not exert much influence on serum CRP; indeed previous analyses have suggested that most individuals with SLE have basal levels of CRP, with a preserved relationship between basal CRP and cardiovascular risk (Szalai *et al* 2005a; Toloza *et al* 2004). Although there is only limited published data it would appear that the

dysregulation of CRP in SLE is related to a failure to mount an adequate acute-phase response (Pereira Da Silva *et al* 1980). Since the normal correlation between levels of CRP and IL-6 appears to be lost in patients with SLE, the block in maximal expression of acute-phase CRP would seem to be a fairly late stage event in the cytokine pathways culminating in CRP expression (Williams, Jr. *et al* 2005). These studies provide evidence that there is a problem mounting an acute-phase response in SLE, but in an analysis confined to individuals with basal level CRP, SLE may not exert a significant influence. Nonetheless it was clearly necessary to control for potential systematic error through the consistent inclusion of SLE affection as a covariate in all the regression models, and my findings backed up the argument above, with no significant SLE effect observed when the cohort was restricted to individuals with serum CRP <10mg/L. I am therefore confident that the potential systematic bias introduced by the inclusion of individuals with SLE was, firstly, minimal and, secondly, controlled for in my analyses. I was able to add further weight to this argument, through additional analyses presented in section 3.4.3, where all individuals with SLE were excluded from analysis. If the role of *CRP* polymorphism in determining basal CRP is different in people with and without SLE then the effect is clearly very small.

A related issue that could be raised regarding the use of patients with SLE is whether any adjustment was made for the level of disease activity. As I stated in section 2.1.1 the disease phenotype of the patients varied, but data collection took place in an outpatient setting from patients who were not “acutely unwell”. Although, broadly speaking, these were patients not in an active stage of their disease, there remains the potential for differences in overall disease activity (and hence potential differences in the effect of SLE on these patients CRP levels). This variation however falls within the “random error” category, because the data collection was performed in a cross-sectional study design in which venepuncture was not

performed at any particular time-point in relation to disease activity. We would therefore expect the “error” associated with the effect of disease activity on CRP levels across the cohort, over and above that due to genetic variation, to be random.

A similar, but slightly different issue concerns the possibility of intra-individual variation in CRP concentration, irrespective of the presence or absence of disease. Clearly the greater the natural intra-individual fluctuations in CRP concentration with time, the greater the random measurement error associated with a single “snap-shot” CRP measurement as collected for this study, and hence the lower the power to detect a genetic association. The only available published data suggests remarkably consistent intra-individual serum CRP levels over time, nonetheless it should be recognised that an analysis of multiple CRP readings collected at different time-points is likely to significantly improve the prospects of identifying genetic influences on serum CRP levels (Danesh *et al* 1998; Kivimaki *et al* 2007; Ridker *et al* 1999).

It should also be recognised that the CRP assay itself may be subject to measurement error, although I am confident that the amount of error, whether random or systematic will be insignificant. The immunoturbimetric assay itself is well validated and, in the setting of a clinical service, was subject to stringent local quality control. CRP itself is a robust molecule with limited degradation occurring either due to room temperature storage (as occurs when collected whole blood is posted to the laboratory for processing) or as a result of freeze-thaw cycles (Aziz *et al* 2003).

A further issue concerns the potential for anti-CRP antibodies in SLE. Although not widely reproduced or validated, the presence of anti-CRP antibodies has been reported in 36-78% with SLE; most notably in patients with active disease, who are also strongly positive for

other antibodies (e.g. anti-dsDNA) which vary in parallel with disease activity (Bell *et al* 1998; Sjowall *et al* 2002; Sjowall *et al* 2005). Current data suggests, however, that these antibodies are directed against monomeric subunits of CRP, and do not react with the intact pentamer (Sjowall & Wettero 2007). Furthermore no correlation between anti-CRP reactivity and CRP levels has been observed (Figueredo *et al* 2006). It would seem unlikely either that anti-CRP antibodies would interfere with the immunoturbimetric assay or that there would be any systematic link between the presence of anti-CRP antibodies and alterations in serum CRP, other than those due to the presence of SLE itself. At most, the presence of anti-CRP antibodies, if having any effect, will simply add to the random error associated with SLE measurement.

The use of a family-based cohort also presents the potential for systematic study bias if analysed using standard population based techniques, since CRP levels are likely to cluster within families of genetically related individuals. Numerous ways have been developed to overcome this problem (as discussed in section 2.5.5.) including, as I have done in this analysis, the incorporation of a random effects parameter into the regression model. Provided the data is appropriately analysed, a family-based design offers advantages in terms of the robust control for biases such as population stratification, thus adding particular weight to my findings. The family-based approach also offers an advantage in allowing the checking of assigned genotypes for Mendelian consistency, thus offering an additional level of quality control. I am therefore confident that genotyping which passed the quality control criteria was of high quality.

By considering each potential problem in terms of whether it is introducing systematic error (which is then adjusted for) or random error (which cannot be controlled for but should not

alter the statistical validity of the study), I believe the association analyses I have undertaken can be considered statistically robust. An understanding of measurement error is important for the interpretation of the data in each of the chapters in this thesis, and I will refer back to this section in later chapters.

Finally, it should be mentioned that one area in which random error can result in erroneous results is in the estimate of heritability (Fisher 1951). We have already defined heritability in section 1.5.3 as the proportion of variance in phenotype that can be attributed to variance in genotype or, expressed another way:

Heritability = Variance due to genetics/Total variance.

Excessive total variance in the trait of interest introduced through measurement error (rather than genuine environmental causes of variation) will therefore reduce the overall estimate of heritability. With this in mind it is likely that we underestimate CRP heritability.

### *3.7.6: Conclusions*

The data presented in this chapter confirm that genetic variation at *CRP* influences serum basal CRP levels, although it is likely that the magnitude of this effect is relatively small. It remains, however, uncertain which variant or variants are likely to play a direct functional role. The studies presented in chapters 4 and 5 use genetic analyses to attempt to answer this important question, firstly through the study of diverse ethnic populations, and secondly, in chapter 5, through the use of a model choice approach which considers the possibility of more than one functional effect at this locus.

## **CHAPTER 4**

# ***CRP* ASSOCIATIONS IN FILIPINO SLE FAMILIES**

#### 4.1: *CRP* Associations in Filipino SLE Families – Study Aims

The Filipino population has been poorly studied genetically. Population based studies suggest a reasonably close genetic relationship with the Han Chinese, but few studies have fine mapped individual loci, or looked for disease associations (Nei & Roychoudhury 1993; Tian *et al* 2008b). The aim of the work presented in this chapter was to delineate the *CRP* genetic architecture in a previously unstudied cohort of Filipino SLE families, encompassing a similar genetic interval to the European data presented in chapter 3. I aimed to evaluate the association between *CRP* variants and serum CRP expression, although the limited number of larger families in this dataset meant the analysis was, by necessity, more restricted. Once again I also aimed to consider potential effects arising from *trans*-acting genetic variants. The overall expectation was that the Filipino genetic architecture may allow me to refine associations found in the white European populations.

Although this cohort is being collected with the purpose of analysing genetic associations with SLE, the number of complete, informative families, at the time this study was performed was insufficient to have the power for a meaningful analysis of the association between *CRP* and SLE (less than half the number of families as the European dataset, which itself only had a 45% power to detect a disease association). The power to detect an association with serum CRP is also likely to be low, although difficult to formalise as explained in section 3.1. Nonetheless, given a fairly strong association between CRP variation and serum CRP was observed in the European cohort, I was keen to pursue this additional study. The data presented in this chapter may be found in published form (Rhodes *et al* 2008c).

## 4.2: Filipino SLE Cohort – Demographics and CRP distribution

As I have already outlined in section 2.1.2. the cohort employed for the analyses presented in this chapter consisted of 328 ethnic Filipino families. Each family consisted as a minimum of one individual with SLE. The availability of data from unaffected relatives varied so that the overall family structure was complex. The mean age was 38.2 years, 79% were female and 47% had SLE. The median CRP was 1.88mg/L (range 0.19 to 203.05).

Therefore the final basal CRP cohort contained 509 individuals with a median CRP of 1.50mg/L. The mean age was 38.2 years, 79% were female and 47% had SLE. In an initial regression model I evaluated the effect of age, sex and SLE affection on basal CRP, while incorporating a random effects factor to allow for clustering of CRP levels within families (**table 4.1**). Basal CRP increased with age and was higher in women and patients with SLE. These three factors were therefore included as covariates in all subsequent regression models. Because it yet to be determined whether basal and acute-phase CRP expression are regulated by the same genetic factors I excluded all individuals with a CRP>10mg/L from further analysis.

**Table 4.1.** Effect of age, sex and SLE affection on CRP

| Variable       | $\beta$ -coefficient | standard error | t     | P-value |
|----------------|----------------------|----------------|-------|---------|
| Age            | 0.015                | 0.006          | 3.226 | 0.006   |
| Sex (Female)   | 0.334                | 0.145          | 2.295 | 0.023   |
| SLE (Affected) | 0.286                | 0.139          | 2.059 | 0.041   |

### 4.3: Filipino *CRP* Genetic Architecture

Serum CRP and genotype data was available for all 285 Filipino SLE probands and 288 unaffected family members. Nineteen *CRP* SNPs passed quality control checks and were included for analysis. Detailed marker information is presented in **table 4.2**.

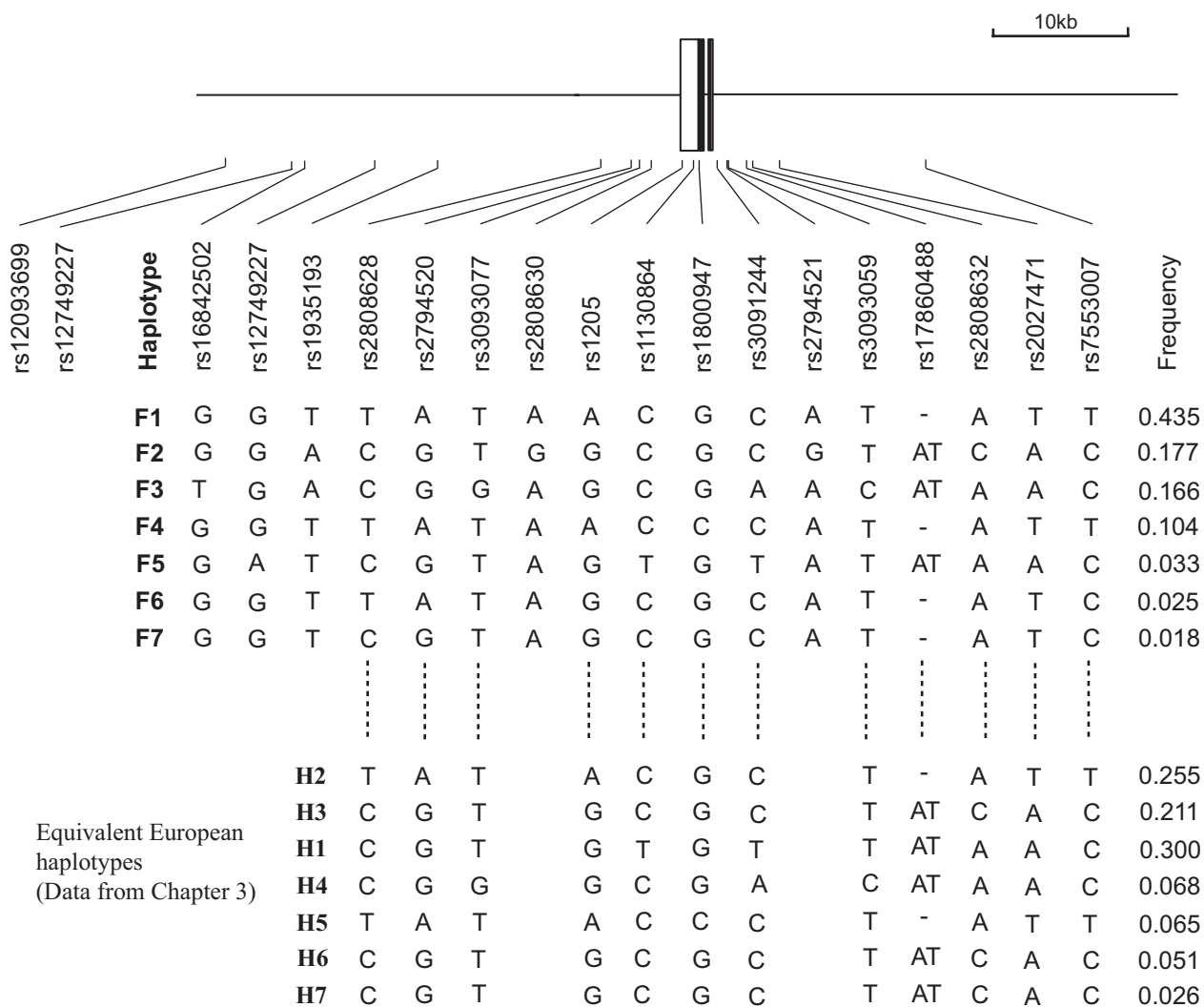
The *CRP* haplotype structure is outlined in **figure 4.1**, with haplotypes labelled F1 to F7 in order of their frequency within this population. Between *rs16842502* and *rs7553007* four haplotypes occur with frequency > 5%. Downstream SNPs *rs2592887* and *rs12093699* were not contained within this haplotype block. **Figure 4.3** demonstrates the pairwise LD between markers (a graphical representation of the LD statistics presented in **table 4.3**.) Unfortunately downstream coverage was somewhat different to that used for the European analysis in chapter 3 (due to Sequenom pooling constraints) and it is therefore difficult to say whether the physical extent of the *CRP* haplotype block is different in this population. **Figure 4.3** shows the most likely phylogenetic relationship between these haplotypes on the basis of maximum parsimony for the mutational steps. A simple clade structure was predicted with no mutations duplicated on different branches of the tree. Haplotype F1 had the highest outgroup weighting probability (0.33) and was therefore proposed as the ancestral haplotype (Castelloe & Templeton 1994). A full pairwise comparison between markers is presented in **table 4.3**.

**Table 4.2** Typed SNPs demonstrating alleles and rare allele frequencies

| rs number                | SNP   | Location            | Rare allele frequency | HWP <sup>a</sup> | missing data (%) |
|--------------------------|-------|---------------------|-----------------------|------------------|------------------|
| <i>CRP</i> SNPs          |       |                     |                       |                  |                  |
| <i>rs7553007</i>         | C/T   | 5' flank            | 0.44                  | 0.76             | 2.2              |
| <i>rs2027471</i>         | A/T   | 5' flank            | 0.41                  | 0.92             | 0.9              |
| <i>rs2808632</i>         | A/C   | 5' flank            | 0.19                  | 0.31             | 1.7              |
| <b><i>rs17860488</i></b> | AT/-  | 5' flank            | 0.40                  | 0.86             | 0.8              |
| <b><i>rs3093059</i>*</b> | T/C   | 5' flank            | 0.17                  | 0.65             | 2.2              |
| <i>rs2794521</i>         | A/G   | 5' flank            | 0.19                  | 0.40             | 2.2              |
| <i>rs3091244</i> *       | C/A/T | upstream promoter   | 0.17(A)/ 0.04(T)      | 0.78             | 0.6              |
| <b><i>rs1800947</i>*</b> | G/C   | exon 2 (synonymous) | 0.12                  | 0.65             | 3.0              |
| <i>rs1130864</i> *       | C/T   | 3'-UTR              | 0.04                  | 0.23             | 7.9              |
| <b><i>rs1205</i>*</b>    | G/A   | 3'-UTR              | 0.46                  | 0.84             | 2.6              |
| <b><i>rs2808630</i></b>  | A/G   | 3' flank            | 0.20                  | 0.41             | 4.0              |
| <b><i>rs3093077</i>*</b> | T/G   | 3' flank            | 0.18                  | 0.76             | 2.2              |
| <i>rs2794520</i>         | G/A   | 3' flank            | 0.43                  | 0.70             | 2.5              |
| <i>rs2808628</i>         | C/T   | 3' flank            | 0.43                  | 0.97             | 2.0              |
| <i>rs1935193</i>         | T/A   | 3' flank            | 0.37                  | 0.40             | 4.5              |
| <i>rs12749227</i>        | G/A   | 3' flank            | 0.04                  | 0.34             | 1.2              |
| <i>rs16842502</i>        | G/T   | 3' flank            | 0.18                  | 0.67             | 3.0              |
| <i>rs2592887</i>         | A/G   | 3' flank            | 0.43                  | 0.80             | 2.8              |
| <i>rs12093699</i>        | C/T   | 3' flank            | 0.04                  | 0.27             | 6.7              |
| <i>IL1B</i> SNPs         |       |                     |                       |                  |                  |
| <i>rs4849124</i>         | C/T   | 3' flank            | 0.06                  | 0.28             | 7.0              |
| <i>rs3917368</i>         | G/A   | 3' flank            | 0.42                  | 0.75             | 4.2              |
| <i>rs1143634</i>         | C/T   | exon 5              | 0.03                  | 0.59             | 2.6              |
| <i>rs16944</i>           | C/T   | 5' flank            | 0.42                  | 0.27             | 3.3              |
| <i>IL6</i> SNPs          |       |                     |                       |                  |                  |
| <i>rs1800796</i>         | G/C   | 5' flank            | 0.33                  | 0.81             | 3.9              |
| <i>rs2069837</i>         | A/G   | intron 2            | 0.25                  | 0.71             | 2.6              |
| <i>rs2069840</i>         | C/G   | intron 3            | 0.07                  | 0.66             | 3.4              |
| <i>APOE</i> SNPs         |       |                     |                       |                  |                  |
| <i>rs429358</i>          | T/C   | non-synonymous      | 0.05                  | 0.54             | 4.7              |
| <i>rs7412</i>            | C/T   | non-synonymous      | 0.11                  | 0.32             | 10.0             |

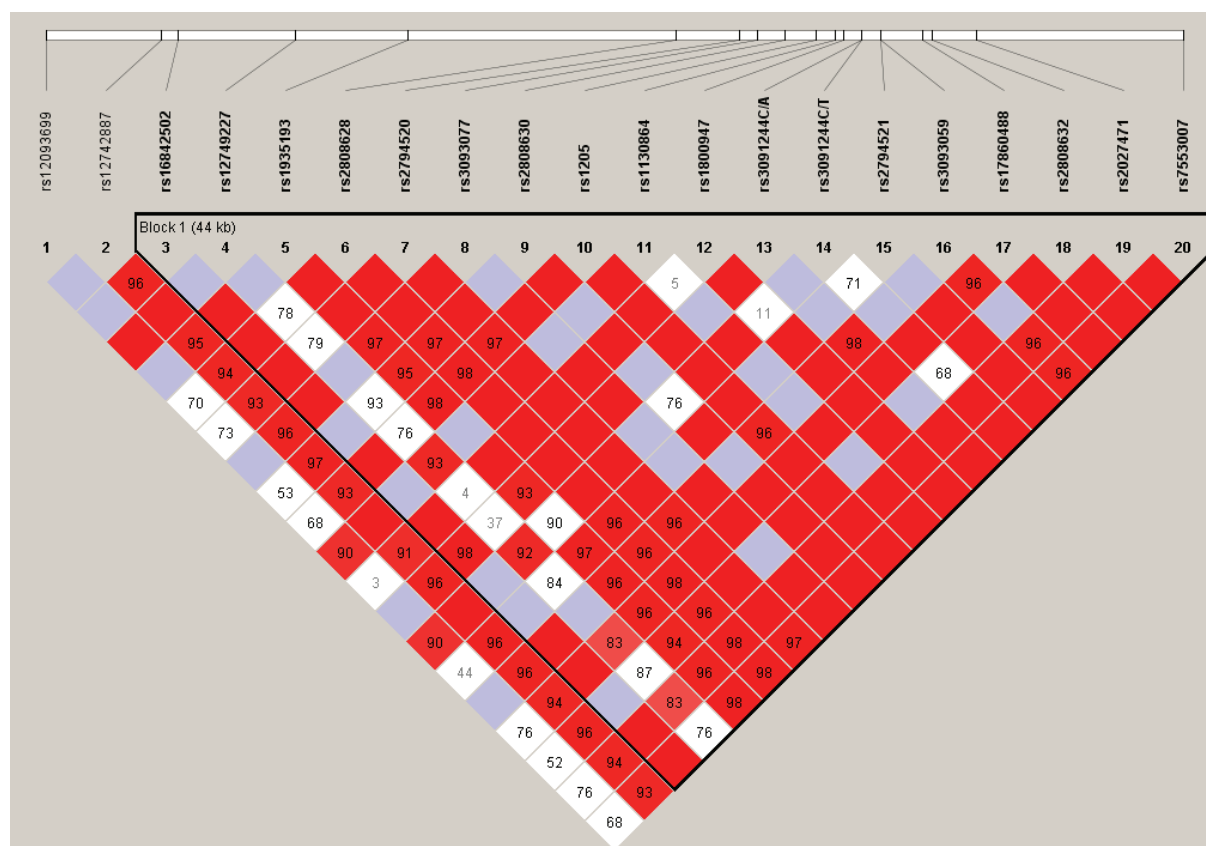
SNPs highlighted in bold have been selected to tag the key *CRP* haplotype block. SNPs marked with a \* have been previously associated with serum CRP levels in any population (see **table 1.2**) <sup>a</sup>Hardy-Weinberg probability *P*-value calculated by  $\chi^2$

**Figure 4.1.** Filipino *CRP* haplotypes.



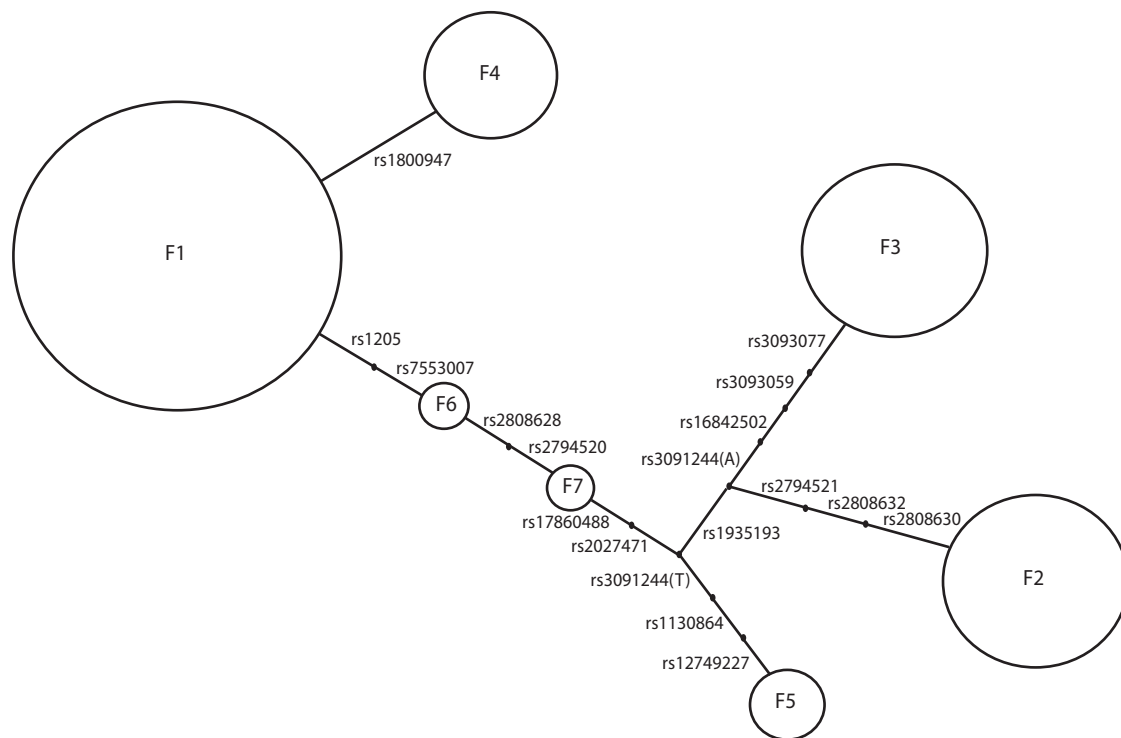
The position of the typed SNPs relative to *CRP* is demonstrated in the top part of the diagram. SNPs rs12093699 and rs12749227 are predicted to lie downstream of the main *CRP* haplotype block. In the lower section of the diagram the equivalent European haplotypes have been illustrated, using the data presented in chapter 3, for SNPs that have been genotyped in common.

**Figure 4.2.** Graphical representation of LD at the Filipino *CRP* locus



This diagram demonstrates the pairwise LD between markers (listed across the top) it is therefore a graphical representation of the data in **table 4.3**. The haplotype block is defined by the default algorithm of Gabriel *et al* (see methods section 2.4.4). The pairwise LD between any two SNPs is indicated by the colour coding of the corresponding square of the LD plot. High LD is indicated by a pink/red colour scheme (the dark red squares are linked with a  $D' = 1$  and a LOD score  $\geq 2$ ). Numbers within the squares correspond to the  $D'$  score, with 96 representing a  $D' = 0.96$  and 44 a  $D' = 0.44$  etc. Where squares have no number the  $D' = 1$

**Figure 4.3.** Cladogram of *CRP* haplotypes.



Haplotype F1 is proposed as the ancestral haplotype. Other haplotypes differ from F1 by the SNPs annotating the connecting lines. Small dots represent haplotypes which are no longer seen but can be inferred to have existed in the population history. The size of the circles is roughly proportional to the haplotype frequency in the population. The length of the connecting lines has no significance

**Table 4.3.** Pairwise linkage disequilibrium between *CRP* SNPs, calculated from the unrelated parents.\*

|              |              |              |              |              |              |           |              |              |              |           |            |              |              |           |           |            |           |           |           |
|--------------|--------------|--------------|--------------|--------------|--------------|-----------|--------------|--------------|--------------|-----------|------------|--------------|--------------|-----------|-----------|------------|-----------|-----------|-----------|
| rs12093699   | rs2592887    | rs16842502   | rs12749227   | rs1935193    | rs2808628    | rs2794520 | rs3093077    | rs2808630    | rs1205       | rs1130864 | rs1800947  | rs3091244C/A | rs3091244C/T | rs2794521 | rs3093059 | rs17860488 | rs2088632 | rs2027471 | rs7443007 |
| X            | 0.041        | 0.007        | 0.742        | 0.022        | 0.018        | 0.022     | 0.007        | 0.002        | 0.017        | 0.592     | 0.000      | 0.001        | 0.684        | 0.002     | 0.008     | 0.025      | 0.002     | 0.025     | 0.016     |
| 1.000        | X            | 0.233        | 0.052        | 0.681        | 0.881        | 0.870     | 0.241        | 0.298        | 0.828        | 0.052     | 0.092      | 0.253        | 0.112        | 0.291     | 0.228     | 0.793      | 0.293     | 0.792     | 0.826     |
| 1.000        | 0.962        | X            | 0.010        | 0.345        | 0.267        | 0.254     | 0.983        | 0.060        | 0.255        | 0.010     | 0.027      | 0.962        | 0.000        | 0.058     | 0.982     | 0.294      | 0.057     | 0.292     | 0.253     |
| 1.000        | 1.000        | 1.000        | X            | 0.032        | 0.030        | 0.033     | 0.010        | 0.011        | 0.027        | 0.873     | 0.001      | 0.003        | 0.854        | 0.009     | 0.011     | 0.041      | 0.010     | 0.041     | 0.027     |
| 1.000        | 0.957        | 1.000        | 1.000        | X            | 0.755        | 0.742     | 0.331        | 0.374        | 0.693        | 0.031     | 0.085      | 0.298        | 0.016        | 0.380     | 0.318     | 0.758      | 0.370     | 0.758     | 0.694     |
| rs1935193    | rs2808628    | rs2794520    | rs3093077    | rs2808630    | rs1205       | rs1130864 | rs1800947    | rs3091244C/A | rs3091244C/T | rs2794521 | rs3093059  | rs17860488   | rs2088632    | rs2027471 | rs7443007 |            |           |           |           |
| 0.702        | 0.944        | 1.000        | 0.782        | 1.000        | X            | 0.978     | 0.277        | 0.289        | 0.927        | 0.049     | 0.108      | 0.309        | 0.117        | 0.277     | 0.257     | 0.888      | 0.284     | 0.887     | 0.928     |
| 0.735        | 0.933        | 1.000        | 0.797        | 1.000        | 1.000        | X         | 0.259        | 0.313        | 0.923        | 0.051     | 0.108      | 0.281        | 0.119        | 0.305     | 0.228     | 0.886      | 0.308     | 0.885     | 0.923     |
| rs3093077    | rs2794520    | rs3093077    | rs2808630    | rs1205       | rs1130864    | rs1800947 | rs3091244C/A | rs3091244C/T | rs2794521    | rs3093059 | rs17860488 | rs2088632    | rs2027471    | rs7443007 |           |            |           |           |           |
| 1.000        | 0.963        | 1.000        | 1.000        | 0.971        | 1.000        | 1.000     | X            | 0.062        | 0.264        | 0.010     | 0.028      | 0.964        | 0.000        | 0.059     | 0.982     | 0.305      | 0.058     | 0.303     | 0.262     |
| rs2808630    | rs1205       | rs1130864    | rs1800947    | rs3091244C/A | rs3091244C/T | rs2794521 | rs3093059    | rs17860488   | rs2088632    | rs2027471 | rs7443007  |              |              |           |           |            |           |           |           |
| 0.536        | 0.970        | 1.000        | 0.935        | 0.950        | 0.970        | 1.000     | 1.000        | X            | 0.289        | 0.013     | 0.035      | 0.067        | 0.010        | 0.969     | 0.064     | 0.337      | 0.984     | 0.336     | 0.291     |
| rs2808630    | rs1205       | rs1130864    | rs1800947    | rs3091244C/A | rs3091244C/T | rs2794521 | rs3093059    | rs17860488   | rs2088632    | rs2027471 | rs7443007  |              |              |           |           |            |           |           |           |
| 0.682        | 0.931        | 1.000        | 0.767        | 0.986        | 0.989        | 0.977     | 1.000        | 1.000        | X            | 0.045     | 0.114      | 0.294        | 0.106        | 0.277     | 0.244     | 0.861      | 0.284     | 0.860     | 1.000     |
| rs1205       | rs1130864    | rs1800947    | rs3091244C/A | rs3091244C/T | rs2794521    | rs3093059 | rs17860488   | rs2088632    | rs2027471    | rs7443007 |            |              |              |           |           |            |           |           |           |
| 0.904        | 1.000        | 1.000        | 0.934        | 1.000        | 1.000        | 1.000     | 1.000        | 1.000        | 1.000        | X         | 0.001      | 0.012        | 1.000        | 0.012     | 0.011     | 0.057      | 0.013     | 0.057     | 0.045     |
| rs1800947    | rs3091244C/A | rs3091244C/T | rs2794521    | rs3093059    | rs17860488   | rs2088632 | rs2027471    | rs7443007    |              |           |            |              |              |           |           |            |           |           |           |
| 0.034        | 0.919        | 1.000        | 0.047        | 1.000        | 1.000        | 1.000     | 1.000        | 1.000        | 1.000        | 0.051     | X          | 0.027        | 0.000        | 0.034     | 0.028     | 0.098      | 0.035     | 0.100     | 0.114     |
| rs3091244C/A | rs3091244C/T | rs2794521    | rs3093059    | rs17860488   | rs2088632    | rs2027471 | rs7443007    |              |              |           |            |              |              |           |           |            |           |           |           |
| 1.000        | 0.961        | 0.981        | 0.379        | 0.935        | 1.000        | 1.000     | 1.000        | 1.000        | 1.000        | 1.000     | 1.000      | X            | X            | 0.064     | 0.946     | 0.339      | 0.065     | 0.337     | 0.291     |
| rs3091244C/A | rs3091244C/T | rs2794521    | rs3093059    | rs17860488   | rs2088632    | rs2027471 | rs7443007    |              |              |           |            |              |              |           |           |            |           |           |           |
| 0.903        | 1.000        | 1.000        | 0.924        | 0.908        | 1.000        | 1.000     | 1.000        | 1.000        | 1.000        | 1.000     | 1.000      | 1.000        | 1.000        | 0.713     | X         | 0.061      | 0.324     | 0.984     | 0.105     |
| rs3091244C/T | rs2794521    | rs3093059    | rs17860488   | rs2088632    | rs2027471    | rs7443007 |              |              |              |           |            |              |              |           |           |            |           |           |           |
| 0.446        | 0.969        | 1.000        | 0.848        | 0.974        | 0.969        | 1.000     | 1.000        | 1.000        | 1.000        | 1.000     | 1.000      | 0.981        | 1.000        | 1.000     | X         | 0.061      | 0.324     | 0.984     | 0.280     |
| rs2794521    | rs3093059    | rs17860488   | rs2088632    | rs2027471    | rs7443007    |           |              |              |              |           |            |              |              |           |           |            |           |           |           |
| 1.000        | 0.960        | 1.000        | 1.000        | 0.969        | 0.965        | 0.960     | 1.000        | 1.000        | 0.963        | 1.000     | 1.000      | 1.000        | 0.981        | 1.000     | X         | 0.275      | 0.061     | 0.274     | 0.243     |
| rs3093059    | rs17860488   | rs2088632    | rs2027471    | rs7443007    |              |           |              |              |              |           |            |              |              |           |           |            |           |           |           |
| 0.764        | 0.940        | 1.000        | 0.837        | 0.962        | 0.989        | 1.000     | 1.000        | 1.000        | 1.000        | 1.000     | 1.000      | 1.000        | 1.000        | 1.000     | 0.967     | X          | 0.333     | 1.000     | 0.862     |
| rs17860488   | rs2088632    | rs2027471    | rs7443007    |              |              |           |              |              |              |           |            |              |              |           |           |            |           |           |           |
| 0.524        | 0.968        | 1.000        | 0.873        | 0.948        | 0.968        | 1.000     | 1.000        | 1.000        | 1.000        | 1.000     | 1.000      | 1.000        | 0.681        | 1.000     | 1.000     | 1.000      | X         | 0.331     | 0.287     |
| rs2088632    | rs2027471    | rs7443007    |              |              |              |           |              |              |              |           |            |              |              |           |           |            |           |           |           |
| 0.762        | 0.940        | 1.000        | 0.836        | 0.962        | 0.988        | 1.000     | 1.000        | 1.000        | 1.000        | 1.000     | 1.000      | 1.000        | 1.000        | 1.000     | 0.967     | 1.000      | 1.000     | X         | 0.861     |
| rs2027471    | rs7443007    |              |              |              |              |           |              |              |              |           |            |              |              |           |           |            |           |           |           |
| 0.682        | 0.931        | 1.000        | 0.767        | 0.986        | 0.989        | 0.977     | 1.000        | 1.000        | 1.000        | 1.000     | 1.000      | 1.000        | 1.000        | 1.000     | 0.963     | 1.000      | 1.000     | 1.000     | X         |

\*  $r^2$  values presented at the top right and D' values at the bottom left.

## 4.4: Genetic Associations with basal CRP

### 4.4.1. Association of CRP SNPs with basal CRP

Each SNP within the *CRP* haplotype block was tested for association with basal CRP in a multiple linear regression model. Age, sex and SLE affection were incorporated as covariates along with a random effects factor to adjust for familial clustering of data. Biallelic SNPs were modelled assuming alleles acted in an additive manner. For triallelic SNP *rs3091244* each allele was assessed against a pool of the remaining alleles, with the most significant model (C vs. A/T) shown. The results of these analyses are presented in **table 4.4**. An individual  $P < 0.014$  was required to achieve study wide significance if a partial Bonferroni correction is applied (17 SNPs mean correlation,  $r = 0.558$ ). Two SNPs, *rs3091244* (C vs. A/T) and *rs3093077* therefore demonstrated an association with basal CRP expression, with the rare allele associated with higher CRP. SNPs *rs3093059* and *rs16842502* approached significance.

### 4.4.2: Trans-acting Genetic Effects.

I evaluated *trans*-acting effects from three genes, *APOE*, *IL1B*, and *IL6*, which are potentially associated with basal CRP expression. Once again, age, sex and SLE affection were included as standard covariates. Because I suspected the dominant genetic effect on basal CRP arose from variation at *CRP* itself I also included the most associated *CRP* SNP, namely *rs3093077* as a genetic covariate.

**Table 4.4.** Single SNP associations with CRP.

| Marker                        | $\beta$ coefficient | Standard Error | t      | Regression $P$ -value <sup>a</sup> |
|-------------------------------|---------------------|----------------|--------|------------------------------------|
| <i>rs7553007</i>              | 0.141               | 0.117          | 1.210  | 0.228                              |
| <i>rs2027471</i>              | 0.193               | 0.119          | 1.617  | 0.107                              |
| <i>rs2808632</i>              | -0.095              | 0.115          | -0.614 | 0.540                              |
| <i>rs17860488</i>             | 0.187               | 0.119          | -1.574 | 0.117                              |
| <i>rs3093059</i>              | 0.359               | 0.151          | 2.369  | 0.019                              |
| <i>rs2794521</i>              | -0.088              | 0.167          | -0.528 | 0.598                              |
| <i>rs3091244</i> <sup>b</sup> | 0.363               | 0.146          | 2.480  | 0.014                              |
| <i>rs1800947</i>              | 0.238               | 0.179          | 1.327  | 0.186                              |
| <i>rs1130864</i>              | 0.265               | 0.291          | 0.909  | 0.364                              |
| <i>rs1205</i>                 | 0.138               | 0.116          | 1.187  | 0.236                              |
| <i>rs2808630</i>              | -0.093              | 0.151          | -0.614 | 0.540                              |
| <i>rs3093077</i>              | 0.385               | 0.153          | 2.509  | 0.013                              |
| <i>rs2794520</i>              | 0.124               | 0.118          | 1.058  | 0.291                              |
| <i>rs2808628</i>              | 0.152               | 0.115          | 1.318  | 0.189                              |
| <i>rs1935193</i>              | 0.130               | 0.123          | 1.055  | 0.292                              |
| <i>rs12749227</i>             | 0.364               | 0.273          | 1.336  | 0.183                              |
| <i>rs16842502</i>             | 0.350               | 0.156          | 2.239  | 0.026                              |

$\beta$ -coefficient assuming increasing copies of the rare allele

<sup>a</sup>Unadjusted  $P$ -values. A  $P < 0.014$  is required to achieve study wide significance at the  $P = 0.05$  level after applying a partial Bonferroni correction (see text).

<sup>b</sup>SNP modelled as C vs. A/T alleles.

In an initial analysis I assumed the derived apolipoprotein  $\epsilon$  (Apo  $\epsilon$ ) serotypes acted in an additive manner on basal CRP expression. **Table 4.5** shows regression parameters for this model, with increasing copies of the rarer  $\epsilon 2$  or  $\epsilon 4$  type modelled by reference to the common  $\epsilon 3$  type.

**Table 4.5.** Trans-acting effects with a significant CRP association

| Variable                 | B coefficient | Standard Error | t      | Regression <i>P</i> -value <sup>a</sup> |
|--------------------------|---------------|----------------|--------|-----------------------------------------|
| <i>APOE</i>              |               |                |        |                                         |
| Additive model           |               |                |        |                                         |
| ε2                       | 0.198         | 0.227          | 0.869  | 0.386                                   |
| ε4                       | -0.089        | 0.255          | -0.348 | 0.728                                   |
| Recessive model          |               |                |        |                                         |
| ε2                       | 1.495         | 0.531          | 2.815  | 0.005                                   |
| ε4                       | 1.106         | 0.690          | 1.603  | 0.111                                   |
| <i>IL6</i>               |               |                |        |                                         |
| <i>rs1800796</i>         | -0.241        | 0.189          | -1.275 | 0.204                                   |
| <i>rs1800796</i> *no SLE | 0.387         | 0.187          | 2.063  | 0.040                                   |

<sup>a</sup>Regression model adjusted for age, sex, SLE affection and *rs3093077* genotype

The trend was towards a higher basal CRP in carriers of ε2 and a lower basal CRP in carriers of ε4 although this was not significant. I also investigated the possibility that Apo ε types acted in a dominant or recessive manner. In this model a highly significant effect on basal CRP was seen in individuals with ε2 ε2 type acting recessively. This remained significant at the  $P < 0.05$  level even after applying a conservative Bonferroni correction for 6 tests (three Apo ε types each tested in additive and recessive models). Using these regression parameters it could be predicted that an average unaffected 30 year old female, homozygous for the common allele of *rs3093077*, would have a basal CRP of 2.86mg/L if she carried the ε2 ε2 Apo ε type, compared with 1.56mg/L if she carried any other combination of Apo ε type.

Each *IL1B* and *IL6* SNP was also examined for its effect on basal CRP. Furthermore, and in view of the findings presented in section 3.4.2, I also assessed for interaction between genotype and SLE. No significant association was seen between genotype at any *IL1B* SNP and basal CRP. I also failed to observe an association between *IL6* genotype and basal CRP

in the cohort as a whole, but a weakly significant interaction between SLE and genotype at *rs1800796* was observed (**table 4.5**). This interaction was such that, in unaffected individuals, the rare allele of *rs1800796* was associated with higher basal CRP, while in SLE patients no such effect was observed, and indeed the rare allele of *rs1800796* appeared to have a negative effect (albeit not significant) on basal CRP.

#### 4.4.3: Mixed Genetic models

The association of *CRP* SNPs with basal CRP was then reassessed in the multiple linear regression model, now incorporating the key *trans*-acting genetic effects presented in section 4.4.2. This analysis is presented in **table 4.6**. The significance level attached to the four SNPs identified in the initial analysis (*rs3091244* (C vs. A/T), *rs3093077*, *rs3093059* and *rs16842502*) increased notably, with all four now falling well below the threshold for significance required according to the correction for multiple hypothesis testing.

It was noted as a chance observation that the inclusion of *rs1800947* as a covariate increased the significance of all the associated SNPs (also shown in **table 4.6**). A highly significant association with basal CRP was then also seen for *rs2027471*, *rs17860488*, *rs2794520* and *rs2808628* with weaker effects for *rs7553007*, *rs1205* and *rs1935193*. In addition *rs1800947* itself demonstrated a greater significance with basal CRP when modelled with another independently associated SNP; the minor allele being associated with higher basal CRP (for example if modelled with *rs3093077* the regression parameters for *rs1800947* were:  $\beta$  0.440, SE 0.208,  $t$  2.118,  $P=0.036$ ).

**Table 4.6.** Single SNP associations with serum CRP with genetic covariates

| Marker                        | $\beta$ coefficient | Standard Error | t      | Regression $P$ -value <sup>a,b</sup> | 1800947 adjusted $\beta$ | 1800947 adjusted $P$ -value <sup>a,c</sup> |
|-------------------------------|---------------------|----------------|--------|--------------------------------------|--------------------------|--------------------------------------------|
| <i>rs7553007</i>              | 0.274               | 0.135          | 2.027  | 0.044                                | 0.353                    | 0.011                                      |
| <i>rs2027471</i>              | 0.331               | 0.135          | 2.456  | 0.015                                | 0.397                    | 0.004                                      |
| <i>rs2808632</i>              | -0.031              | 0.169          | -0.185 | 0.853                                | 0.070                    | 0.704                                      |
| <i>rs17860488</i>             | 0.331               | 0.135          | -2.448 | 0.015                                | 0.396                    | 0.004                                      |
| <i>rs3093059</i>              | 0.452               | 0.158          | 2.860  | 0.005                                | 0.504                    | 0.003                                      |
| <i>rs2794521</i>              | -0.014              | 0.184          | 0.078  | 0.938                                | 0.001                    | 0.996                                      |
| <i>rs3091244</i> <sup>d</sup> | 0.433               | 0.158          | 2.743  | 0.007                                | 0.499                    | 0.002                                      |
| <i>rs1800947</i>              | 0.251               | 0.202          | 1.244  | 0.215                                | NA                       | NA                                         |
| <i>rs1130864</i>              | 0.155               | 0.310          | 0.501  | 0.617                                | 0.104                    | 0.738                                      |
| <i>rs1205</i>                 | 0.274               | 0.135          | 2.027  | 0.044                                | 0.353                    | 0.011                                      |
| <i>rs2808630</i>              | -0.015              | 0.170          | -0.088 | 0.930                                | 0.079                    | 0.668                                      |
| <i>rs3093077</i>              | 0.470               | 0.161          | 2.912  | 0.004                                | 0.506                    | 0.003                                      |
| <i>rs2794520</i>              | 0.255               | 0.132          | 1.935  | 0.055                                | 0.351                    | 0.009                                      |
| <i>rs2808628</i>              | 0.299               | 0.131          | 2.282  | 0.024                                | 0.384                    | 0.005                                      |
| <i>rs1935193</i>              | 0.268               | 0.140          | 1.909  | 0.058                                | 0.362                    | 0.014                                      |
| <i>rs12749227</i>             | 0.285               | 0.302          | 0.945  | 0.346                                | 0.257                    | 0.394                                      |
| <i>rs16842502</i>             | 0.451               | 0.164          | 2.752  | 0.007                                | 0.474                    | 0.006                                      |

$\beta$ -coefficient assuming increasing copies of the rare allele

<sup>a</sup>Unadjusted  $P$ -values. A  $P < 0.014$  is required to achieve study wide significance at the  $P = 0.05$  level after applying a partial Bonferroni correction (see text).

<sup>b</sup>Regression model corrected for age, sex, SLE affection, Apo  $\epsilon 2$  and *rs1800796* genotype.

<sup>c</sup>Regression model additionally adjusted for *rs1800947* genotype

<sup>d</sup>SNP modelled as C vs. A/T alleles.

#### 4.4.4: Haplotypic Association with basal CRP

The association between *CRP* haplotype and basal CRP was assessed for the five most common haplotypes. The full regression model, also showing the covariate parameters is presented in **table 4.7**.

**Table 4.7.** The association of CRP with *CRP* haplotypes

| Variable                      | $\beta$ coefficient | Standard Error | t-test | Regression <i>P</i> -value |
|-------------------------------|---------------------|----------------|--------|----------------------------|
| Age                           | 0.014               | 0.005          | 2.689  | 0.008                      |
| Gender (female)               | 0.381               | 0.167          | 2.280  | 0.024                      |
| SLE (affected)                | 0.540               | 0.216          | 2.499  | 0.013                      |
| Apo $\epsilon$ 2 $\epsilon$ 2 | 1.474               | 0.584          | 2.526  | 0.013                      |
| <i>rs1800796</i>              | 0.272               | 0.222          | 1.227  | 0.222                      |
| <i>rs1800796</i> *no SLE      | -0.563              | 0.216          | -2.618 | 0.010                      |
| F1                            | Ref <sup>a</sup>    |                |        |                            |
| F2                            | 0.205               | 0.184          | 1.111  | 0.268                      |
| F4                            | 0.441               | 0.222          | 1.984  | 0.049                      |
| F5                            | 0.445               | 0.342          | 1.304  | 0.194                      |
| F3                            | 0.596               | 0.187          | 3.185  | 0.002                      |
| Fx                            | 0.276               | 0.288          | 0.960  | 0.339                      |

<sup>a</sup>Haplotype parameters modelled with reference to commonest haplotype F1. Fx represents the group of rare haplotypes in the population (<3%).

There is clearly a significant difference in basal CRP levels associated with commonest haplotype F1 (low basal CRP) and F3 (high basal CRP). F2 appears associated with intermediate basal CRP, while F4 and F5 are also associated with higher CRP (this is of borderline significance for F4 but not significant for F5, presumably because of the small numbers in this group). During haplotype inference it was noted that the sixth and seventh commonest haplotypes (present on 2.5 and 1.8% of chromosomes respectively) shared significant sequence homology with F1, and could be assigned to the F1 clade (**figure 4.1** and **figure 4.3**). Neither of these haplotypes, individually or combined, was common enough to draw any significant conclusions, however we did find that the H6/H7 group was associated with a similar, low,  $\beta$ -coefficient to F2 (compared to F3:  $\beta$  -0.564, S.E. 0.371,  $t$  -1.369,  $P=0.173$ ), suggesting they may be genuinely “low producing” haplotypes.

## 4.5: CRP Genetics in Filipino SLE Families – Discussion

### 4.5.1. Summary of Findings

In this chapter I explored the relationship between genotype and basal CRP in a cohort of Filipino SLE families. I was unable to re-examine the association between *CRP* genotype and disease as, with a rather unconventional cohort structure, there was insufficient power to perform this analysis. Instead I aimed to explore the genetic architecture of *CRP* in this previously unstudied Filipino population, with the intention of identifying novel haplotypes that aid the identification of functional SNPs or to utilise differences in population allele frequency to identify new associations. Four SNPs were found to be associated with higher basal CRP. The minor allele of three of these (*rs3093059*, *rs3093077* and *rs16842502*) defined a haplotype, found in 17% of chromosomes, which was also associated with higher basal CRP levels in comparison with the commonest population haplotype. A similar strong association was seen with tri-allelic SNP *rs3091244* with the commonest “C” allele associated with lower serum CRP than a combination of the “T” allele (carried on the above mentioned haplotype and the much rarer “A” allele (itself defining a low frequency haplotype). The association, albeit borderline ( $P=0.049$ ), of the uncommon haplotype F4 (characterised by the minor allele of *rs1800947*), with higher CRP than the reference haplotype (F1) prompted me to re-analyse the single SNP data proposing a functional role for *rs1800947*. In support of this hypothesis, the modelling of *rs1800947* as an independent covariate increased the significance of all the associated SNPs. Trans-acting associations were also observed, in particular the association of lower serum CRP in individuals homozygous for the *APOE*  $\epsilon 2$  serotype.

#### 4.5.2: A comparison with existing studies – cis-effects

A detailed comparison of inter-ethnic differences in *CRP* genetic architecture will be presented later (section 8.4.3). Nonetheless it is important to compare the data presented in this chapter with that presented in chapter 3 and elsewhere in the published literature. Each of the key associated SNPs identified in this Filipino cohort have also been associated with basal *CRP* expression elsewhere in the literature, again with the minor alleles associated with higher serum *CRP* (Kathiresan *et al* 2006; Miller *et al* 2005; Wang *et al* 2006). It is notable that the set of key associated SNPs identified in the analysis presented in chapter 3 was not significantly associated with serum *CRP* in this study, although the same alleles were observed to have a trend towards lower serum *CRP* (note that the minor/major alleles are swapped in this Filipino group, hence the positive  $\beta$ -coefficient on the regression analyses). A detailed discussion of why studies may highlight significant associations with different groups of SNPs (although all tend to agree on the trend in serum *CRP* seen with the different SNP alleles) is presented at a later stage (section 8.4.).

The weak association of haplotype F4 and its unique SNP, *rs1800947*, with basal *CRP* was interesting. Existing studies typically find an association between low basal *CRP* and the rare “C” allele of *rs1800947* along with the haplotype it defines (Kathiresan *et al* 2006; Lange *et al* 2006b; Miller *et al* 2005). A single study demonstrates a significant association between homozygosity for the minor allele and higher basal serum *CRP* (Crawford *et al* 2006). In this Filipino population *rs1800947* “C”, along with haplotype F4 were almost twice as common as their white European equivalents and we therefore had greater power to detect an association. We found F4 to be associated with higher basal *CRP* than F1 ( $P=0.049$ ). The inclusion of *rs1800947* as an independent genetic covariate in the SNP association model increased the significance of *all* the associated SNPs, supporting a hypothesis that there is an

independent genetic effect arising from this SNP, or one in close LD with it. Phylogenetically this is also feasible if, as we propose, haplotype F4 forms a distinct clade diverging from the ancestral haplotype F1 independently of all the other haplotypes. There are a number of explanations why this effect is not commonly observed in the white European cohorts. It is possible that *rs1800947* “C” is in linkage disequilibrium (LD) with a true functional SNP in Filipinos but not white Europeans: sequencing these haplotypes may be required to demonstrate this. In both white Europeans and Filipinos *rs1800947* “C” is always found in association with the *rs1205* “A” allele and other SNPs in strong LD with *rs1205*. Therefore an alternative explanation is that *rs1800947* is truly functional but in white Europeans, because it is less frequent (approximate reported frequency 5-6%), its effect is swamped by an opposite functional effect arising on the *rs1205* “A” clade. On the other hand in Filipinos where we observed *rs1800947* at higher frequency (12%) we have greater power to observe an independent functional effect.

#### *4.5.3: A comparison with existing studies – trans-effects*

As with the analysis presented in chapter 3, the primary aim of including genetic data from *trans*-acting loci in the analysis was to reduce the levels of “background noise” (random measurement error) arising from other genetic influences, thus increasing the power of our analysis to identify *cis*-acting *CRP* associations. Of particular interest was my identification of an association between *APOE* polymorphism (defining Apo ε serotype) and serum CRP. The link between Apo ε variants and levels of serum CRP has been reported for nearly ten years, although there remains no convincing mechanistic explanation for this association. The knowledge that apolipoprotein E plays an important role in lipid metabolism and cholesterol transport, while lipid lowering statins also lower serum CRP, suggests a link through lipid regulatory mechanisms, although we lack any more detailed evidence. Of the three serotypes,

the Apo  $\epsilon 2$  variant is rare in European populations, such that many of the published studies have no individuals of  $\epsilon 2 \epsilon 2$  serotype in their cohorts. While the observation that carriers of the Apo  $\epsilon 4$  variant have the lowest serum CRP is almost universal, the role of the Apo  $\epsilon 2$  variants is less certain because European studies are underpowered to detect any effect over and above the difference between Apo  $\epsilon 3$  vs. Apo  $\epsilon 4$  carriers (Chasman *et al* 2006; Judson *et al* 2004; Lange *et al* 2006a; Manttari *et al* 2001; Rontu *et al* 2006). In populations of an East Asian origin the Apo  $\epsilon 2$  variant appears to be much more common. A study in Japanese-Americans showed a higher frequency of the Apo  $\epsilon 2$  serotype and found an association between Apo  $\epsilon 2$  type and higher levels of CRP (Austin *et al* 2004). The finding from our Filipino population where the Apo  $\epsilon 2$  variant was also more common appears to support this finding, although only when analysed in a recessive model. I suspect that ApoE type exerts a graded effect on basal CRP levels with Apo  $\epsilon 2$  associated with highest levels, Apo  $\epsilon 3$  with intermediate levels and Apo  $\epsilon 4$  with lowest levels. Certainly a functional explanation limited to Apo  $\epsilon 4$  would appear to be inadequate.

I also demonstrated an association between *IL6* variation and serum CRP, although only when modelled with SLE as an interaction term. Interestingly this same *IL6* SNP was also highlighted in the European analysis presented in chapter 3. It is difficult to make a direct comparison of the two groups because of marked differences in allele frequency and patterns of linkage disequilibrium, nonetheless the demonstration of a consistent association between basal CRP and an *IL6* variant adds weight to this being a true effect. Future work in basal CRP cohorts and further fine-mapping is required to fully delineate this genetic effect.

#### *4.5.4: Key study features and potential limitations.*

The Filipino population has, to date, been poorly characterised genetically, with no existing reports looking at *CRP*: this, in itself, makes the data presented in this chapter of interest. It should be recognised, however, that this is a potential problem when aiming to ensure that genetic variation has been comprehensively captured using a tagSNP approach. As outlined in the methods chapter (section 2.2.1), the HapMap Han Chinese was used as the reference population for the selection of Filipino tagSNPs. Although the Filipino population is clearly closer related to Han Chinese than any of the other HapMap reference populations, the relationship is not that close, raising the possibility that there may be differences in haplotype structure, or even entirely novel variants that have not been identified in the current study. This could really only be addressed by the complete resequencing of the *CRP* locus in a representative sample of, say 50-100 Filipino's. Also not addressed was the possibility of underlying population admixture, for example with recent southern European (Spanish) elements, although to some extent the family-based design of the study should overcome this potential problem.

As with the data presented in chapter 3, I am confident in the quality of genotyping and, since this was a family based cohort, a Mendel error check provided an additional level of confidence in the assigning of genotypes. There are, however, clear limitations to this analysis, in particular related to the familial structure of the cohort, and the reduced level of power to detect genetic associations that occurs as a consequence. As stated in the introduction to this chapter, the Filipino cohort is currently under recruitment and at the time of analysis was limited, particularly in terms of data from family members. As outlined in section 2.1.2 only 27 complete proband/parent trios were available for analysis. Even by inferring parental genotype by making use of genotype data from other siblings (as used by

TRANSMIT) at most 100 informative trios would be available for analysis. This clearly was inadequate to identify even reasonably strong genetic associations using a transmission disequilibrium test. For this reason a disease-based genetic association analysis was not performed. The cohort structure also impacted on the power to perform a quantitative association analysis with serum CRP. In particular there were a limited number of families with data on additional siblings (again barely 300), although a few families had data on >5 siblings. This lack of sibling data results in an imprecise variance components estimate of the intra-family correlation of serum CRP levels, and hence introduces significant random error into the SNP association analysis.

There are clearly other significant sources of random error in this analysis, including the mix of patients affected and unaffected by SLE, and potential random error introduced by the potential for sampling affected patients at different stages of disease activity. These aspects, along with others of relevance to this analysis have already been fully covered in the discussion in section 3.7.5.

#### *4.5.5: Conclusions*

I have mapped the *CRP* locus in a novel cohort of ethnic Filipinos confirming strong LD across the locus in this population. I have replicated the association of a number of SNPs with serum basal CRP, although these SNPs were different from those found in the analysis of white European individuals presented in chapter 3. Nonetheless, with the exception of SNP *rs1800947* a broadly similar trend in CRP levels was found for each SNP typed in common between the two studies. SNP *rs1800947* was interesting because there was some evidence that a separate genetic effect may be arising from this SNP, or one in close LD with it. A further discussion of why different SNPs appear significant in different studies, and whether

there is evidence from other studies of multiple genetic effects arising from *CRP* may be found in section 8.4.

## **CHAPTER 5**

# ***CRP* GENETICS IN AN AFRICAN-AMERICAN COHORT**

### 5.1: *CRP* Genetics in an African-American Cohort – Study Aims

The data presented in chapters 3 and 4 confirm an association between variation at the *CRP* locus and serum CRP levels. However, for reasons that will be discussed further in chapter 8, it remains difficult to establish which SNP or SNPs are likely to be functional variants, and in this respect my data simply replicates and extends existing published reports. In this chapter I aim to unpick the multiple reported *CRP* SNP associations to identify the most likely functional variants that can then be targeted for functional study. I do this both through the selection of the cohort for study, and through the analysis that I employed.

As I have already discussed (section 1.5.3), populations of African ancestry generally demonstrate increased genetic diversity, with shorter and more diverse haplotype blocks. It is therefore likely that a *CRP* association analysis in an African-American population will, at least, limit the number of indirect associations observed, and thus will generate a more restricted range of potential functional variants. The number of existing studies in this population is limited.

I also adopted a different analytic approach to help highlight key variants and resolve uncertainty in the literature. The focus of this study was not, therefore, to demonstrate an association between serum CRP and a string of individual variants; rather I attempted, through a formal model choice analysis, to identify the optimal genetic model, containing one or more SNPs which explain the overall *CRP* association. By using a Bayesian approach to analysis I was able to utilise existing data to make some simple prior assumptions about the nature of the *CRP* genetic effect. These prior assumptions were then built into a Bayesian model choice framework to identify the genetic model “best” explaining the CRP

distribution. The model could then be retested to assess its robustness to changes in prior assumptions.

In a parallel analysis I adopted a clade-based haplotype association approach. The use of a haplotype based approach may itself offer increased power over an analysis of single SNPs, particularly if the overall phenotype is influenced by more than one SNP at a locus. The data presented in this chapter can be found in published form (Rhodes *et al* 2008b).

## **5.2: African-American Population Cohort – Demographics and CRP distribution**

As I have already outlined in section 2.1.3. the cohort employed for the analyses presented in this chapter was an amalgamation of patients derived from two separate patient collections.

As each cohort was relatively small it was felt that a study restricted to each cohort individually was likely to be underpowered. Subjects from cohort 3a (Chicago Health and Aging Project) had a mean age of 78.3 years (standard deviation (s.d.) 7.3) and 37.4% were male. Phenotypic data were available for Body Mass Index (BMI) (mean 28.24kg/m<sup>2</sup>, s.d. 6.3), smoking status (10.5%) and an individual's history of stroke (11.6%), cancer (23.8%), hypertension (81.0%) ischaemic heart disease (12.6%) or diabetes mellitus (30.3%). Median (95th percentile) CRP was 3.38 mg/L (0.18, 27.50). Subjects from Cohort 3b (UCSF Multiple Sclerosis Genetics Group) had a mean age of 43.5 years (s.d. 11.3) and 39.5% were male. 50% had multiple sclerosis but otherwise no other phenotype data were available. Median (95th percentile) CRP was 1.79 mg/L (0.13, 16.81).

Median (95th percentile) CRP was 2.52 mg/L (0.24, 22.75) in the combined dataset. A Bayesian model choice analysis was used to evaluate the combination of phenotypic covariates that best explained the CRP data. The highest posterior probability ( $P_{(M/D)}$ ) was

observed for a model incorporating age and BMI only ( $P_{(M/D)}=0.98$ , representing overwhelming support). There was little support for models including sex or a diagnosis of multiple sclerosis (MS) and the marginal probabilities ( $P_{\text{marg}}$ ) for these covariates were small ( $P_{\text{marg}}=0.002$  and  $P_{\text{marg}}=0.0003$  respectively). There was even less support for other diseases including ischaemic heart disease and diabetes mellitus, although given the small numbers of affected individuals and the absence of data from cohort 2, this was not surprising. We looked for additional cohort specific effects and found strong evidence for no cohort effect (age, BMI, sex and MS adjusted:  $P_{(M/D)}=0.008$  for model with cohort effect,  $P_{(M/D)}=0.992$  for model with no cohort effect, Bayes factor  $>100$ )

### 5.3: African-American *CRP* genetic architecture

The 44 SNPs genotyped in the whole dataset are listed in **table 5.1**. Eight SNPs failed genotyping according to predetermined criteria and were excluded from further analysis. Due to slightly different SNP pooling constraints at each of the two sites at which the genotyping was performed, data were also available for an additional five SNPs for either cohort 3a or 3b only (**table 5.2**). With the exception of SNP *rs11585798*, which is highlighted in the table, all SNPs were in Hardy-Weinburg equilibrium both in the combined cohort and in cohorts 3a and 3b separately. Initially haplotypes were inferred both for cohorts 3a and 3b separately, which demonstrated that haplotype size, structure and diversity was similar in both cohorts. The only notable difference appeared to be a higher frequency of haplotype A2 in cohort 3a (15.3% vs. 20.1%) and this was largely accounted for by a lower frequency of the related haplotype A8 (2.6% vs. 4.4%). Since haplotypes A2 and A8 were closely related and formed part of the same haplotype clade I felt that this small difference in relative haplotype frequency was unlikely to impact on the validity of an analysis of the combined cohort. In view of this, haplotypes were re-inferred for the combined dataset, incorporating the

additional data from SNPs that were genotyped in either cohort 3a or 3b only. **Figure 5.1** demonstrates the position of these genotyped SNPs alongside the inferred haplotype structure, while **table 5.3** shows the pairwise LD statistics between markers. For both these analyses individuals with one or more haplotype of predicted European origin were excluded (see Section 5.4.3). It should be noted that none of the additional SNPs typed in only one cohort defined a new haplotype or altered overall haplotype structure in any way. It should be emphasised that these additional SNPs were not included in the genetic association analysis but I felt that their inclusion in the haplotype inference allowed a useful direct comparison with data presented in other chapters of this thesis.

**Table 5.1.** Typed SNPs demonstrating alleles and rare allele frequencies

| rs number                      | Location            | alleles | Rare allele frequency (%) | HWP <sup>a</sup> | missing data (%) |
|--------------------------------|---------------------|---------|---------------------------|------------------|------------------|
| <i>CRP</i> SNPs                |                     |         |                           |                  |                  |
| <i>rs11265263</i> <sup>§</sup> | 5' flank            | T>G     | 0.03                      | 0.43             | 1.1              |
| <i>rs12728740</i>              | 5' flank            | G>T     | 0.10                      | 0.96             | 0.9              |
| <i>rs7553007</i>               | 5' flank            | C>T     | 0.21                      | 0.20             | 2.5              |
| <i>rs1341665</i>               | 5' flank            | C>T     | 0.19                      | 0.24             | 4.8              |
| <i>rs12079772</i>              | 5' flank            | G>T     | 0.20                      | 0.09             | 1.3              |
| <i>rs2027471</i>               | 5' flank            | A>T     | 0.20                      | 0.56             | 0.7              |
| <b><i>rs12084589</i></b>       | 5' flank            | G>T     | 0.28                      | 0.80             | 2.9              |
| <i>rs3093058</i> *             | 5' flank            | A>T     | 0.17                      | 0.94             | 5.6              |
| <b><i>rs3093062</i></b>        | upstream promoter   | G>A     | 0.16                      | 0.64             | 0.9              |
| <i>rs3091244</i> *             | upstream promoter   | C>T>A   | 0.31(T)/0.25(A)           | 0.48             | 1.1              |
| <i>rs3093063</i> <sup>§</sup>  | upstream promoter   | C>T     | 0.03                      | 0.27             | 0.9              |
| <b><i>rs1417938</i></b>        | intron 1            | A>T     | 0.14                      | 0.78             | 0.0              |
| <i>rs1800947</i> <sup>§</sup>  | exon 2 (synonymous) | G>C     | 0.02                      | 0.67             | 0.5              |
| <b><i>rs3093066</i></b>        | 3'-UTR              | C>A     | 0.22                      | 0.26             | 2.0              |
| <i>rs1205</i> *                | 3'-UTR              | G>A     | 0.18                      | 0.92             | 0.9              |
| <i>rs3093069</i>               | 3' flank            | T>G     | 0.20                      | 0.12             | 1.6              |
| <b><i>rs2808630</i></b> *      | 3' flank            | A>G     | 0.18                      | 0.57             | 0.2              |
| <b><i>rs3093077</i></b> *      | 3' flank            | T>G     | 0.29                      | 0.55             | 0.7              |
| <i>rs2794520</i>               | 3' flank            | G>A     | 0.21                      | 0.75             | 0.4              |
| <i>rs2808628</i>               | 3' flank            | C>T     | 0.21                      | 0.84             | 1.4              |
| <b><i>rs11265259</i></b>       | 3' flank            | A>G     | 0.06                      | 0.27             | 8.8              |
| <b><i>rs1572970</i></b>        | 3' flank            | C>T     | 0.35                      | 0.76             | 1.4              |
| <i>rs12755606</i>              | 3' flank            | G>C     | 0.10                      | 0.82             | 0.9              |
| <b><i>rs11265257</i></b>       | 3' flank            | G>A     | 0.20                      | 0.96             | 0.5              |
| <b><i>rs2808626</i></b>        | 3' flank            | C>A     | 0.08                      | 0.80             | 0.2              |
| <i>rs2369146</i>               | 3' flank            | C>T     | 0.37                      | 0.84             | 5.9              |
| <i>rs17459580</i> <sup>§</sup> | 3' flank            | T>G     | 0.03                      | 0.43             | 0.7              |
| <b><i>rs1935193</i></b>        | 3' flank            | A>T     | 0.46                      | 0.76             | 0.7              |
| <i>rs10465953</i> <sup>§</sup> | 3' flank            | C>T     | 0.02                      | 0.67             | 0.2              |
| <i>rs12749227</i>              | 3' flank            | A>G     | 0.12                      | 0.56             | 0.7              |
| <i>rs16842502</i> <sup>§</sup> | 3' flank            | G>T     | 0.04                      | 0.78             | 15.6             |
| <b><i>rs1470515</i></b>        | 3' flank            | G>A     | 0.25                      | 0.81             | 1.3              |
| <b><i>rs16842493</i></b>       | 3' flank            | T>C     | 0.20                      | 0.20             | 0.9              |
| <i>rs12093699</i>              | 3' flank            | C>T     | 0.33                      | 0.43             | 2.0              |
| <i>rs12744244</i>              | 3' flank            | G>T     | 0.05                      | 0.68             | 1.4              |
| <i>rs11585798</i> <sup>§</sup> | 3' flank            | T>A     | 0.04                      | <0.001           | 2.7              |
| <i>IL6</i> SNPs                |                     |         |                           |                  |                  |
| <i>rs1554606</i>               | 5' flank            | G>T     | 0.36                      | 0.71             | 2.6              |
| <i>rs1800796</i>               | 5' flank            | G>C     | 0.08                      | 0.89             | 0.2              |
| <i>rs1800795</i>               | 5' flank            | G>C     | 0.08                      | 0.64             | 0.9              |
| <i>rs2069837</i>               | intron 2            | A>G     | 0.14                      | 0.16             | 2.1              |
| <i>rs2069840</i>               | intron 3            | C>G     | 0.15                      | 0.77             | 0.4              |
| <i>IL1B</i> SNPs               |                     |         |                           |                  |                  |
| <i>rs3917368</i>               | 3' flank            | G>A     | 0.20                      | 0.56             | 2.3              |
| <i>rs4849124</i> <sup>§</sup>  | 3' flank            | A>G     | 0.02                      | <0.001           | 11.0             |
| <i>rs1143634</i>               | exon 5              | C>T     | 0.14                      | 0.09             | 1.2              |

SNPs highlighted in bold have been selected to tag the key *CRP* haplotype block. SNPs marked with a \* have been previously associated with serum CRP levels in African-American populations (see **table 1.2**)

<sup>§</sup>excluded from analysis by pre-determined quality control criteria.

<sup>a</sup>Hardy-Weinburg probability *P*-value calculated by  $\chi^2$

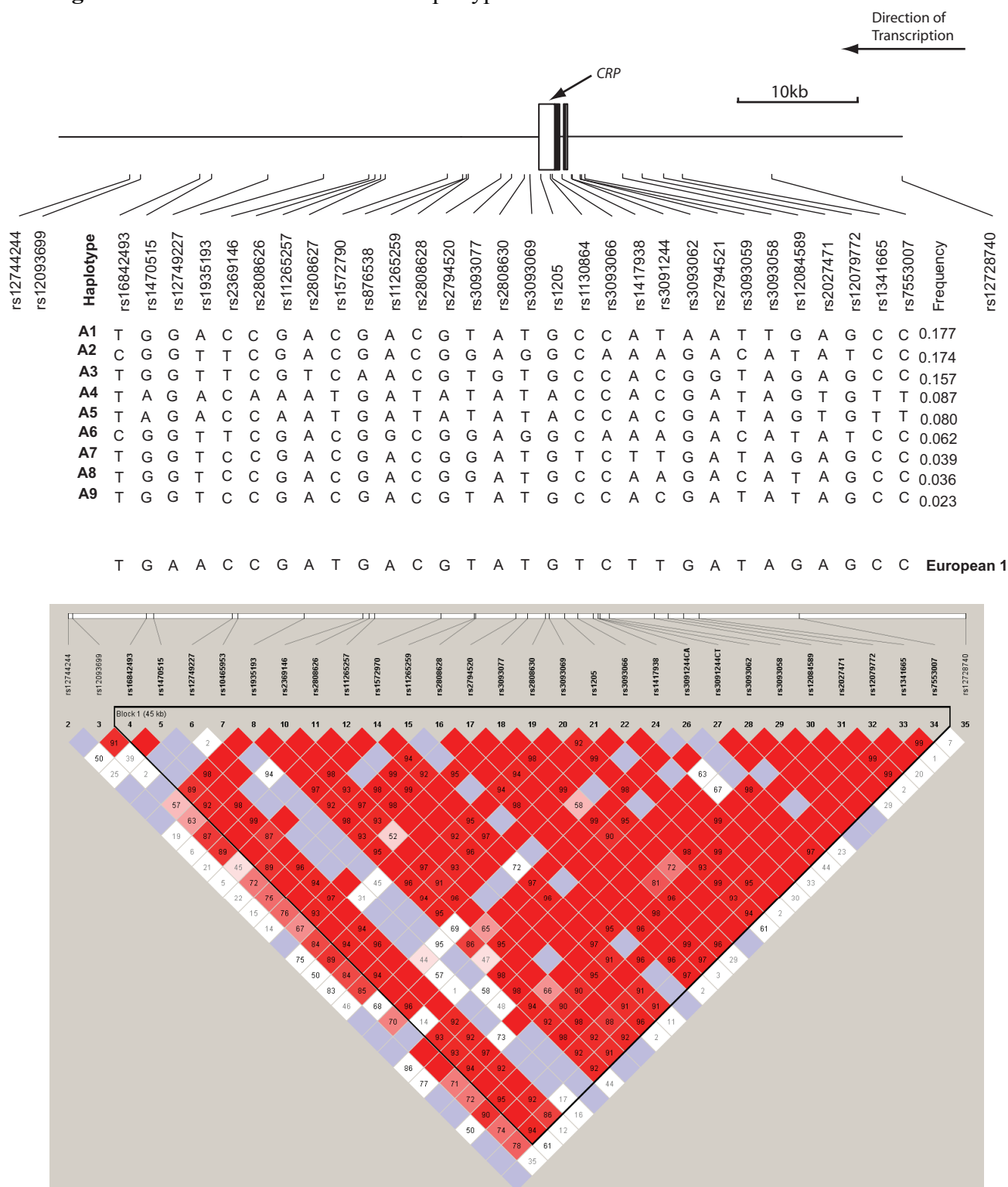
**Table 5.2.** *CRP* SNPs typed in one cohort only<sup>a</sup>

| rs number        | Location | Alleles | minor allele frequency (%) | HWP <sup>b</sup> | missing data (%) |
|------------------|----------|---------|----------------------------|------------------|------------------|
| Cohort 3a        |          |         |                            |                  |                  |
| <i>rs3093059</i> | 5' flank | T>C     | 24.15                      | 0.13             | 5.7              |
| <i>rs2794521</i> | 5' flank | A>G     | 12.41                      | 0.08             | 3.9              |
| <i>rs876538</i>  | 3' flank | G>A     | 12.18                      | 0.96             | 2.1              |
| Cohort 3b        |          |         |                            |                  |                  |
| <i>rs1130864</i> | 3'-UTR   | C>T     | 16.79                      | 0.50             | 9.2              |
| <i>rs2808627</i> | 3' flank | A>T     | 15.46                      | 0.06             | 0.3              |

<sup>a</sup>These SNPs were incorporated into the inferred haplotype structure as outlined in section 5.3 but were not included in the genetic association analysis.

<sup>b</sup>Hardy-Weinburg probability *P*-value calculated by  $\chi^2$

**Figure 5.1.** African-American *CRP* haplotypes.



The position of SNPs relative to *CRP* is shown in the top section of this diagram. Individuals predicted to carry haplotypes of a recent European origin have been excluded from this analysis, although the commonest of these haplotypes is shown at the bottom of the diagram for reference. The middle section demonstrates inferred haplotype structure with haplotype frequencies are shown to the right of each block. The lower section demonstrates the pairwise LD between markers with haplotype blocks defined by the default algorithm of Gabriel *et al* (see methods section 2.4.4). The pairwise LD between any two SNPs is indicated by the colour coding of the corresponding square of the LD plot. High LD is indicated by a pink/red colour scheme (the dark red squares are linked with a  $D' = 1$  and a LOD score  $\geq 2$ ). Numbers within the squares correspond to the  $D'$  score, with 96 representing a  $D'=0.96$  and 44 a  $D'=0.44$  etc. Where squares have no number the  $D'=1$

**Table 5.3:** Pairwise linkage disequilibrium between *CRP* SNPs.

|              |       |       |       |       |       |       |       |       |       |       |        |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |            |       |       |              |              |            |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------------|-------|-------|--------------|--------------|------------|
| rs12744244   | X     | 0.012 | 0.004 | 0.001 | 0.000 | 0.004 | 0.000 | 0.000 | 0.001 | 0.000 | 0.003  | 0.000 | 0.004 | 0.001 | 0.003 | 0.000 | 0.010 | 0.000 | 0.001 | 0.001 | 0.010 | 0.001 | 0.004 | 0.001 | 0.001 | 0.000 | rs12728740 |       |       |              |              |            |
| rs12093699   | 1.000 | X     | 0.638 | 0.020 | 0.000 | 0.118 | 0.220 | 0.037 | 0.078 | 0.028 | 0.091  | 0.063 | 0.066 | 0.372 | 0.067 | 0.613 | 0.069 | 0.610 | 0.011 | 0.459 | 0.047 | 0.099 | 0.099 | 0.416 | 0.059 | 0.620 | 0.055      | 0.073 | 0.002 | rs7443007    |              |            |
| rs16842493   | 0.500 | 0.912 | X     | 0.100 | 0.007 | 0.223 | 0.352 | 0.037 | 0.075 | 0.104 | 0.185  | 0.085 | 0.088 | 0.554 | 0.072 | 0.892 | 0.075 | 0.82  | 0.018 | 0.656 | 0.001 | 0.067 | 0.066 | 0.565 | 0.086 | 0.898 | 0.078      | 0.083 | 0.001 | rs1341665    |              |            |
| rs1470515    | 0.259 | 0.391 | 1.000 | X     | 0.006 | 0.348 | 0.245 | 0.359 | 0.746 | 0.735 | 0.022  | 0.802 | 0.781 | 0.149 | 0.064 | 0.089 | 0.703 | 0.111 | 0.017 | 0.288 | 0.253 | 0.060 | 0.054 | 0.149 | 0.739 | 0.099 | 0.704      | 0.697 | 0.000 | rs12079772   |              |            |
| rs12749227   | 1.000 | 0.028 | 1.000 | 0.986 | X     | 0.025 | 0.018 | 0.002 | 0.005 | 0.069 | 0.002  | 0.005 | 0.005 | 0.011 | 0.000 | 0.007 | 0.005 | 0.008 | 0.076 | 0.003 | 0.000 | 0.005 | 0.005 | 0.006 | 0.005 | 0.006 | 0.007      | 0.005 | 0.009 | rs3093062    |              |            |
| rs1935193    | 1.000 | 0.572 | 0.897 | 1.000 | 1.000 | X     | 0.671 | 0.133 | 0.270 | 0.355 | 0.053  | 0.297 | 0.278 | 0.406 | 0.177 | 0.261 | 0.241 | 0.279 | 0.022 | 0.267 | 0.067 | 0.270 | 0.272 | 0.400 | 0.264 | 0.265 | 0.243      | 0.287 | 0.002 | rs3091244C/T |              |            |
| rs2369146    | 0.197 | 0.637 | 0.925 | 0.981 | 1.000 | 1.000 | X     | 0.088 | 0.182 | 0.242 | 0.079  | 0.197 | 0.184 | 0.185 | 0.270 | 0.393 | 0.152 | 0.424 | 0.045 | 0.196 | 0.168 | 0.185 | 0.185 | 0.293 | 0.172 | 0.398 | 0.163      | 0.187 | 0.005 | rs3091244C/A |              |            |
| rs2808626    | 0.063 | 0.879 | 1.000 | 0.991 | 1.000 | 1.000 | 1.000 | X     | 0.491 | 0.367 | 0.008  | 0.415 | 0.406 | 0.058 | 0.024 | 0.037 | 0.425 | 0.041 | 0.006 | 0.096 | 0.090 | 0.025 | 0.023 | 0.060 | 0.350 | 0.037 | 0.357      | 0.349 | 0.001 | rs1417938    |              |            |
| rs11265257   | 0.210 | 0.894 | 1.000 | 0.876 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.749 | 0.016  | 0.862 | 0.842 | 0.119 | 0.048 | 0.065 | 0.933 | 0.083 | 0.012 | 0.217 | 0.186 | 0.052 | 0.048 | 0.122 | 0.793 | 0.075 | 0.761      | 0.757 | 0.000 | rs3093066    |              |            |
| rs1572970    | 0.057 | 0.459 | 1.000 | 1.000 | 1.000 | 0.974 | 0.981 | 1.000 | 1.000 | 1.000 | X      | 0.022 | 0.762 | 0.741 | 0.163 | 0.067 | 0.094 | 0.706 | 0.113 | 0.009 | 0.268 | 0.255 | 0.070 | 0.063 | 0.156 | 0.713 | 0.102      | 0.671 | 0.752 | 0.000        | rs1205       |            |
| rs11265259   | 0.229 | 0.721 | 0.898 | 0.966 | 1.000 | 0.924 | 0.939 | 1.000 | 1.000 | 1.000 | 1.000  | X     | 0.018 | 0.019 | 0.130 | 0.017 | 0.235 | 0.016 | 0.208 | 0.004 | 0.163 | 0.005 | 0.017 | 0.017 | 0.132 | 0.019 | 0.223      | 0.018 | 0.020 | 0.001        | rs3093077    |            |
| rs2808628    | 0.150 | 0.752 | 1.000 | 0.971 | 1.000 | 0.984 | 0.976 | 0.982 | 0.991 | 0.949 | 1.000  | 1.000 | X     | 0.977 | 0.136 | 0.056 | 0.075 | 0.846 | 0.095 | 0.014 | 0.247 | 0.222 | 0.060 | 0.055 | 0.138 | 0.930 | 0.085      | 0.890 | 0.879 | 0.000        | rs2808630    |            |
| rs2794520    | 0.144 | 0.762 | 1.000 | 0.942 | 1.000 | 0.938 | 0.930 | 0.982 | 0.991 | 0.925 | 1.000  | 1.000 | 1.000 | X     | 0.139 | 0.057 | 0.078 | 0.857 | 0.097 | 0.014 | 0.254 | 0.224 | 0.061 | 0.055 | 0.142 | 0.954 | 0.087      | 0.917 | 0.902 | 0.000        | rs3093069    |            |
| rs3093077    | 1.000 | 0.677 | 0.937 | 0.970 | 1.000 | 0.958 | 0.529 | 1.000 | 1.000 | 1.000 | 0.956  | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.113 | 0.607 | 0.119 | 0.680 | 0.036 | 0.972 | 0.088 | 0.122 | 0.123 | 0.654 | 0.127      | 0.624 | 0.124 | 0.138        | 0.001        | rs1205     |
| rs2808630    | 0.750 | 0.846 | 1.000 | 1.000 | 0.314 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.955  | 1.000 | 0.940 | 0.941 | 0.982 | 1.000 | X     | 0.064 | 0.897 | 0.018 | 0.689 | 0.005 | 0.077 | 0.077 | 0.604 | 0.076 | 0.979      | 0.068 | 0.083 | 0.001        | rs3093066    |            |
| rs3093069    | 0.501 | 0.895 | 0.945 | 0.945 | 1.000 | 0.966 | 0.975 | 1.000 | 0.926 | 0.955 | 1.000  | 0.931 | 0.966 | 1.000 | 1.000 | 1.000 | 0.923 | X     | 0.083 | 0.012 | 0.225 | 0.183 | 0.052 | 0.048 | 0.122 | 0.854 | 0.075      | 0.828 | 0.817 | 0.000        | rs1205       |            |
| rs3093066    | 0.837 | 0.844 | 1.000 | 0.942 | 1.000 | 0.945 | 0.915 | 0.931 | 0.966 | 0.971 | 1.000  | 0.982 | 1.000 | 1.000 | 1.000 | 1.000 | 0.993 | 1.000 | X     | 0.020 | 0.804 | 0.005 | 0.084 | 0.084 | 0.679 | 0.096 | 0.904      | 0.086 | 0.103 | 0.002        | rs3091244C/A |            |
| rs1417938    | 0.462 | 0.851 | 0.949 | 1.000 | 1.000 | 0.954 | 0.966 | 1.000 | 1.000 | 1.000 | 0.9729 | 1.000 | 1.000 | 1.000 | 0.582 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.000 | 0.148 | 0.005 | 0.006 | 0.029 | 0.014 | 0.018      | 0.013 | 0.015 | 0.013        | rs3091244C/T |            |
| rs3091244C/A | 1.000 | 0.704 | 0.962 | 1.000 | 0.580 | 0.867 | 0.654 | 1.000 | 1.000 | 0.973 | 1.000  | 0.966 | 1.000 | 1.000 | 0.991 | 1.000 | 0.987 | 1.000 | 1.000 | 1.000 | X     | 0.000 | 0.000 | 0.822 | 0.260 | 0.707 | 0.243      | 0.261 | 0.002 | rs3093062    |              |            |
| rs3091244C/T | 1.000 | 1.000 | 0.147 | 1.000 | 0.013 | 0.473 | 0.951 | 1.000 | 1.000 | 0.966 | 1.000  | 1.000 | 1.000 | 1.000 | 0.902 | 0.951 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 0.709 | 0.693 | 0.036 | 0.222 | 0.000 | 0.204      | 0.245 | 0.001 | rs3093058    |              |            |
| rs3093062    | 0.866 | 1.000 | 0.932 | 0.929 | 1.000 | 0.982 | 1.000 | 1.000 | 1.000 | 1.000 | 0.966  | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | X     | 1.000 | 0.124 | 0.060 | 0.076 | 0.055      | 0.063 | 0.001 | rs3093058    |              |            |
| rs3093058    | 0.778 | 1.000 | 0.930 | 0.926 | 1.000 | 0.981 | 1.000 | 1.000 | 1.000 | 1.000 | 0.972  | 0.956 | 1.000 | 1.000 | 0.811 | 1.000 | 1.000 | 1.000 | 0.992 | 1.000 | 0.989 | 1.000 | 1.000 | 1.000 | 1.000 | 0.989 | 1.000      | 1.000 | 0.001 | rs12084589   |              |            |
| rs12084589   | 1.000 | 0.716 | 0.946 | 0.971 | 0.735 | 0.947 | 0.667 | 1.000 | 1.000 | 0.972 | 0.956  | 1.000 | 0.968 | 0.984 | 0.965 | 1.000 | 0.982 | 1.000 | 0.992 | 1.000 | 0.989 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000      | 1.000 | 0.001 | rs2027471    |              |            |
| rs2027471    | 1.000 | 0.723 | 1.000 | 0.924 | 1.000 | 0.922 | 0.905 | 0.908 | 0.955 | 0.914 | 1.000  | 0.967 | 1.000 | 1.000 | 0.811 | 1.000 | 0.938 | 0.991 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000      | 1.000 | 0.001 | rs12079772   |              |            |
| rs12079772   | 0.505 | 0.903 | 0.951 | 1.000 | 1.000 | 0.983 | 0.988 | 1.000 | 1.000 | 0.955 | 0.914  | 1.000 | 0.968 | 0.984 | 0.965 | 1.000 | 0.993 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000      | 1.000 | 0.001 | rs1341665    |              |            |
| rs1341665    | 1.000 | 0.743 | 1.000 | 0.923 | 1.000 | 0.928 | 0.920 | 0.885 | 0.913 | 0.913 | 1.000  | 0.964 | 0.991 | 1.000 | 1.000 | 0.933 | 0.952 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000      | 1.000 | 0.001 | rs7443007    |              |            |
| rs7443007    | 1.000 | 0.780 | 0.945 | 0.863 | 0.179 | 0.927 | 0.913 | 0.926 | 0.963 | 0.911 | 1.000  | 0.975 | 0.976 | 0.968 | 1.000 | 0.946 | 1.000 | 1.000 | 1.000 | 1.000 | 0.973 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.992      | 1.000 | 0.991 | X            | 0.000        | rs12728740 |
| rs12728740   | 1.000 | 0.352 | 0.617 | 0.128 | 0.167 | 0.440 | 1.000 | 1.000 | 0.020 | 0.114 | 1.000  | 0.027 | 0.038 | 0.294 | 1.000 | 0.610 | 0.020 | 0.304 | 0.331 | 0.447 | 0.233 | 1.000 | 1.000 | 0.291 | 0.022 | 0.208 | 0.016      | 0.071 | X     |              |              |            |

\*  $r^2$  values at the top right and D' values at the bottom left.

## 5.4. African-American Genetic models

### 5.4.1: Primary Bayesian Analysis

It can be seen from **table 5.3** that a number of SNPs formed highly correlated clusters (a consequence of our deliberate policy to select a broad number of SNPs to encompass existing reports from different ethnic populations). The presence of highly correlated variables can, however, cause problems fitting and interpreting linear models, and may even give the impression no variable is associated with the outcome, even if a strong association is seen for each variable individually. SNPs that were correlated with  $r^2 > 0.8$  were therefore not simultaneously entered into any genetic model choice procedure. For these clusters of correlated SNPs a single tagSNP was selected for analysis. SNP *rs2808628* was selected to tag *rs2794520*, *rs2027471*, *rs1341665*, *rs7553007* and *rs1205* (mean pairwise  $r^2 = 0.89$ ); *rs3093066* to tag *rs3093069*, *rs12079772* and *rs16842493* (mean pairwise  $r^2 = 0.90$ ); and *rs3093062* to tag *rs3093058* ( $r^2 = 1.00$ ). A particular feature of this model choice analysis was that the procedure selected over alternative codings of the triallelic SNP *rs3091244*, with genotypes encoded both for increasing copies of the T allele and the A allele (a model selecting both codings would imply that the baseline C allele was most important). This analysis was therefore able to objectively resolve inconsistency in the literature over the likely functional allele at this SNP. For each genetic analysis age, BMI, sex and MS affection were specified as covariates (sex and MS affection were thought important to include despite the lack of evidence for an effect).

**Table 5.4** shows the results of the Bayesian model choice analysis incorporating all genetic markers (but not the 5 SNPs genotyped in only cohort 3a or 3b). Highest posterior probability was attached to a model incorporating *rs3091244*(T) alone, representing overwhelming

evidence for a genetic effect (Bayes factor >100 vs. null model (no SNPs)). A secondary model incorporating both *rs3091244*(T) + *rs12728740* had a smaller posterior probability. Evidence was weak for models using the alternative codings for *rs3091244* [*rs3091244*(A)  $P_{(M/D)} < 1 \times 10^{-4}$ ; *rs3091244*(A) + *rs3091244*(T)  $P_{(M/D)} = 0.003$ ].

**Table 5.4.** Bayesian models with >5% posterior probability.

| Model                                      | $P_{(M/D)}$ | $P_{\text{marg}}$ | Effect estimate (95% CI)                 |
|--------------------------------------------|-------------|-------------------|------------------------------------------|
| <i>rs3091244</i> (T)                       | 0.14        | 0.65              | 0.37 (0.21, 0.53)                        |
| <i>rs3091244</i> (T)+<br><i>rs12728740</i> | 0.07        | 0.65<br>0.20      | 0.45 (0.29, 0.63)<br>-0.32(-0.59, -0.05) |
| null                                       | <0.01       |                   |                                          |

$P_{(M/D)}$  - Posterior probability of the model given the data,  $P_{\text{marg}}$  – Marginal probability associated with each SNP in the model, Effect estimate (95% CI) – Effect estimate with 95% credible intervals.

#### 5.4.2: Secondary Bayesian Analyses

Additional analyses were performed to explore and substantiate these primary findings. Firstly, having identified *rs3091244* as the key SNP I investigated whether additive, dominant or recessive effects best explained this CRP association. This could be assessed directly by allowing the model algorithm to select over SNPs encoded for each of these modes of genetic effect. There was substantial evidence in favour of an additive model (Bayes factor 5.7 compared with dominant model). Secondly, the model choice analysis was re-run having specified only vague prior assumptions regarding the nature of the genetic model (see section 2.4.7). These data are presented in **table 5.5**. The highest posterior probability was again observed for an *rs3091244*(T) alone model, representing strong evidence for a genetic effect (Bayes factor 16.5 compared with the null model). The next

most likely model contained *rs3093062* alone, although the evidence to select the *rs3091244*(T) model over this secondary one was relatively weak (Bayes factor 2.75). No other model had a posterior probability >0.05. Notably the posterior probabilities were small for models using alternative codings of *rs3091244* [*rs3091244*(A)  $P_{(M/D)} < 1 \times 10^{-4}$ ; *rs3091244*(A) + *rs3091244*(T)  $P_{(M/D)} = 7.4 \times 10^{-4}$ ] or for the model incorporating *rs3091244*(T) + *rs3093062* [ $P_{(M/D)} = 0.0017$ ]. Compared with these models the evidence is overwhelmingly in favour of *rs3091244*(T) alone (Bayes factors >100).

**Table 5.5.** Models with > 5% posterior probability (vague priors)

| Model                | $P_{(M/D)}$ | $P_{\text{marg}}$ | Effect estimate (95% CI) |
|----------------------|-------------|-------------------|--------------------------|
| <i>rs3091244</i> (T) | 0.66        | 0.69              | 0.38 (0.22,0.54)         |
| <i>rs3093062</i>     | 0.24        | 0.26              | 0.46 (0.25,0.67)         |
| null                 | 0.04        |                   |                          |

$P_{(M/D)}$  - Posterior probability of the model given the data,  $P_{\text{marg}}$  – Marginal probability associated with each SNP in the model, Effect estimate (95% CI) – Effect estimate with 95% credible intervals.

Thirdly, as the total dataset consisted of individuals from two original cohorts I explored the possibility of cohort specific effects. Models were re-run with the addition of “cohort” as a fixed phenotype effect, but this made negligible difference to the overall results (data not shown). Models were also re-run with the inclusion of cohort\*genotype interactive terms. All the top models selected across these terms included genotype at *rs3091244*(T) +/- *rs12728740*, however there was some evidence to support a cohort\*genotype interaction with genotype exerting a smaller effect in cohort 1 (**Table 5.6**). Consistent with this was an analysis of each cohort separately, with *rs3091244* being the most significant SNP marginally in both cohorts, but with lower significance in cohort 1 ( $P_{\text{marg}} = 0.58$  in cohort 2 and  $P_{\text{marg}} = 0.10$  in cohort 1).

**Table 5.6** – Models with >5% posterior probability, (with genotype\*cohort interactions)

| Model                         | $P_{(M/D)}$ | $P_{\text{marg}}$ | Effect estimate (95% CI)* |
|-------------------------------|-------------|-------------------|---------------------------|
| <i>rs3091244</i> (T) +        |             | 1.00              | 0.58 (0.33, 0.84)         |
| <i>rs12728740</i> +           | 0.43        | 0.85              | -0.34 (-0.79, 0.00)       |
| <i>rs3091244</i> (T)*Cohort + |             | 0.81              | -0.28 (-0.64, 0.00)       |
| <i>rs12728740</i> *Cohort     |             | 0.52              | 0.15 (-0.18, 0.72)        |
| <i>rs3091244</i> (T) +        | 0.28        |                   |                           |
| <i>rs12728740</i> +           |             |                   |                           |
| <i>rs3091244</i> (T)*Cohort   |             |                   |                           |
| <i>rs3091244</i> (T) +        | 0.10        |                   |                           |
| <i>rs12728740</i> +           |             |                   |                           |

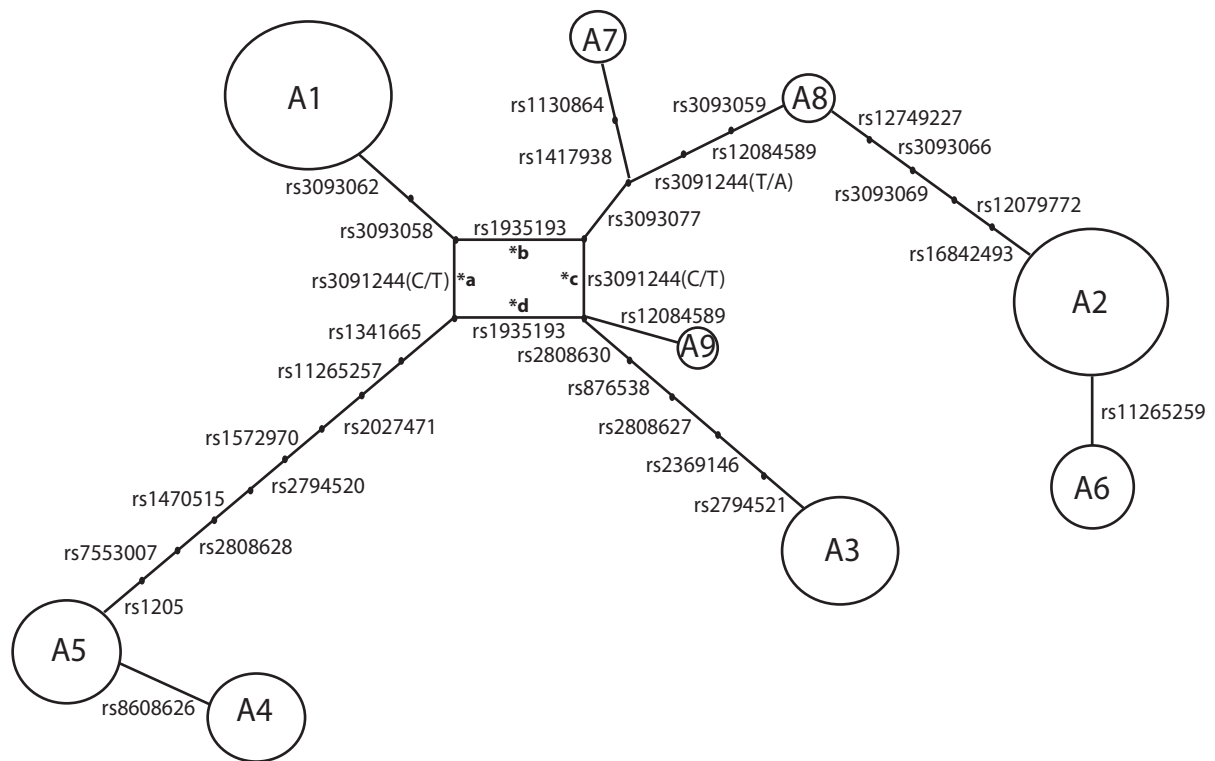
\*Effect estimates for each variable averaged over all models and expressed as the effect in cohort 1 (less effect) compared with cohort 2.

Finally, to ensure that the use of inferred phenotype data was not distorting the results I re-ran the analysis while excluding BMI as a covariate (BMI being the only important phenotypic effect with substantial missing data). In these models the significance attached to both the key SNPs actually increased [for *rs3091244*(T)  $P_{\text{marg}} = 0.90$  and for *rs12728740*  $P_{\text{marg}} = 0.26$ ]. The posterior probability of the two top models also increased [*rs3091244*(T) alone  $P_{(M/D)}=0.19$  and for *rs3091244*(T) + *rs12728740*  $P_{(M/D)}=0.11$ ].

#### 5.4.3: Haplotype Association Analysis

Admixture analysis predicted 10.8% of the observed haplotypes were of recent European origin. 20.2% of the cohort carried one European haplotype and 0.73% carried two (European haplotypes being defined as those carrying the minor allele of SNPs *rs1255606*, *rs11585798* or *rs17459580*). **Figure 5.2** presents a phylogenetic relationship between haplotypes, having excluded these individuals.

**Figure 5.2.** Cladogram of phylogenetic relationships between African-American haplotypes.



Each haplotype is represented by a circle whose size is roughly proportional to the population frequency. Connecting lines are annotated by the SNPs which distinguish one haplotype from another. Small dots along these lines represent haplotypes which are no longer observed but can be inferred to have existed in the population history. The loop at the centre of the tree represents uncertainty over the clade structure. Either *rs3091244* (C/T) or *rs1935193* may be the first mutational event, and consequently either *rs3091244* (C/T) or *rs1935193* must have mutated twice, from the wild type to the new allele, and then back to the wild type. These four explanations may be envisaged by removing any of the four branches labelled \*a, \*b, \*c or \*d.

As explained in the figure caption there are four possible resolutions of the tree, labelled \*a, \*b, \*c or \*d. Significant clade-based partitions were only observed for trees \*a and \*c (those where the *rs3091244* C/T mutation occurs only once, but the *rs1935193* mutation must have occurred twice, on different branches of the tree). The most significant partition was defined by the *rs3091244* C/T mutation, segregating haplotypes A1, A2, A6, A7 and A8 as clade one and haplotypes A3, A4, A5 and A9 as clade two (For tree \*a,  $F=8.91$ ,  $P=0.006$ ). Clade one was associated with higher CRP ( $\beta=0.312$  95%CI 0.146, 0.491), although the overall proportion of CRP variance explained by this partition was quite small (5.20%). A weakly significant effect was also observed for a partition defined by *rs3093062/rs3093058*, separating haplotype A1 from all the others (for tree \*a  $F=6.385$ ,  $P=0.038$ ), although this effect may be largely due to correlation with the *rs3091244* C/T partition. The A1 haplotype was associated with higher CRP level than all the rest ( $\beta=0.448$  95%CI 0.241, 0.655). No significant secondary partition was detected, although it should be noted the power to detect such an effect is dramatically reduced because of the smaller sample size consequent on the primary partition.

## 5.5 CRP Genetics in an African-American Cohort – Discussion

### 5.5.1 Summary of Findings

The analysis presented in this chapter attempts to rationalise the multiple *CRP* SNPs found associated with serum basal CRP both in the analyses presented in chapters 3 and 4 and elsewhere in the published literature. I aimed to do this using a cohort of African-American origin, where the relatively lower LD would potentially narrow the interval over which a genetic association is observed, and by employing a formal model choice analysis, backed up by a clade-based haplotype analysis to highlight the most likely functional SNP at *CRP*.

This analysis provides good evidence that a single key variant, *rs3091244* (or an unmeasured variant in strong LD with *rs3091244*), regulates CRP expression in African Americans. In particular the T allele acting in an additive manner explained most of the genetic component of variation in serum CRP levels. Consistent with previous reports this effect was modest, explaining only 5.20% of CRP variance.

A model incorporating both *rs3091244* and *rs12728740* should also be kept in consideration. The posterior probability of this two-SNP model was smaller than the model with *rs3091244* alone but, because there are more possible two-SNP combinations, the Bayes factor gave it slightly more weight (Bayes factor = 6 in favour of the *rs3091244* + *rs12728740* model). The incorporation of *rs12728740* into a model is interesting. Although not originally selected to tag the admixed European genome it became apparent that 96% of the minor alleles of *rs12728740* were segregating with haplotypes of inferred European origin. In addition it was predicted to lie outside the main *CRP* haplotype block in our population. By reference to the HapMap CEPH panel it is clear that in Europeans *rs1272870* is in much stronger LD with the main *CRP* SNPs, in particular with *rs1417938* ( $r^2=0.857$ ), which tags the commonest European haplotype, and has been associated with higher *CRP* production in existing European studies . This does not downplay the importance of *rs3091244* as the key regulator of *CRP*, but it does provide some evidence for a separate functional effect of recent European origin.

We failed to demonstrate any role for SNPs at the *IL6* and *IL1B* loci.

### 5.5.2. A comparison with existing studies

Five studies to date have considered the role of *CRP* genetics on serum basal CRP in the African American population. The first concentrated on CRP levels associated with *rs3091244* and *rs3093062* only (Szalai *et al* 2005c). The authors' key argument hinged on an interaction between *rs3091244*(T) and *rs3093062* such that haplotype *rs3091244*(T)/*rs3093062*(G) was associated with high CRP, while haplotype *rs3091244*(T)/*rs3093062*(A), the equivalent to my haplotype A1, was associated with low CRP. There are therefore key differences between this paper and our data; we found little evidence for a role for both *rs3091244* and *rs3093062*. Furthermore we found the *rs3093062*(A) haplotype to be associated with high rather than low CRP.

Shortly after this initial report, data from the biracial CARDIA cohort was published (Carlson *et al* 2005). Tag SNPs were selected on the basis of sequencing a small number of individuals and hence failed to capture all the variation we detected. The authors also analysed both European and African ethnicities together (albeit with a "race" term in their regression models), which may be a suboptimal approach given the different SNP and haplotype frequencies between groups. Two subsequent studies in different cohorts used the same combination of SNPs, but with a separate analysis of African American participants (Crawford *et al* 2006; Lange *et al* 2006b). All these studies generated results broadly similar to ours, with haplotypes tagged by the C and T alleles of *rs3091244* being associated with low and high levels of CRP respectively. Notably haplotype *rs3091244*(T)/*rs3093058*(T) (equivalent to my haplotype A1) was associated with highest overall CRP expression, as suggested by the data presented in this chapter.

The final study to report on CRP associations in an African-American cohort did not genotype *rs3091244*, making direct comparison with my data difficult (Lee *et al* 2009). Nonetheless a comparison of the haplotypes reported in this paper with the *rs3091244* carrying haplotypes in my data suggests a similar pattern of haplotype associations with higher and lower CRP as I observed.

### 5.5.3. *Key study features and potential limitations*

It should be recognised that, once again, the *CRP* analysis presented in this chapter represents a *post-hoc* analysis of data on a cohort recruited primarily for other purposes. The structure of this cohort offers some advantages compared with the cohorts presented in chapters 3 and 4, but there also some clear limitations which need to be discussed in detail.

As a population based group of unrelated individuals, the nature of this cohort allowed for a dramatic simplification of analysis, using linear regression models with no requirement to adjust for the familial clustering of data. Secondly, none of the individuals was recruited specifically because they had a disease known to influence *CRP* expression. Multiple sclerosis is not generally associated with elevated inflammatory markers unless in a particular active phase of disease (confirmed in our analysis), while the coincidental recruitment of individuals with cardiovascular disease simply added to random measurement error (Giovannoni *et al* 2001; Soilu-Hanninen *et al* 2005).

The two key difficulties arising from the analysis of this group of patients concerned, firstly, the use of a cohort derived from an amalgamation of two separate collections of African-American samples and, secondly, the ethnic background of the cohort itself. The first

problem was therefore unique to this study, while the second is a generic problem with all studies in admixed populations.

Ideally if cohorts are to be merged for study such as this then it is important that both are heterogeneous groups, and that there are no significant differences in their baseline phenotypic characteristics. It can be seen from the cohort summary in section 2.1.3 that the patients in these two source cohorts differed both in their phenotypic characteristics, and in the degree of available phenotypic data. In particular there were considerable age differences between the two groups (mean 78.9 years vs. 43.5 years), while there was minimal phenotypic information available for the younger group compared with the older. To compound the potential problems was the fact that the biological assays were performed at different sites and, in the case of the CRP quantification, by different assays (albeit highly validated). This is a situation that has been noted as a frequent source of bias in genetic studies (Clayton *et al* 2005). We explored this problem using a “cohort” term as a fixed phenotype effect, using genotype\*cohort interaction terms and, in addition, we examined each cohort separately. These secondary analyses suggested that there were some quantitative differences between cohorts in terms of the magnitude of SNP effect, but there were no qualitative differences in terms of which were the key SNPs, or SNP combinations, associated with CRP levels. The likely explanation for the less extreme SNP effects in cohort 1 (older individuals) is that with higher disease burden (both documented and occult) a greater proportion of the observed CRP level is due to stimulated acute-phase expression rather than truly basal CRP. This increase in “background noise” may decrease our power to detect *CRP* associations. Although these secondary analyses provide some reassurance regarding the validity of the results presented in this chapter, the construction of the study cohort should still be recognised as shortcoming.

A further, partly related, problem concerned my use of inferred phenotype data. It is well recognised from all the early *CRP* genetics papers that a significant proportion of the variability in basal CRP can be attributed to measureable phenotypic characteristics, notably those considered typical cardiovascular disease risk factors such as BMI, blood pressure, smoking status etc. Our inability to control for basic phenotypic parameters was notable in the analyses presented in chapters 3 and 4, and may reduce the power to detect true genetic effects. I was therefore keen to maximise the use of available phenotype data, without omitting genetic data from individuals on whom phenotypes were not available: this was done through the inference of missing phenotype data. The technical details are provided in section 2.5.3 but to summarise in lay terms an inferred phenotypic value is “randomly” assigned for each individual where data is missing, such that the overall mean and distribution of these inferred values matches that seen for the true measured phenotypic values. Theoretically the error introduced by inserting these missing values will be the equivalent of random measurement error (see section 3.7.5), and will not introduce systematic bias. Clearly, however, the inference process is not perfect, in particular if the amount of missing data that is inferred is large. In practice, in the analysis presented here, the only key variable for which levels of missing data were great enough for this to be an issue was BMI. To assess the potential effect of the inferred data I performed a secondary analysis, with the complete omission of BMI. The greatest significance was still attached to a model containing the same key SNPs; in fact the significance of the top models increased somewhat. While recognising the limitations of having incomplete BMI data, I have no evidence to suggest that if these additional data were available it would change the overall genetic conclusions.

The African-American ancestry of the cohort itself also presented issues that require further discussion. I have already outlined the problems that can be associated with genetic association studies in admixed populations (section 1.5.4). At the level of a single locus

genetic admixture may also complicate analyses by distorting the inferred haplotype structure if the pattern of LD differs between the background and the admixed genetic elements. It may therefore affect our ability to detect an indirect genetic association. I did not have the resources to type an extensive panel of genome-wide ancestry informative markers; indeed such panels were not publically available at the time this project was planned. Instead I adopted the approach of at least attempting to highlight *CRP* haplotypes in this population that may have a European origin. Three SNPs were identified from the HapMap dataset that were predicted to be monomorphic in African Populations but polymorphic in white Europeans. Individuals in the study cohort were defined as having haplotypes of recent European origin if they carried the minor allele of one or more of these SNPs. These individuals were excluded from further analysis. This is therefore a rather crude approach as clearly not all European haplotypes will be identified this way, while other individuals may have unjustifiably been excluded from analysis: nonetheless even this approach is likely to be more effective than no consideration of admixture at all.

I should also briefly discuss here the clade-based haplotype analysis. A clade-based approach offers distinct advantages over a conventional haplotype analysis, particularly in populations such as this one, where there are quite a number of haplotypes, each of moderate or low frequency, at a particular locus. A conventional analysis would be penalised both in terms of power (because of the low numbers of these rarer haplotypes) and in terms of multiple hypothesis testing (if the levels of CRP associated with each haplotype is compared with the rest). With a cladistic approach, rather than each haplotype being analysed in isolation, haplotypes are nested into clades defined by their proposed evolutionary relationship. The assumption is that the functional effect of a mutation will be found across the whole clade of haplotypes on which it is found. This approach increases the power of an analysis by looking at larger groups of haplotypes, and also limits unnecessary multiple comparisons by

restricting tests only to evolutionarily relevant clades (Templeton *et al* 2005). The key premise of the haplotype approach is that the genetic locus has evolved by single mutational steps, with little or no ancestral recombination, and for a locus such as *CRP* where LD is strong this seems a reasonable assumption to make. Nonetheless it is slightly unsatisfactory that the African-American haplotype tree could not be determined with certainty, and that all of the possible constructions for this tree depended on one SNP or another mutating from one allele to another and then back again as a distinct separate event. As discussed in section 8.4.4, the mutational rate in the human genome is so low that this would have to be a highly improbable event. The question must therefore arise whether the double switching of alleles truly occurred, or whether the initial mutation was followed by a recombination event in the ancient history of the *CRP* locus.

#### 5.5.4. *Conclusions*

The analyses presented in this chapter point to the triallelic promoter variant *rs3091244* (or an untyped variant in strong LD with it) as being the most important SNP regulating *CRP* expression in this African-American cohort. This was highlighted using both a Bayesian model choice approach and a classical haplotype cladistic analysis.

Identifying the genetic determinants of *CRP* expression remains an important question and the use of ethnically diverse cohorts may make this possible.

## **CHAPTER 6**

# **GENETICS OF ACUTE-PHASE CRP EXPRESSION**

## **6.1: The genetics of acute-phase CRP Expression – Study aims**

The analyses presented in chapters three to five provide evidence that genetic variation influences serum CRP level in the setting of the low-level, basal CRP seen in people without active inflammatory disease. However, in the presence of infection or inflammation serum CRP levels increase dramatically due to accelerated hepatocellular transcription. The quantification of this acute-phase CRP is commonly used in routine clinical practice to diagnose and monitor the activity of inflammatory diseases. As I outline in section 1.3.3, very few studies address the question of whether genetic variation influences the magnitude of this acute-phase CRP rise. Of these studies none are conducted in cohorts of individuals with chronically elevated CRP due to persistent inflammatory disease activity.

The aim of the work presented in this chapter was to perform a pilot study to investigate whether the genetic variants I have highlighted in chapters three to six are also associated with the magnitude of acute-phase serum CRP in patients with rheumatoid arthritis, a well characterised inflammatory disease.

## **6.2: Rheumatoid Arthritis Cohort – Demographics and CRP distribution**

As I have already outlined in section 2.1.4. the cohort employed for the analyses presented in this chapter consisted of 281 UK white European rheumatoid arthritis cases, unselected for their level of disease activity. Mean age was 61.2 (s.d. 12.1) and 79% were female. Median serum CRP was 11 mg/L (range, 1 to 195; 5th, 95th percentiles, 5.0, 74.2). The distribution of CRP values had a similar positive skew to that observed with basal CRP, although obviously right-shifted towards higher CRP values. In view of this, log-transformed CRP was again used for all analyses. Non-disease related phenotype data were available and included age, sex and smoking status. An initial linear regression model (**table 6.1**) demonstrated a

negative association between age and acute-phase CRP, but no association with age or smoking status.

**Table 6.1.** Effect of age sex and smoking on acute-phase CRP

| Variable       | $\beta$ -coefficient | standard error | t      | P-value |
|----------------|----------------------|----------------|--------|---------|
| Age            | -0.005               | 0.002          | -2.431 | 0.025   |
| Sex (Female)   | -0.103               | 0.062          | -1.648 | 0.100   |
| Current smoker | 0.033                | 0.060          | 0.553  | 0.581   |

In view of the findings from this preliminary phenotype regression model, age was retained as a covariate for all subsequent genetic association analyses.

### 6.3: CRP genetic architecture.

SNPs were selected for genotyping on this basis that they tagged the main *CRP* haplotypes already defined in the large UK white European cohort presented in chapter 3. Unfortunately, due to financial constraints, only SNPs that could be genotyped simultaneously in a single pool by the Sequenom method were selected, and this precluded the genotyping of *rs3091244* which I will argue later is the most likely functional SNP in the context of basal CRP expression. It can be seen with reference to **figure 3.1** that the T allele of *rs3091244* tags the common white European haplotype H1, and we can therefore infer that effects attributable to the H1 haplotype in this analysis would also be shared with the T allele of *rs3091244*.

Detailed marker information is presented in **table 6.2**. Inferred *CRP* haplotypes are presented in **table 6.3**.

**Table 6.2.** Typed SNPs demonstrating alleles and rare allele frequencies

| rs number         | SNP | Location            | Rare allele frequency | HWP <sup>a</sup> | missing data (%) |
|-------------------|-----|---------------------|-----------------------|------------------|------------------|
| <i>rs2808632</i>  | A>C | 5' flank            | 0.28                  | 0.12             | 0.70             |
| <i>rs3093059</i>  | T>C | 5' flank            | 0.06                  | 0.48             | 0.00             |
| <i>rs1800947</i>  | G>C | exon 2 (synonymous) | 0.08                  | 0.43             | 3.51             |
| <i>rs1205</i>     | G>A | 3' flank            | 0.35                  | 0.48             | 0.00             |
| <i>rs876538</i>   | A>G | 3' flank            | 0.21                  | 0.31             | 0.00             |
| <i>rs11265257</i> | G>A | 3' flank            | 0.39                  | 0.45             | 0.35             |

<sup>a</sup>Hardy-Weinburg probability *P*-value calculated by  $\chi^2$

All SNPs passed basic quality control checks. As expected, marker frequencies in this white European population accorded well with those observed for the UK SLE cohort (refer to **table 3.2**), and the frequency of common inferred *CRP* haplotypes was also similar (refer to **figure 3.1**).

**Table 6.3.** Frequency of common *CRP* haplotypes

| Haplotype | <i>rs11265257</i> | <i>rs876538</i> | <i>rs1205</i> | <i>rs1800947</i> | <i>rs3093059</i> | <i>rs2808632</i> | Frequency |
|-----------|-------------------|-----------------|---------------|------------------|------------------|------------------|-----------|
| H1        | G                 | G               | G             | G                | T                | A                | 0.294     |
| H2        | A                 | G               | A             | G                | T                | A                | 0.271     |
| H3        | G                 | A               | G             | G                | T                | C                | 0.221     |
| H4        | G                 | G               | G             | G                | C                | A                | 0.078     |
| H5        | A                 | G               | A             | C                | T                | A                | 0.062     |
| H6        | A                 | G               | G             | G                | T                | C                | 0.048     |
| H7        | G                 | G               | G             | G                | T                | C                | 0.019     |

## 6.4: Genetic associations with acute-phase CRP

Each genotyped *CRP* SNP was evaluated for association with acute-phase serum CRP, assuming an additive genetic effect. **Table 6.4** demonstrates the regression parameters and significance for each SNP. Even with no consideration for the level of disease activity in these patients, a significant association was seen between genotype at *rs1205* and serum acute-phase CRP. This association remained significant after applying a partial Bonferroni correction adjusted for the correlation between SNPs (6 tests, mean correlation  $r=0.416$ ). A borderline association was also observed between CRP and SNPs *rs11265257* and *rs1800947*.

**Table 6.4.** Single SNP associations with CRP

| Marker            | $\beta$ -coefficient | standard error | t      | Regression*<br>P-value |
|-------------------|----------------------|----------------|--------|------------------------|
| <i>rs2808632</i>  | -0.002               | 0.041          | -0.047 | 0.963                  |
| <i>rs3093059</i>  | 0.094                | 0.067          | 1.398  | 0.163                  |
| <i>rs1800947</i>  | -0.133               | 0.062          | -2.133 | 0.034                  |
| <i>rs1205</i>     | -0.090               | 0.037          | -2.443 | 0.015                  |
| <i>rs976538</i>   | -0.020               | 0.042          | -0.471 | 0.638                  |
| <i>rs11265257</i> | -0.083               | 0.036          | -2.310 | 0.022                  |

\* Unadjusted P-value. An individual P-value  $<0.018$  is required to achieve overall significance at the  $P=0.05$  level after applying a Bonferroni correction adjusted for the correlation between markers).

In a similar manner to the basal CRP analysis in chapter 3, I also modelled the association of acute-phase CRP with *CRP* haplotypes, with reference to the commonest haplotype, H1.

Haplotypes with a frequency  $<5\%$  were pooled for this analysis. Full linear regression parameters for each of the covariates in the model are presented in **table 6.5**. Compared with H1 the lowest CRP was observed for haplotype H5. Lower levels of serum CRP were also seen with haplotype H2, and the highest levels with haplotype H4.

**Table 6.5.** The association of CRP with *CRP* haplotypes

| Marker    | $\beta$ -coefficient | standard error | t      | Regression*<br>P-value |
|-----------|----------------------|----------------|--------|------------------------|
| Intercept | 1.470                | 0.142          |        |                        |
| Age       | -0.004               | 0.002          | -2.061 | 0.040                  |
| H1        | ref                  |                |        |                        |
| H2        | -0.084               | 0.045          | -1.836 | 0.068                  |
| H3        | -0.061               | 0.050          | -1.221 | 0.223                  |
| H4        | 0.044                | 0.073          | 0.545  | 0.568                  |
| H5        | -0.183               | 0.068          | -2.744 | 0.006                  |
| Hx        | -0.005               | 0.070          | -0.074 | 0.941                  |

Haplotypes modelled with reference to haplotype H1. Hx represents pooled rare haplotypes.

The magnitude of difference in acute-phase CRP between carriers of different *CRP* haplotypes can be appreciated by using the regression  $\beta$ -coefficient to back-convert the transformed CRP data to untransformed values. Using this approach, a representative 40 year old patient with rheumatoid arthritis would be predicted to have a geometric mean (95% C.I.) CRP of 20.4mg/L (10.7, 38.9) if they carried two H1 haplotypes but only 8.79mg/L (4.7, 16.0) if they carried two H5 haplotypes. Carriers of two H4 haplotypes would have an even higher CRP of 25.00mg/L (12.7, 47.2). This difference between CRP values between H4 and H5 can also be demonstrated as significant ( $P=0.013$ ) although both are of low frequency, limiting the power of this comparison.

Having demonstrated an association between *CRP* genotype and acute-phase serum CRP, I extended this analysis to include an adjustment for the level of disease activity. This was attempted in two ways, firstly by the use of the incorporation of the ESR, an alternative marker of inflammation which is not, itself, known to be directly influenced by the level of CRP (or *CRP* genetic variants), and secondly by the incorporation of the disease activity (DAS) score for each individual as a covariate in the regression model. The DAS score is a

commonly used composite measure of disease activity in rheumatoid arthritis, and is widely used as a marker of clinical improvement in rheumatoid arthritis drug trials (Prevoo *et al* 1995). The DAS is calculated by an algorithm which incorporates the tender and swollen joint count, a physician and patient “global score” of disease activity, and the ESR, using the following equation:

$$\text{DAS} = (0.56 \times \sqrt{\text{tender joint count}}) + (0.28 \times \sqrt{\text{swollen joint count}}) + (0.7 \times \ln \text{ESR}) + (0.014 \times \text{global score})$$

The DAS therefore incorporates both objective and subjective measures of disease activity. My expectation was that, by incorporating either the ESR or the DAS to control for the level of underlying inflammatory stimulus driving acute-phase CRP production, it would become easier to isolate the effect of genetics on acute-phase serum CRP level.

The association between serum CRP and *CRP* genotype at each of the typed SNPs was again assessed in a linear regression model incorporating age and either ESR or DAS as covariates. Results are presented in **table 6.6**. It can be seen from these analyses that the inclusion of ESR as a covariate results in a marked increase in the significance of the association between genotype and acute-phase CRP for both *rs1205* and *rs11265257*, while the association with the rarer *rs1800947* remains borderline. Interestingly, the inclusion of DAS as a covariate has much less of an effect on the observed associations, suggesting the DAS quantifies the level of acute-phase stimulus less accurately than ESR alone in active rheumatoid arthritis. In all analyses the same pattern of SNP effects was seen, with the lowest overall acute-phase CRP observed with the minor allele of *rs1800947*, and the highest acute-phase CRP observed with the minor allele of *rs3093059*. Neither of these individual SNPs reached significance, however, presumably because both have low minor allele frequency.

**Table 6.6.** Single SNP associations with CRP, controlling for ESR or DAS

| Marker            | $\beta$ -coefficient | standard error | t      | Regression*<br>P-value |
|-------------------|----------------------|----------------|--------|------------------------|
| ESR adjusted      |                      |                |        |                        |
| <i>rs2808632</i>  | -0.028               | 0.033          | -0.849 | 0.397                  |
| <i>rs3093059</i>  | 0.130                | 0.055          | 2.356  | 0.019                  |
| <i>rs1800947</i>  | -0.097               | 0.050          | -1.918 | 0.056                  |
| <i>rs1205</i>     | -0.085               | 0.030          | -2.833 | 0.005                  |
| <i>rs976538</i>   | -0.017               | 0.035          | -0.499 | 0.618                  |
| <i>rs11265257</i> | -0.093               | 0.029          | -3.183 | 0.002                  |
| DAS adjusted      |                      |                |        |                        |
| <i>rs2808632</i>  | -0.017               | 0.037          | -0.446 | 0.656                  |
| <i>rs3093059</i>  | 0.094                | 0.062          | 1.505  | 0.134                  |
| <i>rs1800947</i>  | -0.093               | 0.058          | -1.609 | 0.109                  |
| <i>rs1205</i>     | -0.084               | 0.034          | -2.489 | 0.013                  |
| <i>rs976538</i>   | -0.034               | 0.039          | 0.877  | 0.381                  |
| <i>rs11265257</i> | -0.079               | 0.033          | -2.380 | 0.018                  |

**Table 6.7** and **table 6.8** demonstrate the haplotype association when either ESR or DAS are included in analysis, with ESR again appearing to provide a better control for the level of acute-phase response.

**Table 6.7.** The association of acute-phase CRP with *CRP* haplotype, controlling for ESR.

| Marker    | $\beta$ -coefficient | standard error | t      | Regression*<br>P-value |
|-----------|----------------------|----------------|--------|------------------------|
| Intercept | 1.266                | 0.116          |        |                        |
| ESR       | 0.009                | 0.001          | 11.959 | 0.005                  |
| Age       | -0.005               | 0.002          | -3.126 | 0.002                  |
| H1        | ref                  |                |        |                        |
| H2        | -0.092               | 0.037          | -2.505 | 0.013                  |
| H3        | -0.053               | 0.040          | -1.317 | 0.189                  |
| H4        | 0.066                | 0.059          | 1.120  | 0.264                  |
| H5        | -0.168               | 0.055          | -3.061 | 0.002                  |
| Hx        | -0.101               | 0.057          | -1.766 | 0.079                  |

Haplotypes modelled with reference to haplotype H1. Hx represents pooled rare haplotypes.

**Table 6.8.** The association of acute-phase CRP with *CRP* haplotype, controlling for DAS.

| Marker    | $\beta$ -coefficient | standard error | t      | Regression*<br>P-value |
|-----------|----------------------|----------------|--------|------------------------|
| Intercept | 1.019                | 0.147          |        |                        |
| DAS       | 0.103                | 0.015          | 7.033  | <0.001                 |
| Age       | -0.004               | 0.002          | -2.157 | 0.032                  |
| H1        | ref                  |                |        |                        |
| H2        | -0.093               | 0.042          | -2.216 | 0.028                  |
| H3        | -0.074               | 0.046          | -1.598 | 0.111                  |
| H4        | 0.026                | 0.067          | 0.391  | 0.696                  |
| H5        | -0.160               | 0.063          | -2.542 | 0.012                  |
| Hx        | -0.038               | 0.064          | -0.599 | 0.549                  |

Haplotypes modelled with reference to haplotype H1. Hx represents pooled rare haplotypes.

## 6.5: The genetics of acute-phase CRP Expression – Discussion

### 6.5.1: Summary of findings

To my knowledge this study is the first to consider the association between *CRP* genotype and acute-phase serum CRP in the context of a chronic inflammatory disease. The key study finding is that it is possible to identify *cis*-acting genetic variants influencing serum CRP, even in this relatively small cohort of individuals. Interestingly, these associations can be demonstrated even when making no attempt to adjust for the underlying level of inflammatory stimulus, although there is a clear increase in the level of significance when such an adjustment is made through the inclusion of ESR as a covariate. The magnitude of genetic effect on acute-phase CRP appears considerable, with a threefold difference in geometric mean CRP predicted for carriers of two “lowest-producing” haplotypes (H5) compared with two “highest-producing” haplotypes (H4).

### 6.5.2: A comparison with existing studies

As I have already outlined in section 1.3.3, only a limited number of studies have looked for genetic associations within the acute-phase CRP range. The study of D'Aiuto *et al*, in patients with periodontitis, suggested carriers of the minor allele of *rs1130864* had higher acute-phase CRP levels (D'Aiuto *et al* 2005). The genetic data I present in chapter 3 demonstrates that *rs1130864* tags the commonest *CRP* haplotype (which I have termed H1). In this acute-phase study I did not directly genotype *rs1130864*, or any equivalent SNP, which is unique to H1, although I could use alternative SNPs to infer H1 haplotypes. In my haplotype analysis (**table 6.5**, **table 6.7** and **table 6.8**) I observed a similar pattern, with the H1 haplotype generally associated with higher levels of CRP and significantly higher than the H5 haplotype. My findings are also similar to those found by Suk *et al*, looking at the acute-phase CRP occurring after myocardial infarction (Suk *et al* 2006). They found lowest CRP levels in carriers of the rare alleles of *rs1800947* and *rs1205* and a haplotype containing both of these (equivalent to my haplotype H5). Highest CRP was observed in carriers of the A allele of *rs3091244* (which is in strong LD with the *rs1130864* variant identified by D'Aiuto) and the rare allele of *rs3093059*, with dramatically higher levels in individuals homozygous for these variants. Finally, the recent study by Perry *et al* also found lowest levels of acute-phase CRP rise following coronary artery bypass surgery in carriers of the rare allele of *rs1800947* (Perry *et al* 2009). It is also notable that the general pattern of association between higher and lower serum CRP and specific genetic variants would appear to be preserved in comparison with my earlier studies of basal CRP.

#### 6.4.3: Key study features and potential limitations

It is important to recognise that the analysis presented in this chapter was intended as a pilot study to test the concept that we can demonstrate genetic associations with acute-phase CRP. The study was therefore limited in its scope. The cohort of 281 individuals is relatively small for a genetic study although, without good existing data to suggest what magnitude of genetic effect we would expect to observe in this setting, it was difficult to perform a meaningful power calculation. Current basal CRP studies suggest individual *CRP* variants have a relatively modest effect on serum CRP (see section 8.4.6). Whether or not this would prove to be the case for acute-phase CRP was uncertain, although the report of Suk *et al* suggested genetic effects on acute-phase CRP could be considerable. A relatively small number of SNPs were genotyped, although previous work suggested that these were adequate to capture common *CRP* variation in a white European population. I did not genotype SNPs at any *trans*-acting loci although this would clearly be interesting in future studies, particularly variants within acute-phase cytokine genes. Although the MALDI-TOF genotyping used for all the studies presented in this thesis appears to generate consistent results it could be argued that, in population data at least, a check of genotyping accuracy should be made by duplicate genotyping a proportion of individuals to assess for any inter-assay errors. This was not done in this study.

A genetic study in a cohort of individuals with overtly “acute-phase” levels of CRP is always going to face methodological problems due to the considerable amount of random error introduced into an analysis by the inability to accurately quantify the level of inflammatory stimulus that drives the rise in serum CRP. Our ability to determine the component of CRP variability that is due to *CRP* genetics may therefore be swamped by the variability in CRP levels due to unquantified driving stimuli. There are a number of potential study designs to

help limit this problem. One approach would be to quantify the magnitude of CRP rise at a set time following a standardised inflammatory stimulus. Using this approach in healthy volunteers would be ethically inappropriate because the degree of tissue damage required to induce a notable acute-phase response would be relatively large. Alternatively one could consider the CRP rise following a standard surgical procedure (the approach taken by Perry *et al* 2009) although it would appear difficult to accurately quantify the nature and magnitude of tissue trauma from one operation to another. The approach of quantifying CRP after the acute inflammatory stimulus of a myocardial infarction has yielded some positive findings as we have seen above (Suk *et al* 2005). It would have been interesting to see how the significance of the genetic association changed if these studies were able to additionally quantify the degree of myocardial damage, perhaps by including in their regression models additional covariate terms for the magnitude of creatinine kinase rise (an enzyme released in myocardial necrosis), or perhaps an echocardiographic quantification of infarct size.

A second approach would be to control for the degree of inflammatory stimulus through the measurement of an alternative acute-phase reactant. Ideally we would probably use a single biochemical marker such as serum amyloid A or interleukin-6, whose levels also reflect the degree of inflammatory response but are not directly influenced by *CRP* genetic variation. The level of these alternative markers could then be included as a covariate in the regression model of genetic variation on serum CRP levels. This may not provide the perfect adjustment because the time course and dynamic range of elevation of these proteins may vary compared with CRP. Caution would also need to be applied, particularly when considering *trans*-acting genetic effects, as these may be shared across the spectrum of acute-phase proteins, and hence inclusion of a second acute-phase protein variable may act as a source of confounding. Despite these caveats it is likely this approach would offer a significant advantage over the

alternative of no control at all and may well increase the power of such studies to identify *cis*-acting *CRP* effects.

In the study presented in this chapter we did not have available data for a single, alternative acute-phase reagent, but we did have concurrent erythrocyte sedimentation rate (ESR) data. It was formally recognised in the early years of the twentieth century that inflammatory or infectious states influenced rate of sedimentation of red blood cells (Westergren 1921). The factors contributing to the rise in ESR in inflammatory states are complex and poorly understood. It appears primarily related to changes in the intrinsic property of the red cell membrane, notably the membrane electrical potential, which influences its “stickiness” and tendency to clump together in rouleaux. To a lesser extent ESR may be influenced by the concentration of large plasma proteins such as immunoglobulin and fibrinogen, but there is no evidence that the concentration of small proteins such as CRP has any influence (Ropes *et al* 1939; Zlonis 1993). Although the biology may remain unclear, the use of ESR fulfils the requirement for an alternative marker of acute-phase response to include as a covariate in the *CRP* analysis.

The final approach, also adopted in the study presented here, is to consider the level of acute-phase serum CRP measured in individuals with chronic active inflammation, and to quantify the magnitude of the inflammatory stimulus through the use of a well validated clinical disease activity index that does not rely on serum CRP level itself as a scoring parameter. Rheumatoid arthritis is one such condition, with disease activity commonly quantified through the DAS-28 score. There are also drawbacks to this approach. Firstly, as I discussed when considering the inclusion of a specific alternative marker of acute-phase response, we could argue that shared genetic influences may determine both the level of acute-phase CRP and one or more of the parameters that influence the DAS score. The traditional DAS score

incorporates ESR as its objective measure of inflammation and, as stated above there is no evidence to suggest CRP concentration directly influences ESR. I would therefore think it highly unlikely that this is a source of confounding when looking at *CRP cis*-effects.

The current data suggest that DAS score may actually be a rather poor summary of the level of acute-phase inflammatory stimulus, and not as good as simply measuring ESR on its own. Indeed it has previously been observed that the correlation between DAS score and markers of inflammation is poor (Keenan *et al* 2008). There are a number of possible reasons for this. Firstly the DAS makes no distinction between very large joints such as the knee or very small joints such as the 5<sup>th</sup> proximal interphalangeal joint, when the count of swollen and tender joints is made. Secondly, the inclusion in the score of “global assessments” by patient and physician introduce a subjective element that may be influenced by general factors such as fatigue, which do not specifically reflect the degree of overall inflammatory stimulus. If the relationship between DAS score and acute-phase response is poor then it is not going to fulfil the requirement of reducing random noise due to varying levels of inflammatory stimulus. The observation in the data presented here was that it was easier to detect *CRP* genetic effects by including ESR as a covariate compared with DAS as a whole, suggesting DAS is a less effective measure of inflammation *per se* in rheumatoid arthritis. One could envisage a scoring system that might fulfil this role better (for example one could perhaps count the swollen of tender joints and multiply these scores by a factor representing joint surface area), but since these are not in standard clinical usage they would require prospective collection in a new cohort.

In section 3.7.5 I discussed, in relation to the study of basal CRP the problem of intra-individual variation in CRP concentration and how this may limit our power to detect true genetic associations. For basal CRP it is likely this effect is small with available evidence

suggesting remarkable longitudinal stability in basal CRP. For a study of acute-phase CRP in a chronic setting it is likely that this intra-individual variation will be large due to underlying disease activity, unless highly accurate correction for the level of background inflammatory stimulus can be made and incorporated into the model. In this setting it is likely that a single snap-shot CRP level is going to be associated with considerable measurement error. A much better approach would be to consider longitudinal measurements of CRP each study participant over a period of time; in this way intra-individual variation can be quantified and accounted for in the analysis.

Given the numerous difficulties discussed above, it is perhaps surprising that I was able to identify any genetic effect at all, given that my cohort was small and only a single CRP measurement per person was recorded. The consistency of genetic effect between this study, the limited number of other acute-phase studies, and even the basal CRP studies, makes it likely that the associations that I observed are true. It is possible that, in the context of acute-phase CRP, the genetic effect is actually quite strong and demonstrable even in a study affected by large amounts of random measurement error. There are, however, enough positive findings from this study to warrant a more extensive analysis along similar lines. Certainly a number of well validated rheumatoid arthritis cohorts have been collected by investigators around the world which should offer improvements in terms of cohort size and longitudinal data. As I will discuss further in chapter 8, there are good reasons why an understanding of acute-phase *CRP* genetics should be a priority for further study.

## **CHAPTER 7**

# **COMPLEMENT C3 GENETICS IN UK SLE FAMILIES**

## 7.1: Complement C3 genetics in UK SLE Families – Study Aims

The key physiological roles played by the C3 component of complement and the existing evidence of genetic association with diseases such as bronchial asthma and age relate macular degeneration has been reviewed in sections 1.2 and 1.4. Only a single study has reported an association between C3 genotype and serum C3 levels; in a cohort of 87 Japanese individuals with SLE (Miyagawa *et al* 2008). These authors also reported an association between C3 variation and SLE itself. Given the frequency with which false-positive genetic associations can be observed, these data can not be considered conclusive until replicated further.

The primary aim of the work presented in this chapter was therefore to establish the heritability of serum C3 and, if shown to be a heritable trait, to delineate *cis*-acting genetic variants determining C3 expression. Finally we aimed to assess the association between C3 genotype and SLE. The data presented in this chapter may be found in published form (Hunnangkul *et al* 2008; Rhodes *et al* 2009).

## 7.2: Serum C3 – Distribution and Heritability

### 7.2.1: C3 heritability – Regression Estimate

Serum C3 had already been quantified in individuals from the first 103 families of cohort 1, by previous members of my supervisor's laboratory. Serum C3 data from the parents and their *unaffected* offspring from this cohort was used to calculate a simple regression based estimate of heritability as outlined in section 2.5.2. This approach gave a raw heritability estimate of 40.2% ( $h^2=0.402$ ; 95%CI 15.9, 64.4). Adjusting for age gave a revised heritability estimate of 38.7% ( $h^2=0.387$ ; 95%CI 13.5, 63.9;  $P=0.003$ ).

### 7.2.2: C3 heritability – Replication using Bayesian Methodology

In order to replicate this heritability estimate, serum C3 was quantified on a further 1371 individuals from 393 UK white European families from Cohort 1. This, therefore, represented an entirely independent cohort from that used for the regression estimate of heritability presented above. The mean age of these individuals was 53.9 (standard deviation 16.5) and 33.7% were male. The mean serum C3 was 1054 mg/L (range 294 to 1564 mg/L). **Table 7.1** demonstrates the influence of age, sex and SLE affection on serum C3, assessed using both maximum likelihood and Bayesian methodologies. As expected, serum C3 was significantly lower in SLE probands; it was also lower in male participants and increased marginally with age.

**Table 7.1** Effect of age, sex and SLE affection on serum C3

|            | Maximum Likelihood Analysis |                         |                 | Bayesian Analysis |                       |
|------------|-----------------------------|-------------------------|-----------------|-------------------|-----------------------|
|            | $\beta$                     | 95% confidence interval | <i>P</i> -value | $\beta$           | 95% credible interval |
| Age*       | 0.088                       | 0.029, 0.147            | 0.003           | 0.038             | -0.042, 0.119         |
| Female Sex | 0.132                       | 0.024                   | 0.017           | 0.158             | 0.011, 0.303          |
| SLE        | -0.477                      | -0.602, -0.351          | <0.001          | -0.525            | -0.685, -0.362        |

\*Standardised age. Coefficient is for a change of one standard deviation (16.5 years) from the mean age of 53.9 years)

In a Bayesian variance components model of serum C3, adjusted for SLE affection, age and sex, C3 heritability was estimated to be 39.6% ( $h^2 = 0.396$ ; 95% Credible Interval 0.250, 0.565). This estimate of heritability was robust to a range of prior assumptions in the model,

and did not change by an appreciable extent when alternative prior distributions were specified as outlined in section 2.5.2.

### 7.3: C3 Genetic Architecture

All the 14 genotyped C3 tagSNPs passed quality control checks. Detailed marker information is presented in **table 7.2**.

**Table 7.2.** Typed SNPs demonstrating alleles and rare allele frequencies

| rs number         | SNP | Location                | Rare allele frequency | HWP* | missing data (%) |
|-------------------|-----|-------------------------|-----------------------|------|------------------|
| <i>rs2230199</i>  | C>G | exon 3 (non-synonymous) | 0.22                  | 0.02 | 1.2              |
| <i>rs11672613</i> | A>G | intron 17               | 0.41                  | 0.36 | 1.9              |
| <i>rs11569450</i> | C>G | intron 18               | 0.09                  | 0.96 | 1.3              |
| <i>rs423490</i>   | C>A | exon 21 (synonymous)    | 0.26                  | 0.56 | 1.5              |
| <i>rs2287846</i>  | C>G | intron 22               | 0.40                  | 0.97 | 2.4              |
| <i>rs3745568</i>  | A>C | intron 27               | 0.12                  | 0.79 | 0.9              |
| <i>rs11569520</i> | T>C | intron 27               | 0.14                  | 0.13 | 1.1              |
| <i>rs344540</i>   | C>T | intron 27               | 0.37                  | 0.23 | 1.5              |
| <i>rs344542</i>   | T>C | intron 27               | 0.37                  | 0.25 | 1.3              |
| <i>rs7257062</i>  | G>A | intron 29               | 0.46                  | 0.70 | 1.1              |
| <i>rs344550</i>   | C>G | intron 33               | 0.36                  | 0.99 | 1.2              |
| <i>rs7951</i>     | C>T | exon 35 (synonymous)    | 0.08                  | 0.72 | 1.2              |
| <i>rs2277984</i>  | A>G | intron 36               | 0.50                  | 0.92 | 1.1              |
| <i>rs344555</i>   | G>A | intron 37               | 0.23                  | 0.52 | 1.2              |

\*Hardy-Weinburg probability calculated by  $\chi^2$

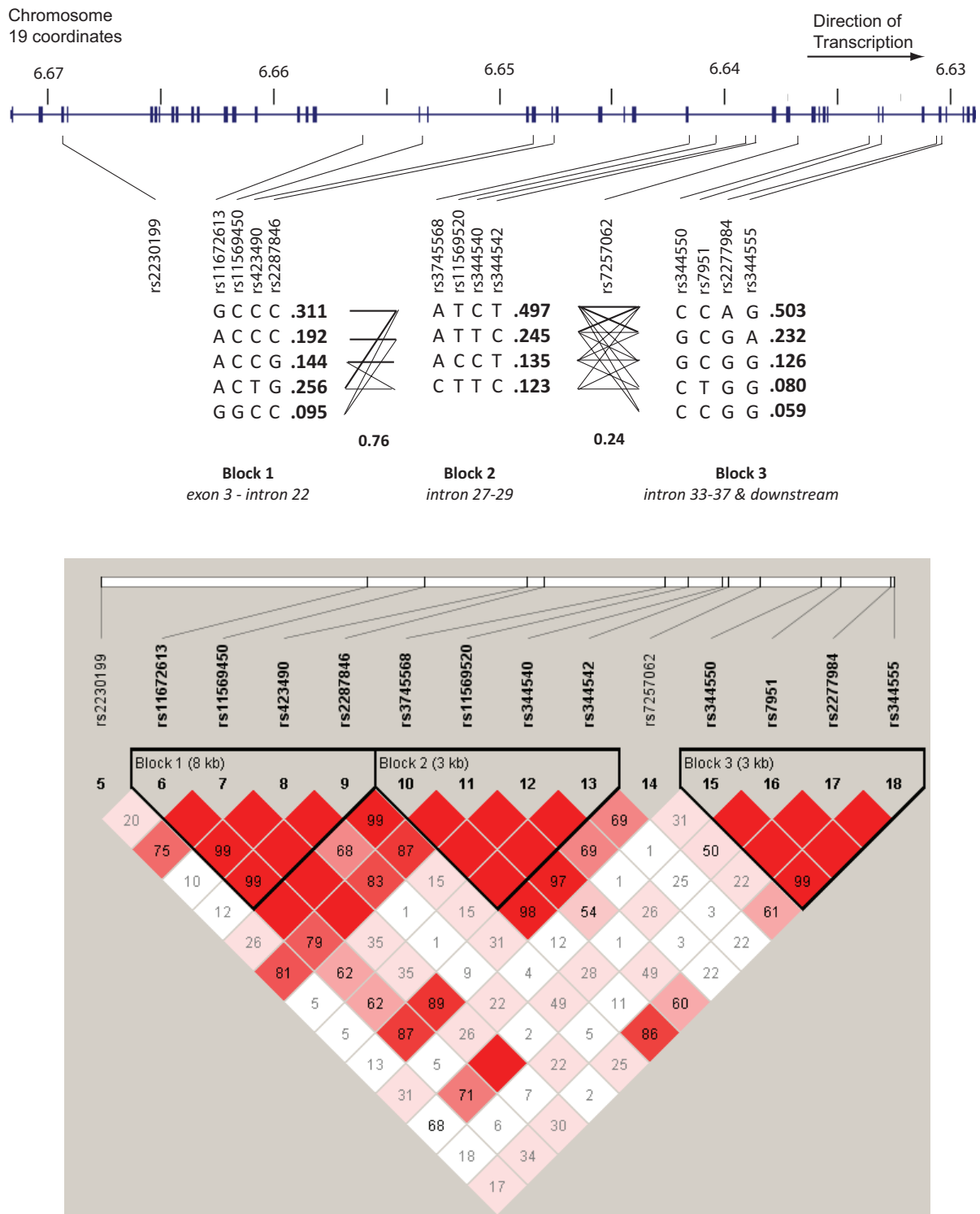
**Table 7.3** demonstrates the pairwise linkage disequilibrium between markers while **Figure 7.1** illustrates the haplotype structure calculated using Haploview. With the exception of *rs344540* and *rs344542* that were in tight linkage disequilibrium ( $r^2 = 0.998$ ), my data confirmed the relatively low level of linkage disequilibrium across C3 demonstrated in the HapMAP CEPH population.

**Table 7.3.** Pairwise linkage disequilibrium between C3 SNPs, calculated from the unrelated parents.

|                   | <i>rs2230199</i> | <i>rs11672613</i> | <i>rs11569450</i> | <i>rs423490</i> | <i>rs2287846</i> | <i>rs3745568</i> | <i>rs11569520</i> | <i>rs344540</i> | <i>rs344542</i> | <i>rs7257062</i> | <i>rs344550</i> | <i>rs7951</i> | <i>rs2277984</i> | <i>rs344555</i> |
|-------------------|------------------|-------------------|-------------------|-----------------|------------------|------------------|-------------------|-----------------|-----------------|------------------|-----------------|---------------|------------------|-----------------|
| <i>rs2230199</i>  | X                | 0.017             | 0.017             | 0.009           | 0.003            | 0.036            | 0.029             | 0.002           | 0.002           | 0.004            | 0.049           | 0.011         | 0.009            | 0.028           |
| <i>rs11672613</i> | 0.204            | X                 | 0.151             | 0.233           | 0.455            | 0.095            | 0.067             | 0.152           | 0.152           | 0.430            | 0.002           | 0.03          | 0.003            | 0.054           |
| <i>rs11569450</i> | 0.756            | 1.000             | X                 | 0.036           | 0.069            | 0.014            | 0.017             | 0.022           | 0.023           | 0.069            | 0.013           | 0.009         | 0.001            | 0.031           |
| <i>rs423490</i>   | 0.103            | 0.992             | 1.000             | X               | 0.519            | 0.190            | 0.038             | 0.000           | 0.000           | 0.002            | 0.031           | 0.000         | 0.017            | 0.000           |
| <i>rs2287846</i>  | 0.128            | 0.992             | 1.000             | 1.000           | X                | 0.203            | 0.177             | 0.010           | 0.010           | 0.078            | 0.001           | 0.031         | 0.002            | 0.013           |
| <i>rs3745568</i>  | 0.267            | 1.000             | 1.000             | 0.686           | 0.990            | X                | 0.022             | 0.240           | 0.240           | 0.162            | 0.001           | 0.053         | 0.002            | 0.031           |
| <i>rs11569520</i> | 0.811            | 0.793             | 1.000             | 0.834           | 0.879            | 1.000            | X                 | 0.091           | 0.091           | 0.176            | 0.026           | 0.000         | 0.039            | 0.018           |
| <i>rs344540</i>   | 0.059            | 0.621             | 0.352             | 0.015           | 0.158            | 1.000            | 1.000             | X               | 0.998           | 0.337            | 0.000           | 0.010         | 0.001            | 0.009           |
| <i>rs344542</i>   | 0.058            | 0.621             | 0.356             | 0.015           | 0.156            | 1.000            | 1.000             | 1.000           | X               | 0.336            | 0.000           | 0.010         | 0.001            | 0.008           |
| <i>rs7257062</i>  | 0.139            | 0.871             | 0.894             | 0.091           | 0.311            | 0.989            | 0.971             | 0.695           | 0.694           | X                | 0.046           | 0.026         | 0.043            | 0.096           |
| <i>rs344550</i>   | 0.312            | 0.053             | 0.262             | 0.224           | 0.042            | 0.128            | 0.545             | 0.010           | 0.011           | 0.316            | X               | 0.048         | 0.556            | 0.540           |
| <i>rs7951</i>     | 0.688            | 0.716             | 1.000             | 0.027           | 0.498            | 0.290            | 0.018             | 0.263           | 0.258           | 0.502            | 1.000           | X             | 0.087            | 0.026           |
| <i>rs2277984</i>  | 0.182            | 0.064             | 0.078             | 0.222           | 0.052            | 0.115            | 0.499             | 0.033           | 0.033           | 0.229            | 1.000           | 1.000         | X                | 0.308           |
| <i>rs344555</i>   | 0.174            | 0.349             | 0.304             | 0.020           | 0.257            | 0.862            | 0.609             | 0.223           | 0.220           | 0.615            | 0.996           | 1.000         | 1.000            | X               |

\*  $r^2$  values presented at the top right and D' values at the bottom left.

**Figure 7.1.** *C3* haplotype structure.



Haplotype construction across *C3* demonstrating the position of the genotyped SNPs relative to *C3*. There are three haplotype blocks across the gene, haplotype frequencies shown to the right of each haplotype. Thick line connect blocks occurring together in >10% of chromosomes and thin lines if in >1% of chromosomes. Inter-block  $D'$  scores are shown below the gaps between haplotype blocks. SNP *rs2230199* falls upstream of haplotype block 1, while *rs7257062* is located between blocks 1 and 2. The lower section demonstrates the pairwise LD between markers with haplotype blocks defined by the default algorithm of Gabriel *et al* (see methods section 2.4.4). The pairwise LD between any two SNPs is indicated by the colour coding of the corresponding square of the LD plot. High LD is indicated by a pink/red colour scheme (the dark red squares are linked with a  $D' = 1$  and a LOD score  $\geq 2$ ). Numbers within the squares correspond to the  $D'$  score, with 96 representing a  $D' = 0.96$  and 44 a  $D' = 0.44$  etc. Where squares have no number the  $D' = 1$

## 7.4: Genetic associations with serum C3

### 7.4.1: Association of C3 SNPs with basal C3

It was noted that *rs344540* and *rs344542* were in tight linkage disequilibrium ( $r^2 = 0.998$ ) so *rs344542* was dropped from further analysis, leaving 13 SNPs for association testing. Each C3 SNP was evaluated for association with serum C3 level using both Bayesian and maximum likelihood multiple linear regression models, incorporating age, sex and SLE affection as covariates. Significant associations are shown in **Table 7.4**. Strong association was observed between serum C3 levels and SNP *rs344555*, with additional weaker associations also seen with SNPs *rs2277984*, *rs7951* and *rs423490*. In particular the rare allele of *rs344555* was associated with higher serum C3 levels in both maximum likelihood and Bayesian models. This association remained significant at  $P < 0.05$  even after applying a Bonferroni correction adjusted for the correlation between markers (13 SNPs mean  $r = 0.2704$ ). A global Chi-Square of association for all SNPs was also significant ( $P = 0.026$ ).

**Table 7.4.** Significant C3 SNP associations with serum C3.

|                  | Maximum Likelihood Analysis |                         |                 | Bayesian Analysis |                       |
|------------------|-----------------------------|-------------------------|-----------------|-------------------|-----------------------|
|                  | $\beta$                     | 95% confidence interval | <i>P</i> -value | $\beta$           | 95% credible interval |
| <i>rs423490</i>  | 0.121                       | -0.004, 0.246           | 0.058           | 0.133             | 0.002, 0.262          |
| <i>rs7951</i>    | -0.220                      | -0.404, -0.037          | 0.019           | -0.211            | -0.402, -0.024        |
| <i>rs2277984</i> | -0.114                      | -0.216, -0.012          | 0.028           | -0.111            | -0.215, -0.005        |
| <i>rs344555</i>  | 0.172                       | 0.048, 0.296            | 0.007           | 0.157             | 0.028, 0.290          |

By reference to **Figure 7.1** it can be seen that *rs344555*, the key SNP associated with serum C3 level, forms part of a four SNP haplotype encompassing the 3' end of *C3* downstream of intron 29. Also incorporated in this haplotype are *rs7951* and *rs2277984*, both of which showed weak association with serum C3, with the rare alleles associated with lower serum C3.

#### 7.4.2: Trans-acting genetic effects

In addition to analysing the association between serum C3 and *C3* genotype, a *post-hoc* association analysis was performed using existing genotype data for *IL6* and *IL1B* (presented in section 3.4.2 and **table 3.2**). No association was observed for the *IL6* genotypes, but a borderline association was seen between serum C3 and two SNPs in *IL1B*, *rs7596461* and *rs1143634* (**Table 7.5**).

**Table 7.5.** Single *IL1B* SNP associations with serum C3.

|                  | Maximum Likelihood Analysis |       |                         |         | Bayesian Analysis |                       |
|------------------|-----------------------------|-------|-------------------------|---------|-------------------|-----------------------|
|                  | $\beta$                     | S.E.  | 95% confidence interval | P-value | $\beta$           | 95% credible interval |
| <i>rs4849124</i> | 0.055                       | 0.060 | -0.063, 0.172           | 0.363   | 0.070             | -0.055, 0.195         |
| <i>rs7596461</i> | -0.168                      | 0.096 | -0.355, 0.019           | 0.079   | -0.134            | -0.331, 0.060         |
| <i>rs1143634</i> | 0.113                       | 0.065 | -0.014, 0.239           | 0.082   | 0.133             | -0.001, 0.267         |
| <i>rs16944</i>   | -0.006                      | 0.058 | -0.119, 0.107           | 0.922   | -0.103            | -0.206, 0.001         |

## 7.5: C3 Genetic Association with SLE

### 7.5.1: Analysis of genotyped SNPs

I performed a family-based analysis of the association of C3 genotype with SLE using TRANSMIT, evaluating each SNP in turn for an association with SLE. A significant association was only observed with SNP *rs3745568*. Of 569 families with transmissions to affected offspring observed, there were 113 observed, compared with 134.58 expected transmissions of the rare allele to affected offspring ( $\chi^2$  (1 d.f.) = 8.026,  $P=0.0046$ ). This corresponded to  $P=0.030$  after correcting for multiple hypothesis testing as above. No other markers approached even borderline significance with SLE.

### 7.5.2: Refinement and replication of SLE association

To evaluate the robustness of the association of *rs3745568* with SLE we performed two additional analyses. Firstly, additional SNP genotypes were imputed across the C3 locus and then tested for association with SLE. Secondly I aimed to replicate this association in an independent study cohort.

Using the pattern of linkage disequilibrium observed in the HapMAP CEPH panel, combined with the genotype data from the 14 SNPs typed in this UK SLE family cohort, the genotype of an additional 287 HapMAP SNPs spanning a 300kb interval incorporating the C3 locus (Chromosome 16: 6503308 – 6799220) was imputed as described in section 2.5.3. None of the imputed SNPs were associated with SLE ( $P>0.1$ ), although *rs3745568* remained significant ( $P=0.0045$ ) with the C allele under-transmitted.

I aimed to replicate the association between *rs3745568* and SLE in an additional independent cohort. I did this using existing genotype data available from the SLEGEN case-control genome wide association study (Harley *et al* 2008). Although *rs3745568* was included among the SNPs on the Illumina 300 chip employed in this study, the assay unfortunately performed poorly, with successful genotyping in <40% of the study population. In an attempt to make some use of this data, C3 SNPs that were successfully genotyped in this SLEGEN GWA study were used to impute genotype at *rs3745568*, along with the 287 additional HapMAP SNPs discussed above. Using the raw, unimputed data (<40% genotyping) no significant association with SLE was seen ( $P=0.135$  for an additive model). Even after data imputation to boost genotyping at *rs3745568* to 769 cases and 3514 controls ( $P=0.350$  for an additive model), no significant association was observed

## **7.6: Complement C3 Genetics in UK SLE Families – Discussion**

### *7.6.1: Summary of findings*

The data presented in this chapter demonstrate that serum C3 is a heritable trait and, to the best of my knowledge, this is the only existing study to quantify C3 heritability. I demonstrate heritability initially in a cohort of healthy individuals ( $h^2=0.387$ ), and replicate this in a larger, independent cohort of SLE families, using alternative statistical methodology ( $h^2=0.396$ ). Using a tagSNP approach to genotyping, to ensure all common genetic variation at C3 was captured, I identify SNPs associated with both serum C3 level (*rs344555*) and with SLE (*rs3745568*).

### 7.6.2: Associations with serum C3 - discussion

The strongest association with serum C3 was observed for SNP *rs344555*, located in intron 37. The rare allele of this SNP was associated with higher level of serum C3, even after correction for the testing of multiple hypotheses. Haplotype analysis demonstrated that *rs344555* was located in a distinct haplotype block, along with *rs2277984*, *rs7951* and *rs344550*. The rare alleles of *rs344555*, *rs2277984* and *rs7951* each defined a distinct haplotype, with two additional haplotypes delineated by the rare allele of *rs344550*. SNP *rs344555* therefore defined a haplotype associated with higher serum C3. SNPs *rs2277984* and *rs7951*, and the haplotypes they defined, were associated with lower serum C3, albeit with lower levels of significance.

As I have discussed elsewhere, while the tagSNP approach to study design maximises the chance of identifying an indirect genetic association, it leaves unresolved the identity of the true functional variant, which may not have been genotyped, or even discovered. By reference to the location of *rs344555* and the pattern of linkage disequilibrium observed both in our own data and in data from the HapMap CEPH panel it is possible, however, to make some broad statements about where the functional variant is likely to lie. There is very weak LD between the key variant *rs344555* and SNPs in the *C3* promoter or proximal coding sequence. There is moderate LD between *rs344555* and sporadic SNPs in the downstream gene, *TNFSF14*, but it would seem unlikely that these SNPs alter serum C3 levels. Most likely, therefore, is that serum C3 is influenced by genetic variation in the distal end of *C3*. Since no amino-acid coding changes have been identified in this region, it is possible that the functional variant alters serum C3 levels by an effect on the stability or trafficking of *C3* mRNA.

Interestingly, the incorporation of *rs344555* genotype as a covariate in our variance components model resulted in only a negligible change to the overall heritability estimate, suggesting that the key genetic effects accounting for C3 variability arise from *trans*-acting loci and currently remain unaccounted for. Since serum C3 level in healthy individuals will be determined both by the rate of C3 synthesis, and by the rate of low-level complement turnover, it is possible that variation in a large number of genes may account for the observed heritability. These may include acute-phase cytokine genes, early complement component genes or complement regulatory factor genes. We were unable to identify a clear association between serum C3 and the well described *IL6* and *IL1B* variants we tested, although it should be noted that these SNPs do not constitute a comprehensive screen of either of these loci.

#### *7.6.3: Associations with SLE - discussion*

A different variant accounted for the effect on SLE susceptibility. The disease associated SNP, *rs3745568*, was located in a haplotype block encompassing the central region of *C3* (between intron 22 and intron 29). The imputation analysis failed to identify a stronger disease association in an extensive array of 287 known SNPs in and around *C3*, while no association was observed with the directly genotyped *rs2230199*, encoding the C3F/C3S structural variant. Interestingly, no association was observed between *rs3745568* and serum C3, while there is no evidence from the HapMap CEPH panel of longer range LD between this SNP and the downstream gene *TNFSF14*. The mechanism underlying the association between this SNP (or variants in LD with it) and SLE therefore remains uncertain but, from our data, it does not appear to be a result of genetically determined alteration in serum C3 level.

#### 7.6.4: A comparison with existing studies

To my knowledge this is the only existing study to report the heritability of serum C3. It is interesting to note, however, that our estimate of 39.6% heritability for serum C3 is of a similar magnitude to that reported for other acute phase plasma proteins, including C-reactive protein and fibrinogen (de Lange M. *et al* 2001; Fox *et al* 2008; Rhodes *et al* 2008a).

Only one other study, in a Japanese population, has reported an association between C3 polymorphisms and both serum C3 and SLE, and there are some notable differences in the genetic associations identified between my study and this one (Miyagawa *et al* 2008). The key SLE associated SNP in the Japanese population was *rs7951*, which we tested but found no disease association. Conversely we observed an SLE association with *rs3745568*, which was not tested in the Japanese study, but with reference to the HapMap Japanese panel does not appear to be in strong LD with either *rs7951* or any other variant typed by these authors. One explanation for this difference is that neither study has typed the true functional variant, and that differing patterns of LD in these ethnically diverse populations explains differences in observed indirect associations. Alternatively our study simply lacked power to identify an association with *rs7951*, which was present at slightly lower frequency in our white European population. Assuming an disease odds ratio of 1.40 as determined in the Japanese study, a power calculation suggests a family based association study of our size would have only 30% power to detect an association at  $P < 0.01$ . It also possible, that all present findings, although plausible, may be false positives, and future larger studies, with detailed fine mapping may help inform on this issue.

In contrast to the disease associations a similar association was observed between C3 variants and serum C3 in the two populations. In the Japanese population an association between

*rs7951* and serum C3 was observed in a cohort of 87 SLE probands, leading to the hypothesis that disease susceptibility is increased by a variant that also leads to lower serum C3 levels. Like this study we also observed a weak association with *rs7951*, although our key association was seen with *rs344555*, which was not typed in the Japanese study. Our data suggest that, in a white European population, distinct high and low C3 associated haplotypes within this block are tagged by *rs344555* and *rs7951* respectively. By reference to the HapMap Japanese population it appears the LD structure in this region of the gene is rather different from ours, with no notable LD between *rs7951* and *rs344555*. Once again it is likely that the true functional SNP remains to be identified, but both studies highlight the 3' end of the gene as its most likely location.

#### *7.6.5: Key study features and potential limitations*

This study is the first to comprehensively map genetic variation across *C3*, and to look for association with both serum C3 levels and SLE. The family based-study design offers advantages in terms of control for bias due to population stratification, as outlined in section 1.5.4. The haplotype tagging strategy I employed should ensure that all common genetic variation at *C3* has been captured. A disadvantage of this approach, however, is that there remains uncertainty over the identity of truly functional alleles (which may not have been genotyped, or even identified). I attempted to address this issue through the use of a SNP imputation approach, utilising the extensive genotype data freely available from HapMAP. A standard maximum likelihood approach was taken for the association of genotype with disease, but for the heritability estimates and association of genotype with serum C3 an additional Bayesian model was employed. Bayesian analyses are increasingly popular in genetic studies and offer advantages in allowing us to test the sensitivity of our results for different prior assumptions regarding, for example, the within family correlation of serum

C3. The Sequenom iPLEX Gold methodology employed for the genotyping presented in this chapter appeared to produce high quality genotyping, with minimal missing data. In addition the family-based study design allowed for the checking of assigned genotypes for Mendelian errors, allowing an additional level of quality control. I am therefore confident the study design and genotyping were as optimal as possible.

I am also confident that the Radial Immunodiffusion assays provided an accurate assessment of serum C3 concentration at the time the assay was performed. Although this is a relatively old and labour intensive technique, and has been replaced in standard clinical usage by newer automated assays such as nephelometry, existing published comparisons suggest it is no less accurate in terms of quantifying plasma protein level (Bouzek & Mancal 1989; Guiguet *et al* 1983). Any assay is associated with random measurement error, but our use of control sera confirmed that the plate to plate variation in readings did not deviate significantly from the normal distribution expected due to this random error (Kologorov-Smirnov  $P=0.20$  – see section 2.3.2). While this random, undifferential, error may reduce the power of a study to demonstrate a genetic effect, it should not alter the interpretation of a statistically significant association if one is still demonstrated. It should also be noted that anti-C3 antibodies may differ in their specificity for antigenic determinants on C3, including a difference in their ability to recognise C3 breakdown products, emphasising the importance of assay standardisation (Spath & Gabl 1976; Vladutiu & Winiarski 1976). The radial immunodiffusion assay technique and antibody reagents I employed were well standardised within our laboratory, and were indeed used for many years to quantify C3 for the clinical service at the Hammersmith Hospital.

While I am confident that the C3 assay provided a reasonably accurate and reproducible quantification of serum C3 at the time the assay was performed, it is important to recognise

that this may not reflect the serum C3 level at the time the blood sample was collected. There is published evidence to suggest that the sample storage conditions prior to C3 quantification influence the measured value; in this respect C3 would appear to be a less robust molecule than CRP. For example there is evidence that intact C3 degrades in serum stored at room temperature (Davis *et al* 1972; Sinosich *et al* 1982). This is not, however, a universal finding (Sjoholm 1975). Given that many of the serum samples used in this study arrived through the post at room temperature, it is clear that this has the potential to introduce quite considerable error into the quantification of C3. However, since the duration of room temperature storage was entirely random with respect to any particular individual, disease manifestation or other phenotypic characteristic, the error it introduces will also be random. This adds to the level of random measurement error for the samples, but will not introduce any systematic bias. Just as I discussed for the measurement error associated with the assay itself, this may significantly reduce the power to detect a genetic association, but it will not affect our interpretation of a statistically significant association if one is still demonstrated. Although the only paper that specifically addresses the issue of freeze-thawing on the quantification of C3 suggests this is of negligible importance, to ensure consistency we only quantified C3 on serum aliquots that had never been thawed and re-frozen (Sjoholm 1975).

It is important to recognise that this analysis of association between *CRP* genetic variation and serum C3 represents a *post-hoc* analysis of a cohort originally collected to study genetic associations with SLE. Consequently the study has limitations, and in particular it is important to consider the impact that the inclusion of individuals with and without SLE may have on my findings. This argument has also been covered in some detail in relation to CRP in section 3.7.2. We have already discussed in section 1.2.3 that there is a complex relationship between serum C3 concentration and SLE, with reduced C3 levels often seen in patients with SLE, and levels also fluctuating according to the level of disease activity. I

recognise that this adds significant and unwanted complexity to the analysis and interpretation of my data, however I do not feel it entirely precludes use of this data. Firstly, by including an “SLE” term as a fixed factor in the regression models, a crude adjustment for the overall effect of disease on serum C3 concentration can be made. Having already made this adjustment for the presence of disease, we can then consider the effect that unmeasured variation in disease activity may have on the results. It should be noted that the collection of serum samples from the SLE patients was performed in a cross-sectional study design in which blood collection was not performed at any particular time-point in relation to disease activity. We would therefore expect the “error” associated with fluctuations in disease activity (as opposed to the presence or absence of disease itself), over and above that due to genetic variation, to be random. Once again, as outlined above, this random error should not affect the validity of our associations. Although providing only indirect evidence, it is perhaps notable that the original, regression based estimate of C3 heritability (37.8%), in a cohort of healthy individuals only, was remarkably similar to the second estimate (39.6%), made in a cohort with SLE probands included, but adjusted for with a regression parameter. This similarity suggests that the generic “SLE” term included as a covariate in our association models was adequate to account for the effect of disease status on serum C3 levels. It should also be considered that serum C3 is a biological trait that, clinically, is really only measured in people with SLE: cohorts of healthy individuals with measured C3 levels do not exist and would have to be collected *de novo*.

#### 7.6.6: Conclusions

The heritability estimates presented in this chapter demonstrate that there is a significant genetic element to the observed variability of serum C3 in our study populations. It should, of course, be recognised that heritability estimates are, strictly, only applicable to the population

for which they are calculated. While the inclusion of individuals with SLE made little relevant difference to the magnitude of the heritability estimate, it should be recognised that both populations we studied consisted of the family members of lupus probands, and hence may be genetically enriched for particular disease susceptibility variant. Despite this note of caution, the level of heritability that was seen, at the very least, suggests it may be feasible to identify genetic variants explaining the population variation in C3.

I was subsequently able to demonstrate a significant association between a C3 SNP and serum C3 and, by reference to the pattern of LD across the gene would argue that any functional variant is likely to lie at the distal, 3', end of the gene. Although I was unable to replicate this association (due to the lack of access to any alternative large cohorts of individuals with both serum and DNA collected), my findings did agree to some extent with the only other published C3 genetic association paper, in a Japanese population (Miyagawa *et al* 2008).

I was also able to identify a significant association between SNP *rs3745568* and SLE. Unlike the Japanese study this SNP did not appear to be associated with serum C3 level. It was unfortunate that I was unable to replicate this association in a separate analysis of the SLEGEN GWA data. There was, of course, problems with the data from this study, not least the extremely poor quality of genotyping at the SNP of interest, which, in most circumstances would preclude its use from any analysis at all. I attempted to circumvent this by imputation analysis but again failed to find an association. This raises the possibility that the original association I demonstrated was a false positive, although of the loss of power in the replication study due to uncertainty around the inferred genotypes may be an alternative explanation. Perhaps some indirect evidence that a genetic interval incorporating *rs3745568* represents a true disease susceptibility effect is that other haplotype associations (in the

absence of any known coding SNPs) have been observed; notably an association between bronchial asthma and a similar central/downstream haplotype to that which we identified (Inoue *et al* 2008).

The findings presented in this chapter are therefore interesting but clearly, before the genetic associations we identify can be considered definitive, further work will be required. In particular it will be necessary to replicate the findings in additional cohorts of similar ethnicity. Narrowing down the location of the true functional variants will require more detailed fine-mapping or gene sequencing.

## **CHAPTER 8**

# **CONCLUSIONS**

## 8.1: Summary of thesis aims

In the introduction to this thesis I reviewed our current understanding of the physiological roles of the serum proteins CRP and C3, and discussed current literature implicating these proteins in a number of disease processes, most notably atherosclerosis and SLE. The data presented in this thesis aimed to expand our understanding of the genetic basis of expression of these two proteins. In particular I aimed to answer the following key genetic questions:

1. To what extent is the variability in serum level of the two studied proteins under genetic control (i.e. determined by an inherited underlying predisposition and not simply related to external factors such as pathogen burden, or subclinical disease processes)?
2. Can we identify specific genetic variants accounting for variation in serum levels and, if we can, are these variants located in the *CRP* or *C3* genes themselves, or at *trans*-acting loci?
3. Is there a single major genetic effect or do multiple genetic variants each account for a small component of the variability in serum protein level?
4. Do the same genetic variants account for the variability in the basal serum levels seen in healthy individuals, compared with the higher levels seen in individuals with disease?
5. Are *cis*-acting genetic variants also associated with the diseases in which these proteins have been implicated?

In the following sections I will summarise the results of the studies presented in this thesis, demonstrating that they make significant progress in answering the above questions. I will then discuss, in detail, points of general relevance to the studies presented here, which have not been covered in the individual chapter discussions. In particular I will discuss the identity of potential genetic variants that determine basal CRP, a topic which has been addressed in

three independent studies presented in this thesis (chapters 3, 4 and 5). Finally I will conclude by discussing the clinical relevance of these findings, and the future directions that this work may take.

## 8.2: Summary of Study Findings

The data and analyses presented in this thesis may be summarised in the following key messages:

1. Both serum CRP and serum C3 are heritable traits with a significant proportion of variation in their levels attributable to genetic effects.
2. *Cis*-acting genetic variants in both *CRP* and *C3* are associated with serum levels, but these account for only a small proportion of the total genetic effect
3. For *CRP*, strong LD across the gene results in the reporting of multiple indirect associations. By adopting a trans-ethnic mapping approach, and utilising appropriate model choice analysis, I identify a tri-allelic promoter SNP (*rs3091244*) as the most likely key regulator of CRP expression at this locus.
4. Genetic variation also determines acute-phase CRP expression in the context of chronic inflammatory disease, a finding that has implications for how we use CRP to diagnose and classify inflammatory conditions.
5. Only weak effects are identified from selected *trans*-acting loci. Recent whole genome association studies are more informative.

### 8.3: Serum CRP and C3 – General Considerations and Heritability

#### 8.3.1: Basal vs. acute-phase levels

As I outlined in section 1.3.2, CRP expression has a vast dynamic range, from low levels in apparently healthy individuals to very high levels in patients with bacterial infection. The range of levels is so great that it is common to refer to “basal” CRP when talking about low-level expression in healthy people, compared with “acute-phase” CRP in individuals with active inflammation. Indeed it is widespread in much of the literature to give “basal” CRP a different name, “highly sensitive” or “hsCRP”, although, of course, it is an identical molecule to the “acute-phase” protein. C3 levels are also increased in inflammatory conditions but because the change is less dramatic a distinction between basal and acute-phase expression is not usually made. When designing or evaluating a *CRP* genetic study it is important to consider whether this distinction between basal and acute-phase CRP is a matter of semantics, or whether there is a genuine difference in physiological regulation. There are two possibilities. Firstly CRP expression could be regulated in essentially the same way across its entire range, with the same intracellular and extracellular signals driving expression but to a greater or lesser extent. In this model, basal CRP would reflect low levels of ongoing inflammatory stimulus, either as a result of subclinical inflammatory disease, or in response to ongoing encounter with pathogens that are eliminated before causing overt disease. The alternative model is that CRP is constitutively expressed at low levels as a distinct regulated process, unrelated to classical inflammatory stimuli such as IL1 $\beta$  and IL6. In this model the production of acute-phase CRP would be a distinct event with unique regulatory steps.

Although there is little direct evidence, as *in vitro* cell based systems all require treatment with acute-phase cytokines before measurable CRP can be detected, I suspect that the second,

dual regulation, scenario is most likely. This supposition is based on two lines of evidence. Firstly there is evidence from the data presented in chapters 2 and 3 that the association of interleukin 6 SNPs with serum CRP is restricted to individuals with SLE, suggesting genetic regulation of CRP expression may be different in inflammatory states. A recent GWA study (discussed in detail in section 8.5.2) also noted that the association of some of the loci, notable the interleukin 6 receptor gene *IL6R*, was weaker when individuals with serum CRP >10mg/L were excluded from the analysis (Ridker *et al* 2008). Secondly there is some *in vitro* evidence to suggest that the production of CRP in IL-1 $\beta$  and IL-6 stimulated cells is associated with increased binding of the specific transcription factor C/EBP $\beta$ , whereas other transcription factors such as STAT3 and Rel p50 remain bound to the endogenous, unstimulated *CRP* promoter (Young *et al* 2008).

This issue of whether there is a true mechanistic difference between basal and acute-phase protein levels is important, because if these processes are regulated differently then it is possible that different genetic variants, either at *CRP* itself, or at different *trans*-acting loci may account for variability in CRP expressed as part of either pathway: this in turn impacts on the design of studies aiming to highlight variants of interest. Numerous authors apply a cut off of CRP level above which individuals are excluded on the assumption they are mounting an acute-phase response. This seems logical if we assume basal and acute-phase CRP are distinct processes, nonetheless it does present other difficulties. For example, it is likely that some individuals with a CRP within the range of, say, 10-15mg/L are mounting a low grade acute-phase response. Other individuals, however, may not have an acute-phase response but simply have naturally high basal CRP. It is exactly these high basal CRP individuals who may be most informative in a genetic association analysis of basal CRP, and we will lose this useful information if they are excluded. Furthermore, by simply cutting off the end of the tail of the positive distribution of CRP values we may end up, after transformation, with a worse

approximation of a normal distribution, hence violating the assumption of normality that is essential for a parametric approach to analysis.

The alternative approach again used in a number of studies (and adopted by myself in the analysis presented in chapter 5) is simply to include everybody. This approach would seem reasonable for population cohorts collected without reference to a specific inflammatory disease, particularly if a smooth distribution of CRP values is seen within the population. However, if too many of these individuals are mounting an acute-phase response it should be recognised that genetic variants specific to acute-phase production may be highlighted at the expense of variants influencing basal CRP

### *8.3.2: Tissue specific expression*

I have discussed the possibility that genetic regulation of CRP expression may differ in basal and acute-phase situations. A related question is whether the genetic regulation of CRP or C3 expression is tissue specific. As discussed in sections 1.3.2 and 1.4.2, the primary site of production of both these proteins is the liver, but notable sites of extra-hepatic production have been described. Presumably the contribution of these extra-hepatic sources to total circulating serum levels in the acute-phase situation is minimal and can essentially be ignored. Whether they may make a more marked contribution to basal serum levels is unknown but certainly it could be envisaged that, for example, CRP produced by the vascular endothelium may enter the circulation and contribute to serum basal CRP levels. If this extra-hepatic CRP expression is regulated in a different way to hepatic CRP, and if its contribution to total serum CRP is not negligible then this may impact on our ability to identify genetic influences on serum CRP levels. It also has implications for the way we interpret and extend the results of genetic study. For example, genetic variants associated with serum basal CRP

may be irrelevant to our understanding of the role of CRP in atherosclerosis, if it is the local, vascular, production of CRP that is pathogenic, and this is regulated by distinct genetic mechanisms. These questions will be difficult to answer using currently available techniques.

### *8.3.3: Influence of age and gender on serum levels*

Since age and sex were included as covariates in each of the analyses presented in this thesis it is possible to make some observations regarding their association with the levels of C3 and CRP.

In the CRP analyses presented in chapters 2 and 3, and the C3 analysis presented in chapter 7 higher serum protein levels were observed in women. It is interesting that this relationship was not observed for the cohort presented in chapter 5, perhaps because these individuals were generally of an older, post-menopausal, age. Higher serum CRP in females was also observed in the majority, but not all (Crawford *et al* 2006 being the notable exception) of the CRP studies presented in **table 1.2**, although none of the authors commented upon this specifically. It has certainly been noticed that oestrogen hormone replacement therapy increases CRP levels in post-menopausal women (Cushman *et al* 1999; Pradhan *et al* 2002). Circumstantial evidence would therefore suggest a genuine gender influence on serum basal CRP. I can find little evidence that this gender effect has been explicitly recognised before for serum C3, although oestrogen response elements have been demonstrated in the C3 promoter (Fan *et al* 1996). The presence of higher levels of inflammation associated proteins in women appears paradoxical given both the links between atherosclerosis and inflammation, and the reduced risk of cardiovascular disease in pre-menopausal women. It is certainly an observation that supporters of a causative role for CRP in the pathogenesis of atherosclerosis need to explain.

It should also be noticed that both basal CRP and C3 increased with age, although this is perhaps not surprising and probably reflects the increased burden of subclinical inflammatory pathology in an elderly population. On the other hand when we looked at high level CRP in the context of rheumatoid arthritis (chapter 6) we observed a different relationship, with lower acute-phase CRP levels overall in older patients. This is perhaps a reflection of the inability of the senescent liver to mount the high levels of acute-phase protein expression in comparison with younger individuals.

#### *8.3.4: Heritability*

In this thesis I have provided heritability estimates for both basal CRP (27.7%) and C3 (39.6%): what conclusions can be drawn about these estimates?

As I outlined in section 1.3.3 a number of other authors have estimated CRP heritability. I have summarised these existing reports in **table 8.1**. It can be seen that my heritability estimate is towards the lower end of the range of existing estimates and there are probably a number of reasons for this. Firstly twin studies inherently tend to give higher estimates of heritability than family studies. The basic premise of a twin study is that dizygotic and monozygotic twins are raised in an identical shared environment. This, however, is unlikely to be entirely true, with many arguing that intrauterine and external environmental factors are likely to be more closely shared between monozygotic compared with dizygotic twins (Phillips 1993). Twin-based heritability studies may therefore over-estimate the genetic component of trait variability (Harrap & Hopper 2001; Hopper 2000). Family studies may be better at identifying environmental influences on the trait of interest; these are revealed when different trait correlations are observed between different between groups of individuals with

identical genetic relatedness (e.g. parent vs. offspring compared with sib-pair). On the negative side, however, when the number of individuals available for study in a family cohort is relatively small, the uncertainty associated with the estimates of environmental and genetic effect can be large.

**Table 8.1.** Current CRP heritability estimates.

| Study                       | Cohort design                                                       | Heritability (%) |
|-----------------------------|---------------------------------------------------------------------|------------------|
| Pankow <i>et al</i> 2001    | European-American two generation pedigrees                          | 40               |
| Vickers <i>et al</i> 2002   | White European two generation pedigrees (proband with hypertension) | 39               |
| Austin <i>et al</i> 2004    | Japanese-American two generation pedigrees                          | 27               |
| MacGregor <i>et al</i> 2004 | UK female twins                                                     | 52               |
| Lange <i>et al</i> 2006a    | Mixed ethnicity sibships (multiplex for type 2 diabetes mellitus)   | 39               |
| Saunders and Gulliford 2006 | White UK two generation pedigrees                                   | 10               |
| Wessel <i>et al</i> 2007    | European American twins                                             | 56               |
| Fox <i>et al</i> 2008       | African-American two generation pedigrees                           | 45               |

It is, perhaps, pedantic to argue which heritability estimate is “most accurate”, as each will be specific to the population under study in any case. It is probably sufficient to say that somewhere between 25% and 50% of the variation in basal CRP on a population basis can be attributed to genetic influences.

In the case of serum C3, I am unaware of any previous attempts to calculate heritability for this trait. Certainly the estimated 39.6% heritability places it in a similar range to that

estimated for CRP. It is also similar to estimates for another acute-phase protein, fibrinogen (20-48%), suggesting all these acute-phase reactants may show similar levels of heritability (de Lange M. *et al* 2001; Friedlander *et al* 1995; Hamsten *et al* 1987).

The general message that should be taken from the heritability estimates is that the genetic component of variation of both serum CRP and C3 is of a sufficient magnitude that the search for genetic variants accounting for this genetic effect is not only feasible, but also important if we are to maximise the use of these proteins as markers of human disease.

#### **8.4: *Cis*-acting Effects on Basal CRP – Discussion**

Each of the *CRP* studies presented in this thesis, along with multiple studies from other authors have demonstrated an association between *CRP* genetic variants and serum basal CRP. The evidence for a *cis*-acting genetic effect is now overwhelming. The question that remains is which variant, or variants are likely to have a true functional role. As I discussed in section 1.5.3, the final stage in the pathway of functional SNP identification must be functional experimentation to confirm that the SNP can alter the phenotype of interest. However, without a clear candidate identified on the basis of genetic study, a functional investigation of multiple SNPs is likely to be time consuming, expensive and, in many cases, futile. **Table 1.2** summarised SNPs reported to be associated with serum basal CRP in the current literature. It highlights that multiple SNPs have been associated with serum CRP but none are consistently associated in each study where they have been tested. The three *CRP* studies presented in this thesis also highlight different genetic variants so, before we can hypothesise which variants are most likely to have a functional role, it is important to consider why this is the case.

#### 8.4.1: Power

In the sections below I explore specific reasons that may explain the multiple, but inconsistent, associations observed at *CRP*. All these explanations hinge on the issue of study power (Chasman *et al* 2009; Gordon & Finch 2005). The critical factors determining the power of a study to demonstrate a direct genetic association are the effect size (which for a quantitative outcome can be expressed as the proportion of phenotypic variance explained by the genotypes at that SNP), the minor allele frequency of the SNP, and the study sample size. When considering an indirect association (which the majority of studies will be identifying), the degree of LD between marker SNP and functional SNP, along with the relative difference in minor allele frequency between the two also become important considerations (Zondervan & Cardon 2004). There are, of course, many other factors influencing the power of a genetic study, in particular sources of random measurement error resulting in misclassification of either genotype or phenotype. These aspects have been discussed in detail in the various data chapters.

#### 8.4.2: Strong LD across *CRP* in European populations

It is clear from the data presented in **figure 3.1** and **table 3.3** that the LD across *CRP*, in white European populations, is very strong and extends to encompass large upstream and downstream regions. Even in the African-American population we observe relatively high levels of LD, with  $D'$  scores approaching one for the majority of pairwise marker comparisons (**table 5.3**). This provides a simple explanation for the large number of genetic markers identified in different studies: clearly the higher the LD across a locus, the easier to demonstrate an indirect association. In the European population it is also notable that whole clusters of SNPs in high LD also have highly correlated minor allele frequencies, as

demonstrated by  $r^2$  values approaching one. In this situation it is difficult to identify, in all but the largest studies, which, out of a cluster of correlated SNPs, is likely to be a functional SNP. Any observed difference is just as likely to be a consequence of minor differences in genotyping success/accuracy as it is to represent a true difference in the degree of association. This will be an even greater problem where interest focuses on a SNP with relatively low minor allele frequency such as *rs1800947* (MAF 7% in Europeans).

#### *8.4.3: Ethnic differences in CRP gene structure*

As in all areas of genetic study there is a clear bias in published *CRP* studies towards cohorts of a white European ancestry, making the data presented in this thesis of particular interest. Of the more extensive existing *CRP* genetic studies only five are in African-American populations (Carlson *et al* 2005; Crawford *et al* 2006; Lange *et al* 2006b; Lee *et al* 2009; Szalai *et al* 2005c), two in Hispanic Americans (Crawford *et al* 2006; Lee *et al* 2009), one in Japanese (Morita *et al* 2006) and one in Asians/Pacific Islanders (Lee *et al* 2009). This latter group I suspect represents a genetically rather heterogeneous population.

The pattern that emerges from these studies and, in particular, the trans-ethnic mapping studies I present in this thesis is that there is increased genetic diversity in populations of primarily African origin. The overall LD across *CRP* remained strong in my African-American population, to the extent that the physical size of the haplotype block was similar to that seen in the white European cohort (see **figure 3.1** and **figure 5.1**). There was, however, an increased diversity of haplotypes in the African-American cohort and some quite notable differences in allele frequency for a number of markers. For example the frequency of the “A” allele of *rs3091244* in African-Americans is over four times that observed in Europeans and there are additional SNPs that appear unique to the African-American population such as

*rs3093058* and *rs3093063*. Conversely, Europeans appear enriched for other alleles which are over twice as common in this group compared with African-Americans: this group includes *rs1800947*, *rs1417938* and *rs3093077*. These findings are not surprising as we would predict that the more ancient African genome (and the related African-American genome) would offer the most genetic diversity; indeed this is illustrated by the LD architecture of Yoruba subjects from Nigeria, which is more diverse and shows more evidence of recombination than reference populations of European and East Asian ancestry (Gabriel *et al* 2002).

This difference in genetic structure between populations means that we have increased power to detect effects arising from some, more frequent, alleles, but decreased power to detect effects arising from other, less frequent, alleles. If, for example, there were two functional SNPs at *CRP*, with divergent allele frequencies in two different ethnic populations, studies could find an association with one SNP in one population and the other SNP in the second population, even though the true picture may be two individual functional effects. The advantage of an analysis in an African population, however, is that we are less likely to report multiple indirect associations, because of the weaker overall LD.

#### *8.4.4: Problems with a tri-allelic SNP*

SNP *rs3091244* has three alleles, implying that two separate mutational events have occurred at this single site. The rate of mutation of the human genome is so low (approximately  $1.8 \times 10^{-8}$  mutations/bp/generation, but varying according to genomic location through processes that are poorly understood) that the presence of a tri-allelic SNP is an exceptional finding (Baer *et al* 2007; Kondrashov 2003). It also presents problems with analysis. Again there is the issue of power because the population is subdivided on the basis of three rather than two alleles. This is a particular problem when genotype models are used as there are six potential

genotypes and only small numbers of individuals will be homozygous for one or other of the minor alleles. Some authors overcome this by pooling the minor alleles and assessing the effect of the common “C” allele acting as an additive effect over the rare “A” and “T” alleles. However, should one, but not both, of the minor alleles be functional then this approach will result in a significant loss of power to detect the true effect. Alternatively if all possible models are assessed (i.e. “C” vs. “A+T”, “T” vs. “C+A” and “A” vs. “C+T”) then this is going to entail a significant penalty in terms of correcting for the testing of multiple hypotheses. Most authors do not report the testing of all possible models, although one would suspect most perform unreported testing.

The model choice analysis I present in chapter 6 appears to be the only serious attempt to address this issue. A key aspect of this approach is that it allows the selection over models containing *rs3091244* encoded in each possible way, thus providing a real advantage in identifying the most likely functional alleles.

#### *8.4.5: Potential for problems with multiple functional SNPs at CRP*

It is quite possible that *CRP* gene expression is influenced by multiple functional SNPs. To give an example: at the SLE disease susceptibility gene *IRF5*, two, or possibly even more, functional effects are felt to mediate susceptibility to disease (Kozyrev & Alarcon-Riquelme 2007). In the context of *CRP* the conventional approach of testing each genotyped SNP in turn for an association with CRP level may lead to erroneous results in the presence of multiple SNP effects, particularly if the different effects are carried on different *CRP* haplotypes. For example, imagine a locus with 5 haplotypes, A to E. If two different functional SNPs with a minor allele frequency of, say, 5%, are carried on haplotypes D and E, then a non-functional SNP with minor allele frequency of 10% carried on both haplotypes

D and E would show a stronger association than either of the true functional variants (which may indeed not show association themselves due to lack of power). An explanation such as this may also underlie the inconsistency in existing *CRP* studies.

#### *8.4.6: CRP functional SNPs – A hypothesis based on thesis data.*

It can be seen from the discussion in previous sections that there are a number of potential explanations for the lack of clarity over the key *cis*-effects in the current *CRP* literature. To some extent these problems remain despite the work presented in this thesis, although there is now probably enough circumstantial evidence, from my studies and other published data, to formulate a working hypothesis to guide further study.

I propose, on the basis of my analysis in chapter 5 that *rs3091244* is a strong candidate for being a functional SNP at this locus. The C allele of *rs3091244* is commonest in all studied human populations, is found on the Chimpanzee and Macaque genome reference sequences, and has not been identified as polymorphic in these species (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007; The Chimpanzee Sequencing and Analysis Consortium 2005). My cladistic analysis in Filipinos also gave highest outgroup weighting probability to *CRP* haplotypes carrying the C allele. It is therefore likely the C allele is ancestral, and a key functional moment in the evolution of the human *CRP* locus was the mutation of the *rs3091244* alleles from C to T, thus increasing basal serum CRP expression. We can only hypothesise how this enhanced expression may have conferred an advantage to early humans. Perhaps basal CRP has a surveillance role in protecting against the early stages of bacterial infection; alternatively higher basal CRP may enhance clearance of apoptotic debris and thus protect against developing autoimmunity. Another explanation, although untested, is that *rs3091244* also alters the magnitude or time-course of the acute-phase CRP

response. This may have additional functional consequences and is discussed further elsewhere.

SNP *rs3091244* is also an attractive candidate because it alters the CAnnTG E-box consensus sequence centred 497 bp upstream from the transcriptional start site. Szalai *et al* purport to have evidence from electrophoretic mobility shift assay (EMSA) for differential transcription factor binding associated with the “T” allele at this site (Szalai *et al* 2005c). Unfortunately they do not actually present their data, stating “EMSA targeting E-box 2 yielded discernable but very weak mobility shifts only with the -390T version (data not shown)”. One can only assume that the data was indeed weak, but these experiments would be worth revisiting in the light of my data. Another attempt to delineate key *CRP* SNPs used *in vitro* reporter assays with *CRP* promoter constructs varying at known SNP sites (Carlson *et al* 2005). As I have already eluded to in section 8.3.1., it is difficult to know how good a physiological representation of basal *CRP* expression these models are, particularly as the level of expression from the native *CRP* promoter in the absence of IL-6 is very small. The constructs contained several SNPs and there was no convincing segregation of promoter activity based exclusively on one SNP, making it difficult to draw firm conclusions from this study. Nonetheless the published results would certainly not be inconsistent with a functional role for the T allele of *rs3091244*. It has been noted that there are frequently discrepancies between genetic associations and *in vitro* functional data and, as it stands, the existing functional data is by no means conclusive (Ioannidis & Kavvoura 2006).

While my data suggested a rather simple model of genetic effect at *CRP* in African-Americans, we need to explain why other SNPs are consistently associated with *CRP* in other ethnic groups (including the European and Filipino analyses presented in chapters 3 and 4). In addition we need to explain why the T vs. C+A model for *rs3091244* is not the

consistently best association in these other studies. Assuming that *rs3091244* is a genuine functional effect then part of the explanation may be related to power and LD issues as discussed in previous sections. The degree of LD with the T allele of *rs3091244* in African-Americans, as measured by  $r^2$ , was 0.138 for *rs1205*, 0.225 for *rs1417938*, 0.007 for *rs1800947*, and the equivalent in Europeans was 0.256 for *rs1205*, 0.990 for *rs1417938*, and 0.035 for *rs1800947*. The association with these additional SNPs in Europeans may therefore be influenced by both their higher frequency and their stronger LD with *rs3091244*(T), with the true effect only becoming apparent in this African American cohort.

I think, however, the most likely explanation here is that there is more than one functional SNP at *CRP*, and that this additional variant (or variants) is only found with sufficient frequency to be identified in non-African populations. This has, in fact, been formally proposed in a recent meta-analysis of studies published up to 2007 (Verzilli *et al* 2008). Bayesian model selection was used to identify variants exerting strongest independent effect on serum CRP. The model with the highest posterior probability (0.22) contained no less than four SNPs, although the inclusion of *rs1800947* was rather dependant on the prior on the model space. SNPs *rs1205* and *rs3093077* were included in all models with a posterior probability  $>0.02$ . Interestingly *rs3091244* was included in the meta-analysis but did not appear in the top models, probably (although it is actually not specified in the paper) because the C vs. T+A coding was used (in common with the majority of papers incorporated into the analysis).

In their recent genome wide association study, Ridker *et al* performed backward selection models (using both maximum likelihood (AIC) and Bayesian (BIC) models) to identify a minimal set of variants from their GWA study accounting for CRP variability (Ridker *et al* 2008). Although the authors do not provide a detailed methodology (e.g. specifying prior

distributions for the Bayesian models) it appears that even under stringent model conditions a minimum of four *CRP* variants account for the observed *CRP* variation. Interestingly *rs3091244* was included in this group even though it was encoded as G vs. A+T (their assay presumably operating on the opposite strand). It is possible that the size of this cohort was so large that a significant association could still be observed with the inclusion of a non-functioning, low frequency A allele in with the T allele group. Kathiresan *et al* also performed selection models (forward stepwise selection) using the 13 *CRP* SNPs typed in their study (Kathiresan *et al* 2006). Their best model contained one SNP only, *rs3091244*, with the next best model containing two SNPs (*rs1205* and *rs2808630*). They do not, however, provide an analysis such as an AIC to enable us to assess the weight of evidence for their assertions. Finally I should also re-iterate the evidence from my Filipino analysis that there was some evidence for a separate functional effect arising from *rs1800947*. This has been discussed in section 4.4.2.

Given the evidence for multiple SNP effects in non-African populations is there any evidence to highlight functional candidates other than *rs3091244*? In addition to the cross-species conservation analysis that I have presented in **figure 1.4**, I also used three publically available, web-based SNP assessment tools (PupasView (<http://pupasuite.bioinfo.cipf.es>), SNPselector (<http://primer.duhs.duke.edu/>) and TAMAL (<http://neoref.ils.unc.edu/tamal>)) to look for additional evidence of function for any known common *CRP* SNPs (Bhatti *et al* 2006). None were found to lie within known regulatory sequences or to show significant cross-species conservation. Although there are no outstanding candidates on the basis of bioinformatics, two SNPs probably stand out above the rest.

*CRP* does have a notably long 3'-UTR (1238 base pairs), raising the possibility, in theory at least, that control of mRNA degradation, transport or translation is an important regulatory

mechanism for CRP (Chen *et al* 2006). The location of *rs1205* within the 3'UTR therefore makes it a reasonable functional candidate, although there is currently no direct evidence for this. SNP *rs1800947*, as a synonymous variant, could also be considered a potential functional candidate. Recent studies suggest synonymous base pair changes can have functional consequences as the substitution of rare for common codons leads to a delay in ribosomal translation and subsequently alters protein folding; whether this mechanism could also quantitatively alter protein levels is unknown (Tsai *et al* 2008). SNP *rs1205* appears in all the best models proposed in the Bayesian meta-analysis, while *rs1800947* was also in the top model and appears associated with basal CRP levels in many existing studies (Verzilli *et al* 2008).

Ultimately, further functional work will be required to determine which SNP or SNPs truly influences gene expression. A combination of assessing CRP promoter variants by an *in-vitro* reporter assays of promoter mutants (say a luciferase based system), combined with an assessment of transcription factor binding (by EMSA or chromatin immunoprecipitation) seems reasonable. These approaches have been attempted (see above) but crucially these studies introduced multiple promoter mutants rather than considering a single variant, thus adding unnecessary complication. Also of interest would be to move away from hepatocyte cell lines to other cells reported to express CRP.

#### *8.4.7: Cis-effects explain a small proportion of basal CRP variance.*

Although there remains some uncertainty over the identity of the functional SNP or SNPs at *CRP*, all studies, including those presented in this thesis, agree that the proportion of CRP variance explained by known *CRP* variants is small (**table 8.2**). Most authors assess this through the partial  $r^2$  determined through linear regression analysis. My UK white European

family analysis has been the only attempt to address this through a variance components approach, suggesting *CRP* haplotypes accounted for 3.77% of basal *CRP* variability. Given the estimated 28% *CRP* heritability this therefore represented only 15% of the total genetic effect. It is possible, of course, that the mode of association between *CRP* SNPs and serum *CRP* is more complex than the simple models that have been currently analysed. In particular, genotype-environmental interactions may be important, in particular environmental factors acting as inflammatory stimuli (leading us back to the discussion of “basal” vs. “acute-phase” *CRP* in section 8.3.1).

**Table 8.2:** Proportion of *CRP* variance explained by cis-acting effects in current studies

| Study                                 | Proportion of Variance |
|---------------------------------------|------------------------|
| UK SLE families (chapter 3)           | 3.8%                   |
| African-American analysis (chapter 5) | 5.2%                   |
| Miller <i>et al</i> 2005              | <5%*                   |
| Crawford <i>et al</i> 2006            | <3%*                   |
| Kathireson <i>et al</i> 2006          | 1.4%                   |
| Lange <i>et al</i> 2006b              | <6.4%*                 |
| Ridker <i>et al</i> 2008              | 3.4%                   |

\* Where more than one cohort is presented in a publication the largest proportion of variance is presented.

There may also be important epistatic effects, or alternatively there may be a more complex relationship between genotype and serum *CRP*, perhaps with SNPs acting through effects on the variance of *CRP* rather than the mean (Chasman *et al* 2009). Unless these other effects are identified, however, we must assume that a significant proportion of the variance in population serum basal *CRP* is explained by *trans*-acting genetic effects. These will be discussed further in the next section.

## 8.5: *Trans*-acting Effects on Serum Basal CRP – Discussion

As outlined above, *cis*-acting variation appears to explain only a small proportion of total serum CRP variability; this suggests that we must look at variation elsewhere within the genome to explain the remaining genetic effect.

### 8.5.1: *Trans*-acting effects presented in this thesis

When the work presented in this thesis was planned, existing literature on potential *trans*-acting genetic effects focussed on three loci, *APOE*, *IL6* and *IL1B*, and I therefore concentrated on genotyping at these loci. My aim by including this additional genotyping was to increase the power to detect effects arising from *CRP* itself by reducing overall variance due to additional genetic influences. I did not primarily intend a thorough evaluation of these additional loci, but nonetheless some interesting patterns did emerge.

A number of existing studies report an association between variants at *IL6* and *IL1B* and serum CRP (reviewed in section 1.3.3), and given the *in vitro* evidence for the role of these two cytokines as key regulators of CRP expression it is perhaps not surprising that some significant associations have been reported between serum CRP and variants at these two loci.

In both the UK white European cohort and the Filipino cohort we observed an association between *IL6* variant *rs1800796* and basal CRP expression, although interestingly only when modelled as an interacting term with SLE. No such association was observed in the African-American cohort analysis. Neither did we observe an association between the commonly

reported *rs1800795* variant and serum CRP. The *rs1800796* variant was particularly common in our Filipino population (minor allele frequency 0.33), mirroring previous studies in Chinese and Korean populations (Jang *et al* 2008; Paik *et al* 2007; Wong *et al* 2007). The data presented in this thesis therefore add to the inconsistency of *IL6* SNP associations reported in the literature. There are probably a number of explanations for this, many of which would be similar to those already discussed in relation to *CRP*.

I think a key issue is whether study participants are expressing basal or acute-phase levels of CRP. I would argue that the reason why I only observed an association between *rs1800796* and serum CRP in the context of SLE is because these are exactly the individuals mounting a low-grade acute-phase response. I suspect that the association between *IL6* variants and serum CRP is strongly dependant on the level of acute-phase response being mounted. Additional evidence along these lines can be found from a recent GWA study (discussed below in 8.5.2) where the association between *IL6R* variants and serum CRP weakened when only individuals with low level CRP were included (Ridker *et al* 2008). As I have already highlighted in section 1.3.2., numerous *IL6* genetic studies are conducted in cohorts of individuals with diseases such as CHD and diabetes. These are exactly the individuals in which there is going to be considerable ambiguity over what we are really measuring: a true basal CRP or a low grade, inflammatory acute-phase CRP. This issue will really only be resolved by a study in unambiguously healthy young individuals to assess the association with basal CRP, or a study in individuals with unambiguous active inflammatory disease to assess the acute-phase CRP association (as I have attempted in the study presented in chapter 6).

I also suspect that we have not yet fully delineated functional variants in *IL6*. An early *in vitro* functional study suggested that the role of *IL6* promoter variants in determining

transcriptional activity was not simple, but depended on a haplotype effect (Terry *et al* 2000). Recent genetic studies have mapped the *IL6* promoter region in more detail, and have begun to consider more complex functional effects arising from *IL6* promoter haplotypes. Certainly extensive promoter haplotypes have been associated with disease, although the association between haplotypes and serum CRP has again been inconsistent in current studies asking this question, highlighting the need for cohort consistency (Fife *et al* 2005; Walston *et al* 2007; Yanbaeva *et al* 2009). Interestingly *IL6* regulatory elements have also been identified at a considerable distance upstream from the gene, suggesting a more extensive search for functional variants may be also fruitful (Samuel *et al* 2008).

To some extent the findings of existing *IL1B* studies mirror those for *IL6*, with early studies highlighting two SNPs (*rs1143634* and *rs16944*) that demonstrate association with serum CRP, although often only in certain cohort subsets, or in certain ethnic groups identified in secondary analyses (section 1.3.3). I did not observe an association between either of these SNPs and serum CRP in any of the studies I present in this thesis. I did observe a fairly weak association between variant *rs7596461* and serum CRP in the UK SLE family cohort. This SNP has not specifically been genotyped in other studies, so this is a preliminary association that will clearly need replicating. As with *IL6* there have been some recent attempts to use more extensive genotyping efforts to analyse the association between *IL1B* haplotypes and serum CRP. Once again these reports are inconsistent emphasising the need for replication and cohort standardisation (Enquobahrie *et al* 2009; Rogus *et al* 2008).

Our observed association between *APOE* variation has now been confirmed in two large recent genome wide association study (discussed in more detail in section 8.5.2) (Elliott *et al* 2009; Ridker *et al* 2008). Interestingly, in both studies, this association was observed for SNPs closely flanking the *APOE* gene, and not those defining Apo  $\epsilon$  serotype (which were

not genotyped). Unfortunately the Apo  $\epsilon$  defining SNPs are not included in the HapMap dataset either, so it is impossible to infer whether the association observed in the GWA study is likely to be independent of the effect of Apo  $\epsilon$  type. Chasman *et al* also typed additional, non-serotype defining *APOE* SNPs but observed no additional association (Chasman *et al* 2006). They also measured Apo  $\epsilon$  levels in their large cohorts and found that adjusting for Apo  $\epsilon$  levels made little difference to the strength of association between Apo  $\epsilon$  serotype and serum CRP, suggesting it was genuinely a qualitative rather than a quantitative difference in Apo  $\epsilon$  that was influencing CRP expression. The true nature of this association can only be resolved by large scale studies involving extensive *APOE* mapping, quantitation of Apo  $\epsilon$  levels and, ultimately, better functional characterisation of the Apo  $\epsilon$  variants.

#### 8.5.2: Recent GWA studies

With recent technological advances that now permit the simultaneous genotyping of over 100000 SNPs, a whole genome association study, with serum CRP as a quantitative outcome measure, is the ideal way forward in the identification of novel *trans*-acting variants.

Although the UK SLE cohort which I studied in detail has also been incorporated into existing whole-genome association projects with SLE as the outcome of interest unfortunately a pooled-sample methodology was used, which did not permit the identification of individual patient genotypes for a *post-hoc* analysis of serum CRP association.

Since the experimental work presented in this thesis has been completed there have now been four GWA studies published for serum CRP; an example of how rapidly work in this field is progressing. Since the results of these studies both complement and extend my findings, it is worth considering each of them in more detail (data summarised in **table 8.3**). The first GWA study, published in late 2007, performed a family based analysis on 1012 related members of

the Framingham offspring cohort (Benjamin *et al* 2007). The relatively small size of this cohort, combined with the family-based design resulted in a rather under-powered study. In addition it used one of the earlier SNP chips, the Affymetrix 100K gene chip, which, at a genome-wide level, may leave significant gaps in genetic coverage. Other than confirming a *cis*-effect from *CRP* no further known genes were identified with genome-wide significance.

**Table 8.3.** Genes associated with CRP expression in current GWA studies (*P*-values shown)

| Gene/Locus*  | <i>P</i> -value†       |                       |                        |                        |                       |
|--------------|------------------------|-----------------------|------------------------|------------------------|-----------------------|
|              | Benjamin <i>et al.</i> | Melzer <i>et al.</i>  | Ridker <i>et al.</i>   | Sabatti <i>et al.</i>  | Elliott <i>et al.</i> |
| <i>CRP</i>   | $2.83 \times 10^{-8}$  | $6.36 \times 10^{-6}$ | $6.16 \times 10^{-28}$ | $2.92 \times 10^{-22}$ | $7.6 \times 10^{-44}$ |
| <i>LEPR</i>  |                        |                       | $6.52 \times 10^{-21}$ | $3.76 \times 10^{-7}$  | $3.1 \times 10^{-14}$ |
| <i>IL6R</i>  |                        |                       | $1.93 \times 10^{-8}$  |                        | $1.8 \times 10^{-14}$ |
| <i>GCKR</i>  |                        |                       | $6.73 \times 10^{-15}$ |                        |                       |
| <i>HNF1A</i> |                        |                       | $6.75 \times 10^{-17}$ |                        | $1.3 \times 10^{-30}$ |
| <i>APOE</i>  |                        |                       | $8.70 \times 10^{-21}$ |                        | $4.5 \times 10^{-27}$ |
| <i>LEF1</i>  |                        |                       |                        | $2.71 \times 10^{-11}$ |                       |
| 12q23.2      |                        |                       | $1.16 \times 10^{-10}$ |                        |                       |
| 5q23         | $1.2 \times 10^{-6}$   |                       |                        |                        |                       |

† Unadjusted *P*-values are shown. Where studies demonstrate more than one SNP associated at a locus, the smallest *P*-value is shown. Study methodologies differed and only associations judged to have achieved genome-wide significance by the methodology of that paper are shown.

\* Genes encode known proteins: *CRP* – C-reactive protein; *LEPR* – Leptin receptor; *IL6R* – Interleukin-6 receptor; *GCKR* – Glucokinase regulatory protein; *HNF1A* – Hepatic nuclear factor 1a; *APOE* – Apolipoprotein E; *LEF1* – Lymphoid enhancer binding factor 1.

The second study, from May 2008 was in a relatively small cohort of 1200 white European participants; primarily older individuals (Melzer *et al* 2008). Using genotyping from the Illumina HumanHap 500 chip, the authors looked for an association between genotype and 42 biological proteins. The large number of genotypes and proteins tested in this analysis led the authors to select a very high level of genome-wide significance ( $2.4 \times 10^{-9}$ ), and they limited their analysis to SNPs with a minor allele frequency >10%. Again the *CRP* *cis*-effect was confirmed but no significant *trans*-acting variants were identified. Published in the same

month was a much larger, well designed study in 6345 healthy women from the Womens Genome Health Study) (Ridker *et al* 2008). The Illumina HumanHap 300 chip was used to generate 336,108 genotypes, and all typed SNPs with a minor allele frequency >1% were included for analysis. In addition to *CRP*, a number of *trans*-acting effects were also identified. *APOE* was confirmed as an important *trans*-acting locus as discussed above in section 8.5.1. Very strong associations were also seen between *LEPR* variants and serum CRP, confirming the weak association seen in an earlier candidate gene study (Zhang *et al* 2007). Associations were also seen between serum CRP and variants in *HNF1A* (encoding one of the key CRP transcription factors: see section 1.3.2), *GCKR* (encoding the Glucokinase Regulatory Protein) and *IL6R* (encoding the Interleukin 6 receptor). The identified association between serum CRP and *HNF1A* were confirmed in a separate paper, published simultaneously, which followed up the results of a small whole-genome scan (980 individuals) with fine mapping of identified candidate genes in further cohorts of white European ancestry (Reiner *et al* 2008). The most recent GWA study analysed 329,091 genotypes generated using the Illumina 370cnvDUO array, in a population of 4,763 ethnically homogenous individuals from Northern Finland (Sabatti *et al* 2009). A number of quantitative metabolic traits, including CRP were analysed. Along with the now extensively replicated *CRP cis*-effect, the *LEPR* association was replicated, and a new association was identified with the *LEF1* gene, encoding the Lymphoid Enhancer Binding Protein. The largest and most recent GWA study was conducted in a cohort of 17,967 participants (Elliott *et al* 2009). This cohort was homogeneous, containing individuals of Scandinavian, Western European and Asian Indian ethnicity, some of whom has CHD or metabolic syndrome, although this was accounted for by use of a principal components analysis. Replication of top hits was performed in a cohort of 13,615 individuals free from heart disease. Genotyping was by either Illumina 500K or Illumina 317K chip with additional SNP imputation across the whole dataset resulting in approximately 1.4 million direct or imputed genotypes per person.

Results largely mirrors those of the Ridker et al study, with the key genes associated with CRP level being *CRP*, *LEPR*, *IL6R*, *HNFI1A* and *APOE*.

**Table 8.4** highlights the actual CRP SNP found most associated with CRP levels in these genome wide analyses.

**Table 8.4.** Most significant SNPs at *CRP* in the current GWA studies.

| Study          | Key SNP    | Comments                                                      |
|----------------|------------|---------------------------------------------------------------|
| Benjamin et al | rs2794520  | Directly genotyped. In strong LD with rs1205                  |
| Melzer et al   | rs12093699 |                                                               |
| Ridker et al   | rs3091244  | A hand-picked SNP not on GWA chip (C vs. A/T coding reported) |
| Sabatti et al  | rs2794520  | Directly genotyped. In strong LD with rs1205                  |
| Kooner et al   | rs7553007  | An imputed SNP but in strong LD with rs1205                   |

It should be emphasised that the *CRP* coverage on genome-wide arrays is relatively sparse and does not include the key SNPs commonly reported throughout the literature i.e. *rs1205*, *rs3091244* or *rs1800947* (The 8 SNPs on the HapMap 500 panel being *rs2808660*, *rs2592900*, *rs1146973*, *rs12744244*, *rs12083699*, *rs2592887* and *rs2794520*). GWA at its current level of coverage should not be regarded as a substitute for detailed fine mapping. Indeed the “top hit” reported in the Ridker *et al* study (*rs3091244*) was not actually on the GWA chip, but was part of a panel of additional hand-selected SNPs for genotyping. The “top-hit” reported in the Elliott *et al* study (*rs7553007*) was actually one of the imputed genotypes. Despite this, it is still possible to see consistency within the GWA literature. It can be seen with reference to **figure 3.1** and **figure 3.2** and **table 3.3** that *rs7553007* and

rs2794520, which are highlighted in the GWA studies, are both in strong LD with each other and both are in strong LD with the commonly reported rs1205 variant: all three SNPs fall on the same two *CRP* haplotypes.

#### 8.5.3: *Trans-acting effects – the current picture and future work*

Recent GWA studies highlight a number of new *trans*-acting genetic influences on basal CRP expression. I think that the function of these novel candidates, within metabolic pathways, provides good evidence that basal CRP is regulated as a distinct non-inflammatory process. The focus of further work on *trans*-acting effects therefore moves away from cytokine genes such as *IL6* and *IL1B*, which, nonetheless, may play an important role in the variation of acute-phase CRP expression. The Ridker GWA study is interesting in that it estimates the proportion of CRP variance explained by each of the identified loci. The total estimated variance explained was 10.1%. This figure will probably be revised upwards when the genetic effect from each of these loci is refined by detailed mapping studies, nonetheless, it suggests we have still only identified approximately half of the total genetic component of CRP variation. Whether or not we are likely to identify further loci is a point of discussion. It is possible that the remaining genetic influence on CRP variation is determined by multiple genes (for arguments sake say 50-100) each of which are contributing only a small amount to total CRP variability. It is possible that we may never identify these weak effects because to have sufficient power to detect the associations would require a cohort several magnitudes larger than those currently available.

## 8.6 CRP, C3 and Disease

While an increasing number of studies report an association between genetic variants and serum CRP, the number of studies reporting *CRP* associations with disease remains small. A key finding from my analysis presented in chapter 3 was my inability to detect an association between *CRP* variants and SLE, although I have argued that our power to detect such an association was actually quite low. Notably the last 12 months has also seen the publication of three GWA studies in SLE cohorts and in none of these did *CRP* variants approach genome wide significance (Graham *et al* 2008; Harley *et al* 2008; Hom *et al* 2008). Although these studies were large, they did not reach the size of cohort used for GWA studies in diseases such as diabetes mellitus, and it could be argued that the power to detect more modest effects was limited. The coverage of markers across *CRP* in these studies was of course also more limited than that employed in candidate gene studies. As I discussed in section 1.3.4, four relatively small scale candidate gene association studies have reported association between *CRP* variants and SLE, but there is no consistency between the identified genetic variants in each of these studies. A link between CRP expression and SLE is theoretically attractive although on the basis of current evidence we must be cautious about whether there is true genetic effect arising from this gene, or at least accept that any effect is likely to be relatively weak.

While none of the studies presented in this thesis addressed the question of association between CRP and cardiovascular disease, it is important to recognise that there remains a paucity of published data suggesting that such an association exists.

As I outlined in section 1.4.3 the most frequently observed association between *C3* variants and disease are for age-related macular degeneration and renal allograft survival (for the

C3F/S variant), and bronchial asthma (for probable non-coding effects). Although there are a number of theoretical reasons why *C3* variation may be of relevance in SLE, published data supporting this is limited. The data presented in this thesis suggests that there was an SLE association, supporting an existing Japanese study but highlighting a different SNP (Miyagawa *et al* 2008). Unfortunately the different ethnicity of the two cohorts makes a direct comparison difficult. Notably, *C3* variants were not highlighted in the three recent GWA studies in SLE, although with the overall LD across *C3* being relatively weak it is likely that the density of genetic coverage offered by these GWA chips was inadequate to convincingly exclude an association at this locus.

## **8.7 Concluding Remarks and Future Work**

We can use genetics to answer two simple questions regarding serum CRP and *C3*: are genetic variants associated with the levels of these two proteins and are variants in these genes associated with disease. More sophisticated questions, however, that are of greater clinical relevance are: does genetics provide evidence for a direct causal role of these proteins in disease and, even if they are not directly causal factors, can we use genetics to improve the predictive power of these proteins as markers for disease. We have seen from the section above that for CRP, the evidence from genetics to implicate physiological levels of basal CRP (or at least circulating serum CRP) as a causal factor in both atherosclerosis and SLE is weak. Despite the size of the recent negative Mendelian Randomisation study, it remains possible that studies are hampered by a lack of power particularly as the influence of individual loci on CRP variability is small (Zacho *et al* 2008). Plans for a collaborative effort to repeat this sort of experiment on a larger scale (at least 37,000 CHD cases and 120,000 controls) and with comprehensive genetic coverage should answer this question definitively (CRP CHD Genetics Collaboration 2008). This is not to say that unmeasured locally acting

CRP within tissues may have a pathological role. It is also important to stress that even if no genetic effect is found, either supra-physiological administration of CRP, or the use of CRP inhibitors may have a role in the treatment of disease. Both CRP and its inhibitor have indeed been shown in early studies to have a potential therapeutic role in SLE and cardiovascular disease (Pepys *et al* 2006; Rodriguez *et al* 2005; Rodriguez *et al* 2006). What is suggested by current genetic studies is that lifelong, genetically determined minor variation in CRP concentration may have a minimal role in the development of disease.

The second question is whether we can make use of the available *CRP* genetic data to improve the utility of basal CRP levels as a predictive biomarker. In section 1.3.5 I commented on calls for CRP to be used as a routine clinical measure of coronary risk (Danesh *et al* 2004; Pearson *et al* 2003). If we are able to understand the genetic influences on basal CRP level then it is possible that the use of a “genetically adjusted” CRP level in these predictive models will provide improved risk estimates. Unfortunately current evidence, including the work presented in this thesis suggests that routinely adjusting CRP for genetic influences will not be practical in the foreseeable future. Current evidence suggests multiple genetic variants both *cis*- and *trans*-acting each account for a small proportion of CRP variability. Even if we were to conclusively identify all the true functional variants, a genetically adjusted CRP would require the genotyping of multiple variants; a prospect that is not currently feasible for routine clinical practice.

For CRP, I think that interest will now focus much more on the genetics of acute-phase CRP expression. Acute-phase CRP has also been built into clinical algorithms to compare disease activity between patients and to predict future clinical events, most notably in the field of rheumatoid arthritis (as discussed in section 1.3.3). A review of the literature over the past six months suggests a rapid growth in the number of clinical scenarios in which acute-phase CRP

has been proposed to have this predictive role. These include: mortality in oral squamous cell carcinoma, spread of gastric carcinoma, risk of recurrence of atrial fibrillation following cardioversion, survival from adult respiratory distress syndrome, progression to joint replacement in rheumatoid arthritis, all-cause hospital mortality, non-relapse mortality following bone-marrow transplantation and prognosis in colorectal cancer determining need for adjuvant chemotherapy after surgery (Artz *et al* 2008; Bajwa *et al* 2009; Khandavilli *et al* 2009; Kim *et al* 2009; Koike *et al* 2008; Kompoti *et al* 2008; Mazza *et al* 2009; Poole *et al* 2008). Although the number of studies looking at acute-phase CRP are limited, the available evidence from these suggest that the role played by genetics in determining the magnitude of acute-phase CRP may be relatively strong. Despite the problems faced by all these studies in quantifying the magnitude of inflammatory stimulus and hence reducing random error associated with differing levels of inflammatory activity (see section 6.4.3) all studies were able to demonstrate significant genetic associations. In addition there appears to be some consistency in the literature, particularly surrounding the association of *rs1800947* with lower acute-phase CRP and, probably, with *rs3091244* and higher CRP. Studies of acute-phase CRP may be harder to design, but the usefulness of a genetically adjusted acute-phase CRP level is likely to impact across a wide range of medical fields. With this in mind I have already established a further collaboration with the aim of replicating my own acute-phase CRP findings. The new cohort consists of a further 500 patients with rheumatoid arthritis. This cohort has the added benefit of having multiple paired CRP/ESR measurements for each patient, which will considerably increase the power of the study, hopefully confirming and strengthening the preliminary finding I presented in chapter 6.

While the clinical measurement of serum C3 is more restricted, it has a particular role in the monitoring of disease activity in systemic lupus erythematosus. Similar to CRP, it has also been reported to predict risk of future clinical events, namely hypertension, future risk of

diabetes mellitus and outcome in IgM nephropathy (Engstrom *et al* 2005; Engstrom *et al* 2007; Myllymaki *et al* 2006). The study of the genetics of C3 expression and the relationship between common C3 variants and susceptibility to SLE is at a preliminary stage, consisting of the data presented in this thesis and one other study. Unfortunately I was unable to replicate the existing Japanese study findings that a single C3 SNP influences both C3 levels and disease susceptibility, although my findings were not inconsistent with genetic effects reported for other diseases. Clearly further work is required in this field. For the study of genetic effect on serum C3 analysis of a cohort of individuals without SLE would probably provide the most unambiguous results. For the genetic association with disease a detailed fine-mapping study in alternative large SLE cohorts will be required.

## **REFERENCES**

Abecasis, G. R., Cardon, L. R., & Cookson, W. O. (2000) A general test of association for quantitative traits in nuclear families. *Am.J.Hum.Genet.* 66: 279-292.

Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat.Genet.* 30: 97-101.

Agrawal, A., Samols, D., & Kushner, I. (2003) Transcription factor c-Rel enhances C-reactive protein expression by facilitating the binding of C/EBPbeta to the promoter. *Mol.Immunol.* 40: 373-380.

Agrawal, A. & Volanakis, J. E. (1994) Probing the C1q-binding site on human C-reactive protein by site-directed mutagenesis. *J.Immunol.* 152: 5404-5410.

Al-Sharif, W. Z., Sunyer, J. O., Lambris, J. D., & Smith, L. C. (1998) Sea urchin coelomocytes specifically express a homologue of the complement component C3. *J.Immunol.* 160: 2983-2997.

Alarcon-Segovia, D., Alarcon-Riquelme, M. E., Cardiel, M. H., Caeiro, F., Massardo, L., Villa, A. R., & Pons-Estel, B. A. (2005) Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum.* 52: 1138-1147.

Allen, G., Harvald, B., & Shields, J. (1967) Measures of twin concordance. *Acta Genet.Stat.Med* 17: 475-481.

Alper, C. A., Johnson, A. M., Birtch, A. G., & Moore, F. D. (1969) Human C'3: evidence for the liver as the primary site of synthesis. *Science* 163: 286-288.

Alper, C. A. & Propp, R. P. (1968) Genetic polymorphism of the third component of human complement (C'3). *J.Clin.Invest* 47: 2181-2191.

Altmuller, J., Palmer, L. J., Fischer, G., Scherb, H., & Wjst, M. (2001) Genomewide scans of complex human diseases: true linkage is hard to find. *Am.J.Hum.Genet.* 69: 936-950.

Andrews, E., Feldhoff, P., Feldhoff, R., & Lassiter, H. (2003) Comparative effects of cytokines and cytokine combinations on complement component C3 secretion by HepG2 cells. *Cytokine* 23: 164-169.

Armitage, P. & Parmar, M. "Some approaches to the problem of multiplicity in clinical trials", in *Proceedings of the XIIth International Biometrics Conference*.

Arnaout, M. A., Todd, R. F., III, Dana, N., Melamed, J., Schlossman, S. F., & Colten, H. R. (1983) Inhibition of phagocytosis of complement C3- or immunoglobulin G-coated particles and of C3bi binding by monoclonal antibodies to a monocyte-granulocyte membrane glycoprotein (Mol). *J.Clin.Invest* 72: 171-179.

Arnett, F. C., Edworthy, S. M., Bloch, D. A., McShane, D. J., Fries, J. F., Cooper, N. S., Healey, L. A., Kaplan, S. R., Liang, M. H., Luthra, H. S., & . (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* 31: 315-324.

- Artz, A. S., Wickrema, A., Dinner, S., Godley, L. A., Kocherginsky, M., Odenike, O., Rich, E. S., Stock, W., Ulaszek, J., Larson, R. A., & van, B. K. (2008) Pretreatment C-reactive protein is a predictor for outcomes after reduced-intensity allogeneic hematopoietic cell transplantation. *Biol.Blood Marrow Transplant.* 14: 1209-1216.
- Arvilommi, H. (1974) Capacity of complement c3 phenotypes to bind on to mononuclear cells in man. *Nature* 251: 740-741.
- Austin, M. A., Zhang, C., Humphries, S. E., Chandler, W. L., Talmud, P. J., Edwards, K. L., Leonetti, D. L., McNeely, M. J., & Fujimoto, W. Y. (2004) Heritability of C-reactive protein and association with apolipoprotein E genotypes in Japanese Americans. *Ann.Hum.Genet.* 68: 179-188.
- Aziz, N., Fahey, J. L., Detels, R., & Butch, A. W. (2003) Analytical performance of a highly sensitive C-reactive protein-based immunoassay and the effects of laboratory variables on levels of protein in blood. *Clin.Diagn.Lab Immunol.* 10: 652-657.
- Azumi, K., De, S. R., De, T. A., Rigoutsos, I., Yoshizaki, F., Pinto, M. R., Marino, R., Shida, K., Ikeda, M., Ikeda, M., Arai, M., Inoue, Y., Shimizu, T., Satoh, N., Rokhsar, D. S., Du, P. L., Kasahara, M., Satake, M., & Nonaka, M. (2003) Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". *Immunogenetics* 55: 570-581.
- Baechler, E. C., Batliwalla, F. M., Karypis, G., Gaffney, P. M., Ortmann, W. A., Espe, K. J., Shark, K. B., Grande, W. J., Hughes, K. M., Kapur, V., Gregersen, P. K., & Behrens, T. W. (2003) Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc.Natl.Acad.Sci.U.S.A* 100: 2610-2615.
- Baer, C. F., Miyamoto, M. M., & Denver, D. R. (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat.Rev.Genet.* 8: 619-631.
- Bajwa, E. K., Khan, U. A., Januzzi, J. L., Gong, M. N., Thompson, B. T., & Christiani, D. C. (2009) Plasma C-Reactive Protein Levels Are Associated With Improved Outcome in ARDS. *Chest.*
- Balistreri, C. R., Vasto, S., Listi, F., Grimaldi, M. P., Lio, D., Colonna-Romano, G., Caruso, M., Caimi, G., Hoffmann, E., Caruso, C., & Candore, G. (2006) Association between +1059G/C CRP polymorphism and acute myocardial infarction in a cohort of patients from Sicily: a pilot study. *Ann.N.Y.Acad.Sci.* 1067: 276-281.
- Ballou, S. P. & Lozanski, G. (1992) Induction of inflammatory cytokine release from cultured human monocytes by C-reactive protein. *Cytokine* 4: 361-368.
- Barnes, K. C., Grant, A. V., Baltadzhieva, D., Zhang, S., Berg, T., Shao, L., Zambelli-Weiner, A., Anderson, W., Nelsen, A., Pillai, S., Yarnall, D. P., Dienger, K., Ingersoll, R. G., Scott, A. F., Fallin, M. D., Mathias, R. A., Beaty, T. H., Garcia, J. G., & Wills-Karp, M. (2006) Variants in the gene encoding C3 are associated with asthma and related phenotypes among African Caribbean families. *Genes Immun.* 7: 27-35.
- Barnum, S. R., Jones, J. L., & Benveniste, E. N. (1993) Interleukin-1 and tumor necrosis factor-mediated regulation of C3 gene expression in human astrogloma cells. *Glia* 7: 225-236.

- Barrat, F. J., Meeker, T., Gregorio, J., Chan, J. H., Uematsu, S., Akira, S., Chang, B., Duramad, O., & Coffman, R. L. (2005) Nucleic acids of mammalian origin can act as endogenous ligands for Toll-like receptors and may promote systemic lupus erythematosus. *J.Exp.Med* 202: 1131-1139.
- Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 21: 263-265.
- Bartok, I. & Walport, M. J. (1995) Comparison of the binding of C3S and C3F to complement receptors types 1, 2, and 3. *J.Immunol*. 154: 5367-5375.
- Bassuk, S. S., Rifai, N., & Ridker, P. M. (2004) High-sensitivity C-reactive protein: clinical importance. *Curr.Probl.Cardiol*. 29: 439-493.
- Baudino, T. A. & Cleveland, J. L. (2001) The Max network gone mad. *Mol.Cell Biol*. 21: 691-702.
- Baumann, H. (1989) Hepatic acute phase reaction in vivo and in vitro. *In Vitro Cell Dev.Biol*. 25: 115-126.
- Baumann, I., Kolowos, W., Voll, R. E., Manger, B., Gaipf, U., Neuhuber, W. L., Kirchner, T., Kalden, J. R., & Herrmann, M. (2002) Impaired uptake of apoptotic cells into tingible body macrophages in germinal centers of patients with systemic lupus erythematosus. *Arthritis Rheum*. 46: 191-201.
- Beatty, D. W., Davis, A. E., III, Cole, F. S., Einstein, L. P., & Colten, H. R. (1981) Biosynthesis of complement by human monocytes. *Clin.Immunol.Immunopathol*. 18: 334-343.
- Becker, K. G. (2004) The common variants/multiple disease hypothesis of common complex genetic disorders. *Med Hypotheses* 62: 309-317.
- Bell, S. A., Faust, H., Schmid, A., & Meurer, M. (1998) Autoantibodies to C-reactive protein (CRP) and other acute-phase proteins in systemic autoimmune diseases. *Clin.Exp.Immunol*. 113: 327-332.
- Benjamin, E. J., Dupuis, J., Larson, M. G., Lunetta, K. L., Booth, S. L., Govindaraju, D. R., Kathiresan, S., Keaney, J. F., Jr., Keyes, M. J., Lin, J. P., Meigs, J. B., Robins, S. J., Rong, J., Schnabel, R., Vita, J. A., Wang, T. J., Wilson, P. W., Wolf, P. A., & Vasan, R. S. (2007) Genome-wide association with select biomarker traits in the Framingham Heart Study. *BMC.Med Genet*. 8 Suppl 1: S11.
- Bennett, L., Palucka, A. K., Arce, E., Cantrell, V., Borvak, J., Banchereau, J., & Pascual, V. (2003) Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J.Exp.Med* 197: 711-723.
- Berger, P., McConnell, J. P., Nunn, M., Kornman, K. S., Sorrell, J., Stephenson, K., & Duff, G. W. (2002) C-reactive protein levels are influenced by common IL-1 gene variations. *Cytokine* 17: 171-174.
- Beutler, B. A. (2008) TLRs and innate immunity. *Blood*.

- Bhakdi, S., Torzewski, M., Klouche, M., & Hemmes, M. (1999) Complement and atherogenesis: binding of CRP to degraded, nonoxidized LDL enhances complement activation. *Arterioscler.Thromb.Vasc.Biol.* 19: 2348-2354.
- Bhakdi, S., Torzewski, M., Paprotka, K., Schmitt, S., Barsoom, H., Suriyaphol, P., Han, S. R., Lackner, K. J., & Husmann, M. (2004) Possible protective role for C-reactive protein in atherogenesis: complement activation by modified lipoproteins halts before detrimental terminal sequence. *Circulation* 109: 1870-1876.
- Bharadwaj, D., Stein, M. P., Volzer, M., Mold, C., & Du Clos, T. W. (1999) The major receptor for C-reactive protein on leukocytes is fcgamma receptor II. *J.Exp.Med* 190: 585-590.
- Bhatti, P., Church, D. M., Rutter, J. L., Struewing, J. P., & Sigurdson, A. J. (2006) Candidate single nucleotide polymorphism selection using publicly available tools: a guide for epidemiologists. *Am.J.Epidemiol.* 164: 794-804.
- Bienias, J. L., Beckett, L. A., Bennett, D. A., Wilson, R. S., & Evans, D. A. (2003) Design of the Chicago Health and Aging Project (CHAP). *J.Alzheimers.Dis.* 5: 349-355.
- Blaschke, F., Bruemmer, D., Yin, F., Takata, Y., Wang, W., Fishbein, M. C., Okura, T., Higaki, J., Graf, K., Fleck, E., Hsueh, W. A., & Law, R. E. (2004) C-reactive protein induces apoptosis in human coronary vascular smooth muscle cells. *Circulation* 110: 579-587.
- Block, S. R., Winfield, J. B., Lockshin, M. D., D'Angelo, W. A., & Christian, C. L. (1975) Studies of twins with systemic lupus erythematosus. A review of the literature and presentation of 12 additional sets. *Am.J.Med* 59: 533-552.
- Botstein, D. & Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat.Genet.* 33 Suppl: 228-237.
- Botto, M., Fong, K. Y., So, A. K., Koch, C., & Walport, M. J. (1990) Molecular basis of polymorphisms of human complement component C3. *J.Exp.Med* 172: 1011-1017.
- Bouzek, J. & Mancal, P. (1989) Comparability of some serum protein determinations by radial immunodiffusion, laser nephelometric and turbidimetric assays employing Q-antisera SEVAC. *J.Hyg.Epidemiol.Microbiol.Immunol.* 33: 105-111.
- Bronnestam, R. (1973) Studies on the C3 polymorphism. Relationship between phenotype and quantitative immunochemical measurements. *Hum.Hered.* 23: 128-134.
- Brooimans, R. A., Stegmann, A. P., van Dorp, W. T., van der Ark, A. A., van der Woude, F. J., van ES, L. A., & Daha, M. R. (1991) Interleukin 2 mediates stimulation of complement C3 biosynthesis in human proximal tubular epithelial cells. *J.Clin.Invest* 88: 379-384.
- Brown, E. J., Bohnsack, J. F., & Gresham, H. D. (1988) Mechanism of inhibition of immunoglobulin G-mediated phagocytosis by monoclonal antibodies that recognize the Mac-1 antigen. *J.Clin.Invest* 81: 365-375.
- Brown, K. M., Kondeatis, E., Vaughan, R. W., Kon, S. P., Farmer, C. K., Taylor, J. D., He, X., Johnston, A., Horsfield, C., Janssen, B. J., Gros, P., Zhou, W., Sacks, S. H., & Sheerin, N.

S. (2006) Influence of donor C3 allotype on late renal-transplantation outcome. *N.Engl.J.Med* 354: 2014-2023.

Browning, B. L. & Browning, S. R. (2007a) Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. *Genet.Epidemiol.* 31: 365-375.

Browning, B. L. & Browning, S. R. (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am.J.Hum.Genet.* 84: 210-223.

Browning, S. R. & Browning, B. L. (2007b) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am.J.Hum.Genet.* 81: 1084-1097.

Brull, D. J., Serrano, N., Zito, F., Jones, L., Montgomery, H. E., Rumley, A., Sharma, P., Lowe, G. D., World, M. J., Humphries, S. E., & Hingorani, A. D. (2003) Human CRP gene polymorphism influences CRP levels: implications for the prediction and pathogenesis of coronary heart disease. *Arterioscler.Thromb.Vasc.Biol.* 23: 2063-2069.

Buchta, R., Gennaro, R., Pontet, M., Fridkin, M., & Romeo, D. (1988) C-reactive protein decreases protein phosphorylation in stimulated human neutrophils. *FEBS Lett.* 237: 173-177.

Burton, P. R., Tiller, K. J., Gurrin, L. C., Cookson, W. O., Musk, A. W., & Palmer, L. J. (1999) Genetic variance components analysis for binary phenotypes using generalized linear mixed models (GLMMs) and Gibbs sampling. *Genet.Epidemiol.* 17: 118-140.

Burton, P. R., Tobin, M. D., & Hopper, J. L. (2005) Key concepts in genetic epidemiology. *Lancet* 366: 941-951.

Calabro, P., Willerson, J. T., & Yeh, E. T. (2003) Inflammatory cytokines stimulated C-reactive protein production by human coronary artery smooth muscle cells. *Circulation* 108: 1930-1932.

Cantor, R. M., Yuan, J., Napier, S., Kono, N., Grossman, J. M., Hahn, B. H., & Tsao, B. P. (2004) Systemic lupus erythematosus genome scan: support for linkage at 1q23, 2q33, 16q12-13, and 17q21-23 and novel evidence at 3p24, 10q23-24, 13q32, and 18q22-23. *Arthritis Rheum.* 50: 3203-3210.

Carlson, C. S., Aldred, S. F., Lee, P. K., Tracy, R. P., Schwartz, S. M., Rieder, M., Liu, K., Williams, O. D., Iribarren, C., Lewis, E. C., Fornage, M., Boerwinkle, E., Gross, M., Jaquish, C., Nickerson, D. A., Myers, R. M., Siscovick, D. S., & Reiner, A. P. (2005) Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *Am.J.Hum.Genet.* 77: 64-77.

Casas, J. P., Shah, T., Cooper, J., Hawe, E., McMahon, A. D., Gaffney, D., Packard, C. J., O'Reilly, D. S., Juhan-Vague, I., Yudkin, J. S., Tremoli, E., Margaglione, M., Di, M. G., Hamsten, A., Kooistra, T., Stephens, J. W., Hurel, S. J., Livingstone, S., Colhoun, H. M., Miller, G. J. *et al.* (2006) Insight into the nature of the CRP-coronary event association using Mendelian randomization. *Int.J.Epidemiol.* 35: 922-931.

- Casciola-Rosen, L. A., Anhalt, G., & Rosen, A. (1994) Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. *J.Exp.Med* 179: 1317-1330.
- Castelloe, J. & Templeton, A. R. (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Mol.Phylogenet.Evol.* 3: 102-113.
- Cermak, J., Key, N. S., Bach, R. R., Balla, J., Jacob, H. S., & Vercellotti, G. M. (1993) C-reactive protein induces human peripheral blood monocytes to synthesize tissue factor. *Blood* 82: 513-520.
- Chang, M. K., Binder, C. J., Torzewski, M., & Witztum, J. L. (2002) C-reactive protein binds to both oxidized LDL and apoptotic cells through recognition of a common ligand: Phosphorylcholine of oxidized phospholipids. *Proc.Natl.Acad.Sci.U.S.A* 99: 13043-13048.
- Chasman, D. I., Kozlowski, P., Zee, R. Y., Kwiatkowski, D. J., & Ridker, P. M. (2006) Qualitative and quantitative effects of APOE genetic variation on plasma C-reactive protein, LDL-cholesterol, and apoE protein. *Genes Immun.* 7: 211-219.
- Chasman, D. I., Pare, G., & Ridker, P. M. (2009) Population-based genomewide genetic analysis of common clinical chemistry analytes. *Clin.Chem.* 55: 39-51.
- Chen, J., Zhao, J., Huang, J., Su, S., Qiang, B., & Gu, D. (2005) -717A>G polymorphism of human C-reactive protein gene associated with coronary heart disease in ethnic Han Chinese: the Beijing atherosclerosis study. *J.Mol.Med* 83: 72-78.
- Chen, J. M., Ferec, C., & Cooper, D. N. (2006) A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. *Hum.Genet.* 120: 1-21.
- Chi, M., Tridandapani, S., Zhong, W., Coggeshall, K. M., & Mortensen, R. F. (2002) C-reactive protein induces signaling through Fc gamma RIIa on HL-60 granulocytes. *J.Immunol.* 168: 1413-1418.
- Choi, Y. S., Hur, J., & Jeong, S. (2007) Beta-catenin binds to the downstream region and regulates the expression C-reactive protein gene. *Nucleic Acids Res.* 35: 5511-5519.
- Christensen, S. R., Shupe, J., Nickerson, K., Kashgarian, M., Flavell, R. A., & Shlomchik, M. J. (2006) Toll-like receptor 7 and TLR9 dictate autoantibody specificity and have opposing inflammatory and regulatory roles in a murine model of lupus. *Immunity.* 25: 417-428.
- Cipollone, F., Toniato, E., Martinotti, S., Fazia, M., Iezzi, A., Cuccurullo, C., Pini, B., Ursi, S., Vitullo, G., Averna, M., Arca, M., Montali, A., Campagna, F., Uchino, S., Spigonardo, F., Taddei, S., Viridis, A., Ciabattini, G., Notarbartolo, A., Cuccurullo, F. *et al.* (2004) A polymorphism in the cyclooxygenase 2 gene as an inherited protective factor against myocardial infarction and stroke. *JAMA* 291: 2221-2228.
- Circolo, A., Garnier, G., Fukuda, W., Wang, X., Hidvegi, T., Szalai, A. J., Briles, D. E., Volanakis, J. E., Wetsel, R. A., & Colten, H. R. (1999) Genetic disruption of the murine complement C3 promoter region generates deficient mice with extrahepatic expression of C3 mRNA. *Immunopharmacology* 42: 135-149.

Clayton, D. (1999) A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am.J.Hum.Genet.* 65: 1170-1177.

Clayton, D. G., Walker, N. M., Smyth, D. J., Pask, R., Cooper, J. D., Maier, L. M., Smink, L. J., Lam, A. C., Ovington, N. R., Stevens, H. E., Nutland, S., Howson, J. M., Faham, M., Moorhead, M., Jones, H. B., Falkowski, M., Hardenbol, P., Willis, T. D., & Todd, J. A. (2005) Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat.Genet.* 37: 1243-1246.

Clement, M., Posada, D., & Crandall, K. A. (2000) TCS: a computer program to estimate gene genealogies. *Mol.Ecol.* 9: 1657-1659.

Cordell, H. J. & Clayton, D. G. (2005) Genetic association studies. *Lancet* 366: 1121-1131.

Crawford, D. C., Sanders, C. L., Qin, X., Smith, J. D., Shephard, C., Wong, M., Wittrak, L., Rieder, M. J., & Nickerson, D. A. (2006) Genetic variation is associated with C-reactive protein levels in the Third National Health and Nutrition Examination Survey. *Circulation* 114: 2458-2465.

Cree, B. A., Khan, O., Bourdette, D., Goodin, D. S., Cohen, J. A., Marrie, R. A., Glidden, D., Weinstock-Guttman, B., Reich, D., Patterson, N., Haines, J. L., Pericak-Vance, M., DeLoa, C., Oksenberg, J. R., & Hauser, S. L. (2004) Clinical characteristics of African Americans vs Caucasian Americans with multiple sclerosis. *Neurology* 63: 2039-2045.

CRP CHD Genetics Collaboration (2008) Collaborative pooled analysis of data on C-reactive protein gene variants and coronary disease: judging causality by Mendelian randomisation. *Eur.J.Epidemiol.* 23: 531-540.

Cushman, M., Legault, C., Barrett-Connor, E., Stefanick, M. L., Kessler, C., Judd, H. L., Sakkinen, P. A., & Tracy, R. P. (1999) Effect of postmenopausal hormones on inflammation-sensitive proteins: the Postmenopausal Estrogen/Progestin Interventions (PEPI) Study. *Circulation* 100: 717-722.

D'Aiuto, F., Casas, J. P., Shah, T., Humphries, S. E., Hingorani, A. D., & Tonetti, M. S. (2005) C-reactive protein (+1444C>T) polymorphism influences CRP response following a moderate inflammatory stimulus. *Atherosclerosis* 179: 413-417.

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., & Lander, E. S. (2001) High-resolution haplotype structure in the human genome. *Nat.Genet.* 29: 229-232.

Danesh, J., Collins, R., Appleby, P., & Peto, R. (1998) Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. *JAMA* 279: 1477-1482.

Danesh, J., Collins, R., & Peto, R. (1997) Chronic infections and coronary heart disease: is there a link? *Lancet* 350: 430-436.

Danesh, J., Muir, J., Wong, Y. K., Ward, M., Gallimore, J. R., & Pepys, M. B. (1999) Risk factors for coronary heart disease and acute-phase proteins. A population-based study. *Eur.Heart J.* 20: 954-959.

Danesh, J., Wheeler, J. G., Hirschfield, G. M., Eda, S., Eiriksdottir, G., Rumley, A., Lowe, G. D., Pepys, M. B., & Gudnason, V. (2004) C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N.Engl.J.Med* 350: 1387-1397.

Danilova, N. (2006) The evolution of immune mechanisms. *J.Exp.Zoolog.B Mol.Dev.Evol.* 306: 496-520.

Darlington, G. J., Wilson, D. R., & Lachman, L. B. (1986) Monocyte-conditioned medium, interleukin-1, and tumor necrosis factor stimulate the acute phase response in human hepatoma cells in vitro. *J.Cell Biol.* 103: 787-793.

Davey Smith G., Lawlor, D. A., Harbord, R., Timpson, N., Rumley, A., Lowe, G. D., Day, I. N., & Ebrahim, S. (2005) Association of C-reactive protein with blood pressure and hypertension: life course confounding and mendelian randomization tests of causality. *Arterioscler.Thromb.Vasc.Biol.* 25: 1051-1056.

Davey, S. G. & Ebrahim, S. (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int.J.Epidemiol.* 32: 1-22.

Davies, K. A., Erlendsson, K., Beynon, H. L., Peters, A. M., Steinsson, K., Valdimarsson, H., & Walport, M. J. (1993) Splenic uptake of immune complexes in man is complement-dependent. *J.Immunol.* 151: 3866-3873.

Davis, N. C., West, C. D., & Ho, M. (1972) Effect of aging of serum on quantitation of complement component C3. *Clin.Chem.* 18: 1485-1487.

de Beer, F. C., Soutar, A. K., Baltz, M. L., Trayner, I. M., Feinstein, A., & Pepys, M. B. (1982) Low density lipoprotein and very low density lipoprotein are selectively bound by aggregated C-reactive protein. *J.Exp.Med* 156: 230-242.

de Bruijn, M. H. & Fey, G. H. (1985) Human complement component C3: cDNA coding sequence and derived primary structure. *Proc.Natl.Acad.Sci.U.S.A* 82: 708-712.

De Jager, P. L., Richardson, A., Vyse, T. J., & Rioux, J. D. (2006) Genetic variation in toll-like receptor 9 and susceptibility to systemic lupus erythematosus. *Arthritis Rheum.* 54: 1279-1282.

de Lange M., Snieder, H., Ariens, R. A., Spector, T. D., & Grant, P. J. (2001) The genetics of haemostasis: a twin study. *Lancet* 357: 101-105.

Deapen, D., Escalante, A., Weinrib, L., Horwitz, D., Bachman, B., Roy-Burman, P., Walker, A., & Mack, T. M. (1992) A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum.* 35: 311-318.

Dishaw, L. J., Smith, S. L., & Bigger, C. H. (2005) Characterization of a C3-like cDNA in a coral: phylogenetic implications. *Immunogenetics* 57: 535-548.

Dong, Q. & Wright, J. R. (1996) Expression of C-reactive protein by alveolar macrophages. *J.Immunol.* 156: 4815-4820.

Du Clos, T. W. (1989) C-reactive protein reacts with the U1 small nuclear ribonucleoprotein. *J.Immunol.* 143: 2553-2559.

Du Clos, T. W., Zlock, L. T., & Rubin, R. L. (1988) Analysis of the binding of C-reactive protein to histones and chromatin. *J.Immunol.* 141: 4266-4270.

Edberg, J. C., Langefeld, C. D., Wu, J., Moser, K. L., Kaufman, K. M., Kelly, J., Bansal, V., Brown, W. M., Salmon, J. E., Rich, S. S., Harley, J. B., & Kimberly, R. P. (2002) Genetic linkage and association of Fcγ receptor IIIA (CD16A) on chromosome 1q23 with human systemic lupus erythematosus. *Arthritis Rheum.* 46: 2132-2140.

Edberg, J. C., Wu, J., Langefeld, C. D., Brown, E. E., Marion, M. C., McGwin, G., Jr., Petri, M., Ramsey-Goldman, R., Reveille, J. D., Frank, S. G., Kaufman, K. M., Harley, J. B., Alarcon, G. S., & Kimberly, R. P. (2008a) Genetic variation in the CRP promoter: association with systemic lupus erythematosus. *Hum.Mol.Genet.* 17: 1147-1155.

Edberg, J. C., Wu, J., Langefeld, C. D., Brown, E. E., Marion, M. C., McGwin, G., Petri, M., Ramsey-Goldman, R., Reveille, J. D., Frank, S. G., Kaufman, K. M., Harley, J. B., Alarcon, G. S., & Kimberly, R. P. (2008b) Genetic Variation in the CRP Promoter: Association with Systemic Lupus Erythematosus (SLE). *Hum.Mol.Genet.*

Eklund, C., Huttunen, R., Syrjanen, J., Laine, J., Vuento, R., & Hurme, M. (2006a) Polymorphism of the C-reactive protein gene is associated with mortality in bacteraemia. *Scand.J.Infect.Dis.* 38: 1069-1073.

Eklund, C., Jahan, F., Pessi, T., Lehtimäki, T., & Hurme, M. (2003) Interleukin 1B gene polymorphism is associated with baseline C-reactive protein levels in healthy individuals. *Eur.Cytokine Netw.* 14: 168-171.

Eklund, C., Lehtimäki, T., & Hurme, M. (2005) Epistatic effect of C-reactive protein (CRP) single nucleotide polymorphism (SNP) +1059 and interleukin-1B SNP +3954 on CRP concentration in healthy male blood donors. *Int.J.Immunogenet.* 32: 229-232.

Eklund, C., Nenonen, A., Kukkonen-Harjula, K., Borg, P., Fogelholm, M., Laine, S., Huhtala, H., Lehtimäki, T., & Hurme, M. (2006b) Association of the IL6-174(G/C) polymorphism with C-reactive protein concentration after weight loss in obese men. *Eur.Cytokine Netw.* 17: 131-135.

Elliott, P., Chambers, J. C., Zhang, W., Clarke, R., Hopewell, J. C., Peden, J. F., Erdmann, J., Braund, P., Engert, J. C., Bennett, D., Coin, L., Ashby, D., Tzoulaki, I., Brown, I. J., Mt-Isa, S., McCarthy, M. I., Peltonen, L., Freimer, N. B., Farrall, M., Ruukonen, A. *et al.* (2009) Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA* 302: 37-48.

Elston, R., Olson, J., & Palmer, L. 2002, "Multiple Comparisons," in *Biostatistical Genetics and Genetic Epidemiology*, Wiley, pp. 552-559.

Endo, Y., Takahashi, M., & Fujita, T. (2006) Lectin complement system and pattern recognition. *Immunobiology* 211: 283-293.

Engstrom, G., Hedblad, B., Berglund, G., Janzon, L., & Lindgarde, F. (2007) Plasma levels of complement C3 is associated with development of hypertension: a longitudinal cohort study. *J.Hum.Hypertens.* 21: 276-282.

Engstrom, G., Hedblad, B., Eriksson, K. F., Janzon, L., & Lindgarde, F. (2005) Complement C3 is a risk factor for the development of diabetes: a population-based cohort study. *Diabetes* 54: 570-575.

Enquobahrie, D. A., Rice, K., Williams, O. D., Williams, M. A., Gross, M. D., Lewis, C. E., Schwartz, S. M., & Siscovick, D. S. (2009) IL1B genetic variation and plasma C-reactive protein level among young adults: the CARDIA study. *Atherosclerosis* 202: 513-520.

Ewens, W. J. & Spielman, R. S. (1995) The transmission/disequilibrium test: history, subdivision, and admixture. *Am.J.Hum.Genet.* 57: 455-464.

Falus, A., Rokita, H., Walcz, E., Brozik, M., Hidvegi, T., & Meretey, K. (1990) Hormonal regulation of complement biosynthesis in human cell lines--II. Upregulation of the biosynthesis of complement components C3, factor B and C1 inhibitor by interleukin-6 and interleukin-1 in human hepatoma cell line. *Mol.Immunol.* 27: 197-201.

Fan, J. D., Wagner, B. L., & McDonnell, D. P. (1996) Identification of the sequences within the human complement 3 promoter required for estrogen responsiveness provides insight into the mechanism of tamoxifen mixed agonist activity. *Mol.Endocrinol.* 10: 1605-1616.

Fearon, D. T. & Carroll, M. C. (2000) Regulation of B lymphocyte responses to foreign and self-antigens by the CD19/CD21 complex. *Annu.Rev.Immunol.* 18: 393-422.

Fernando, M. M. A., Stevens, C. R., Walsh, E. C., De Jager, P. L., Goiyette P., Plenge, R. M., Vyse, T. J., & Rioux, J. D. (2008) Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS.Genet.* In Press.

Ferrari, S. L., hn-Luong, L., Garnero, P., Humphries, S. E., & Greenspan, S. L. (2003) Two promoter polymorphisms regulating interleukin-6 gene expression are associated with circulating levels of C-reactive protein and markers of bone resorption in postmenopausal women. *J.Clin.Endocrinol.Metab* 88: 255-259.

Fey, G., Domdey, H., Wiebauer, K., Whitehead, A. S., & Odink, K. (1983) Structure and expression of the C3 gene. *Springer Semin.Immunopathol.* 6: 119-147.

Fife, M. S., Ogilvie, E. M., Kelberman, D., Samuel, J., Gutierrez, A., Humphries, S. E., & Woo, P. (2005) Novel IL-6 haplotypes and disease association. *Genes Immun.* 6: 367-370.

Figueredo, M. A., Rodriguez, A., Ruiz-Yague, M., Romero, M., Fernandez-Cruz, A., Gomez-de la, C. E., & Patino, R. (2006) Autoantibodies against C-reactive protein: clinical associations in systemic lupus erythematosus and primary antiphospholipid syndrome. *J.Rheumatol.* 33: 1980-1986.

Finn, J. E. & Mathieson, P. W. (1993) Molecular analysis of C3 allotypes in patients with nephritic factor. *Clin.Exp.Immunol.* 91: 410-414.

Finn, J. E., Zhang, L., Agrawal, S., Jayne, D. R., Oliveira, D. B., & Mathieson, P. W. (1994) Molecular analysis of C3 allotypes in patients with systemic vasculitis. *Nephrol.Dial.Transplant.* 9: 1564-1567.

Fong, K. Y., Botto, M., Walport, M. J., & So, A. K. (1990) Genomic organization of human complement component C3. *Genomics* 7: 579-586.

Ford, E. S. (1999) Body mass index, diabetes, and C-reactive protein among U.S. adults. *Diabetes Care* 22: 1971-1977.

Fox, E. R., Benjamin, E. J., Sarpong, D. F., Rotimi, C. N., Wilson, J. G., Steffes, M. W., Chen, G., Adeyemo, A., Taylor, J. K., Samdarshi, T. E., & Taylor, H. A., Jr. (2008) Epidemiology, heritability, and genetic linkage of C-reactive protein in African Americans (from the Jackson Heart Study). *Am.J.Cardiol.* 102: 835-841.

Frank, M. M. & Fries, L. F. (1991) The role of complement in inflammation and phagocytosis. *Immunol.Today* 12: 322-326.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C.*et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851-861.

Fredslund, F., Jenner, L., Husted, L. B., Nyborg, J., Andersen, G. R., & Sottrup-Jensen, L. (2006) The structure of bovine complement component 3 reveals the basis for thioester function. *J.Mol.Biol.* 361: 115-127.

Friedlander, Y., Elkana, Y., Sinnreich, R., & Kark, J. D. (1995) Genetic and environmental sources of fibrinogen variability in Israeli families: the Kibbutzim Family Study. *Am.J.Hum.Genet.* 56: 1194-1206.

Gabay, C. & Kushner, I. (1999) Acute-phase proteins and other systemic responses to inflammation. *N.Engl.J.Med* 340: 448-454.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., & Altshuler, D. (2002) The structure of haplotype blocks in the human genome. *Science* 296: 2225-2229.

Ganter, U., Arcone, R., Toniatti, C., Morrone, G., & Ciliberto, G. (1989) Dual control of C-reactive protein gene expression by interleukin-1 and interleukin-6. *EMBO J.* 8: 3773-3779.

Garred, P., Voss, A., Madsen, H. O., & Junker, P. (2001) Association of mannose-binding lectin gene variation with disease severity and infections in a population-based cohort of systemic lupus erythematosus patients. *Genes Immun.* 2: 442-450.

Gershov, D., Kim, S., Brot, N., & Elkon, K. B. (2000) C-Reactive protein binds to apoptotic cells, protects the cells from assembly of the terminal complement components, and sustains an antiinflammatory innate immune response: implications for systemic autoimmunity. *J.Exp.Med* 192: 1353-1364.

Giovannoni, G., Miller, D. H., Losseff, N. A., Sailer, M., Lewellyn-Smith, N., Thompson, A. J., & Thompson, E. J. (2001) Serum inflammatory markers and clinical/MRI markers of disease progression in multiple sclerosis. *J.Neurol.* 248: 487-495.

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002) Finding genes that underlie complex traits. *Science* 298: 2345-2349.

Gordon, D. & Finch, S. J. (2005) Factors affecting statistical power in the detection of genetic association. *J.Clin.Invest* 115: 1408-1418.

Graham, R. R., Cotsapas, C., Davies, L., Hackett, R., Lessard, C. J., Leon, J. M., Burt, N. P., Guiducci, C., Parkin, M., Gates, C., Plenge, R. M., Behrens, T. W., Wither, J. E., Rioux, J. D., Fortin, P. R., Graham, D. C., Wong, A. K., Vyse, T. J., Daly, M. J., Altshuler, D. *et al.* (2008) Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat.Genet.* 40: 1059-1061.

Gros, P., Milder, F. J., & Janssen, B. J. (2008) Complement driven by conformational changes. *Nat.Rev.Immunol.* 8: 48-58.

Guiguet, M., Padieu, P., Mack, G., & Dalebroux, M. (1983) Laser nephelometric measurement of seven serum proteins compared with radial immunodiffusion. *J.Clin.Chem.Clin.Biochem.* 21: 217-221.

Haas, P. J. & van, S. J. (2007) Anaphylatoxins: their role in bacterial infection and inflammation. *Immunol.Res.* 37: 161-175.

Hage, F. G. & Szalai, A. J. (2007) C-reactive protein gene polymorphisms, C-reactive protein blood levels, and cardiovascular disease risk. *J.Am.Coll.Cardiol.* 50: 1115-1122.

Halder, I., Shriver, M., Thomas, M., Fernandez, J. R., & Frudakis, T. (2008) A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum.Mutat.* 29: 648-658.

Hamsten, A., Iselius, L., de, F. U., & Blomback, M. (1987) Genetic and cultural inheritance of plasma fibrinogen concentration. *Lancet* 2: 988-991.

Hansen, S. & Holmskov, U. (1998) Structural aspects of collectins and receptors for collectins. *Immunobiology* 199: 165-189.

Harboe, M. & Mollnes, T. E. (2008) The alternative complement pathway revisited. *J.Cell Mol.Med.*

Harley, J. B., Alarcon-Riquelme, M. E., Criswell, L. A., Jacob, C. O., Kimberly, R. P., Moser, K. L., Tsao, B. P., Vyse, T. J., & Langefeld, C. D. (2008) Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat.Genet.* 40: 204-210.

Harrap, S. B. & Hopper, J. L. (2001) Genetics of haemostasis. *Lancet* 357: 83-84.

Harrison, R. A. & Lachmann, P. J. (1980) The physiological breakdown of the third component of human complement. *Mol.Immunol.* 17: 9-20.

Hartung, K., Fontana, A., Klar, M., Krippner, H., Jorgens, K., Lang, B., Peter, H. H., Pichler, W. J., Schendel, D., Robin-Winn, M., & . (1989) Association of class I, II, and III MHC gene products with systemic lupus erythematosus. Results of a Central European multicenter study. *Rheumatol.Int.* 9: 13-18.

Hasegawa, K., Tamari, M., Shao, C., Shimizu, M., Takahashi, N., Mao, X. Q., Yamasaki, A., Kamada, F., Doi, S., Fujiwara, H., Miyatake, A., Fujita, K., Tamura, G., Matsubara, Y.,

Shirakawa, T., & Suzuki, Y. (2004) Variations in the C3, C3a receptor, and C5 genes affect susceptibility to bronchial asthma. *Hum.Genet.* 115: 295-301.

Haverkate, F., Thompson, S. G., Pyke, S. D., Gallimore, J. R., & Pepys, M. B. (1997) Production of C-reactive protein and risk of coronary events in stable and unstable angina. European Concerted Action on Thrombosis and Disabilities Angina Pectoris Study Group. *Lancet* 349: 462-466.

Hegele, R. A., Kraw, M. E., Ban, M. R., Miskie, B. A., Huff, M. W., & Cao, H. (2003) Elevated serum C-reactive protein and free fatty acids among nondiabetic carriers of missense mutations in the gene encoding lamin A/C (LMNA) with partial lipodystrophy. *Arterioscler.Thromb.Vasc.Biol.* 23: 111-116.

Helmy, K. Y., Katschke, K. J., Jr., Gorgani, N. N., Kljavin, N. M., Elliott, J. M., Diehl, L., Scales, S. J., Ghilardi, N., & van Lookeren, C. M. (2006) CR1g: a macrophage complement receptor required for phagocytosis of circulating pathogens. *Cell* 124: 915-927.

Herrmann, M., Voll, R. E., Zoller, O. M., Hagenhofer, M., Ponner, B. B., & Kalden, J. R. (1998) Impaired phagocytosis of apoptotic cell material by monocyte-derived macrophages from patients with systemic lupus erythematosus. *Arthritis Rheum.* 41: 1241-1250.

Heuertz, R. M., Xia, D., Samols, D., & Webster, R. O. (1994) Inhibition of C5a des Arg-induced neutrophil alveolitis in transgenic mice expressing C-reactive protein. *Am.J.Physiol* 266: L649-L654.

Hingorani, A. & Humphries, S. (2005) Nature's randomised trials. *Lancet* 366: 1906-1908.

Hirschfield, G. M., Gallimore, J. R., Kahan, M. C., Hutchinson, W. L., Sabin, C. A., Benson, G. M., Dhillon, A. P., Tennent, G. A., & Pepys, M. B. (2005) Transgenic human C-reactive protein is not proatherogenic in apolipoprotein E-deficient mice. *Proc.Natl.Acad.Sci.U.S.A* 102: 8309-8314.

Hirschfield, G. M., Herbert, J., Kahan, M. C., & Pepys, M. B. (2003) Human C-reactive protein does not protect against acute lipopolysaccharide challenge in mice. *J.Immunol.* 171: 6046-6051.

Hirschhorn, J. N. (2003) Genetic epidemiology of type 1 diabetes. *Pediatr.Diabetes* 4: 87-100.

Hirschhorn, J. N., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002) A comprehensive review of genetic association studies. *Genet.Med* 4: 45-61.

Hochberg, M. C. (1987) The application of genetic epidemiology to systemic lupus erythematosus. *J.Rheumatol.* 14: 867-869.

Hom, G., Graham, R. R., Modrek, B., Taylor, K. E., Ortmann, W., Garnier, S., Lee, A. T., Chung, S. A., Ferreira, R. C., Pant, P. V., Ballinger, D. G., Kosoy, R., Demirci, F. Y., Kamboh, M. I., Kao, A. H., Tian, C., Gunnarsson, I., Bengtsson, A. A., Rantapaa-Dahlqvist, S., Petri, M.*et al.* (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N.Engl.J.Med* 358: 900-909.

- Hopper, J. L. 2000, "Why "common environmental effects" are so uncommon in the literature," in *Advances in twin and sib-pair analysis*, T. D. Spector, H. Snieder, & A. J. MacGregor, eds., Greenwich Medical Media Ltd, London, pp. 151-165.
- Hopper, J. L. & Mathews, J. D. (1982) Extensions to multivariate normal models for pedigree analysis. *Ann.Hum.Genet.* 46: 373-383.
- Huang, Y. F., Wang, W., Han, J. Y., Wu, X. W., Zhang, S. T., Liu, C. J., Hu, Q. G., Xiong, P., Hamvas, R. M., Wood, N., Gong, F. L., & Bittles, A. H. (2003) Increased frequency of the mannose-binding lectin LX haplotype in Chinese systemic lupus erythematosus patients. *Eur.J.Immunogenet.* 30: 121-124.
- Hubacek, J. A., Stavek, P., Pit'ha, J., Skodova, Z., & Poledne, R. (2003) [CD14 (C-159-->T) polymorphism and levels of C-reactive protein]. *Cas.Lek.Cesk.* 142: 174-176.
- Hunnangkul, S., Nitsch, D., Rhodes, B., Chadha, S., Robertson, C. A., Pessoa-Lopes, P., Norsworthy, P. J., Fernando, M. M., Charles, P., kworth-Young, C., Isenberg, D. A., Whittaker, J. C., & Vyse, T. J. (2008) Familial clustering of non-nuclear autoantibodies and C3 and C4 complement components in systemic lupus erythematosus. *Arthritis Rheum.* 58: 1116-1124.
- Hur, J. W., Shin, H. D., Park, B. L., Kim, L. H., Kim, S. Y., & Bae, S. C. (2005) Association study of Toll-like receptor 9 gene polymorphism in Korean patients with systemic lupus erythematosus. *Tissue Antigens* 65: 266-270.
- Inoue, E., Yamanaka, H., Hara, M., Tomatsu, T., & Kamatani, N. (2007) Comparison of Disease Activity Score (DAS)28- erythrocyte sedimentation rate and D. *Ann.Rheum.Dis.* 66: 407-409.
- Inoue, H., Mashimo, Y., Funamizu, M., Shimojo, N., Hasegawa, K., Hirota, T., Doi, S., Kameda, M., Miyatake, A., Kohno, Y., Okamoto, Y., Tamari, M., Hata, A., & Suzuki, Y. (2008) Association study of the C3 gene with adult and childhood asthma. *J.Hum.Genet.* 53: 728-738.
- Ioannidis, J. P. & Kavvoura, F. K. (2006) Concordance of functional in vitro data and epidemiological associations in complex disease genetics. *Genet.Med* 8: 583-593.
- Jabs, W. J., Logering, B. A., Gerke, P., Kreft, B., Wolber, E. M., Klinger, M. H., Fricke, L., & Steinhoff, J. (2003) The kidney as a second site of human C-reactive protein formation in vivo. *Eur.J.Immunol.* 33: 152-161.
- Jang, Y., Kim, O. Y., Hyun, Y. J., Chae, J. S., Koh, S. J., Heo, Y. M., Choi, D., Shin, D. J., Huttner, K., & Lee, J. H. (2008) Interleukin-6-572C>G polymorphism-association with inflammatory variables in Korean men with coronary artery disease. *Transl.Res.* 151: 154-161.
- Janssen, B. J., Huizinga, E. G., Raaijmakers, H. C., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B., & Gros, P. (2005) Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* 437: 505-511.
- Jarva, H., Jokiranta, T. S., Hellwage, J., Zipfel, P. F., & Meri, S. (1999) Regulation of complement activation by C-reactive protein: targeting the complement inhibitory activity of

factor H by an interaction with short consensus repeat domains 7 and 8-11. *J.Immunol.* 163: 3957-3962.

Jeffreys, H. 1961, *The Theory of Probability*, 3rd edn, Oxford.

Jewell, W. S., Marnell, L. L., Rokeach, L. A., & Du Clos, T. W. (1993) C-reactive protein (CRP) binding to the Sm-D protein of snRNPS. Identification of a short polypeptide binding region. *Mol.Immunol.* 30: 701-708.

Jiang, S., Xia, D., & Samols, D. (2006) Expression of rabbit C-reactive protein in transgenic mice inhibits development of antigen-induced arthritis. *Scand.J.Rheumatol.* 35: 351-355.

Johnson, C. A., Densen, P., Hurford, R. K., Jr., Colten, H. R., & Wetsel, R. A. (1992) Type I human complement C2 deficiency. A 28-base pair gene deletion causes skipping of exon 6 during RNA splicing. *J.Biol.Chem.* 267: 9347-9353.

Johnson, E. & Hetland, G. (1988) Mononuclear phagocytes have the potential to synthesize the complete functional complement system. *Scand.J.Immunol.* 27: 489-493.

Johnson, G. C., Esposito, L., Barratt, B. J., Smith, A. N., Heward, J., Di, G. G., Ueda, H., Cordell, H. J., Eaves, I. A., Dudbridge, F., Twells, R. C., Payne, F., Hughes, W., Nutland, S., Stevens, H., Carr, P., Tuomilehto-Wolf, E., Tuomilehto, J., Gough, S. C., Clayton, D. G. *et al.* (2001) Haplotype tagging for the identification of common disease genes. *Nat.Genet.* 29: 233-237.

Jokiranta, T. S., Jaakola, V. P., Lehtinen, M. J., Parepalo, M., Meri, S., & Goldman, A. (2006) Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. *EMBO J.* 25: 1784-1794.

Jonsen, A., Gunnarsson, I., Gullstrand, B., Svenungsson, E., Bengtsson, A. A., Nived, O., Lundberg, I. E., Truedsson, L., & Sturfelt, G. (2007) Association between SLE nephritis and polymorphic variants of the CRP and FcγRIIIa genes. *Rheumatology.(Oxford)* 46: 1417-1421.

Jonsson, H., Sturfelt, G., Martensson, U., Truedsson, L., & Sjöholm, A. G. (1995) Prospective analysis of C1 dissociation and complement activation in patients with systemic lupus erythematosus. *Clin.Exp.Rheumatol.* 13: 573-580.

Juan, T. S., Wilson, D. R., Wilde, M. D., & Darlington, G. J. (1993) Participation of the transcription factor C/EBP delta in the acute-phase regulation of the human gene for complement component C3. *Proc.Natl.Acad.Sci.U.S.A* 90: 2584-2588.

Judson, R., Brain, C., Dain, B., Windemuth, A., Ruano, G., & Reed, C. (2004) New and confirmatory evidence of an association between APOE genotype and baseline C-reactive protein in dyslipidemic individuals. *Atherosclerosis* 177: 345-351.

Kalman, B. & Lublin, F. D. (1999) The genetics of multiple sclerosis. A review. *Biomed.Pharmacother.* 53: 358-370.

Kardys, I., de Maat, M. P., Uitterlinden, A. G., Hofman, A., & Witteman, J. C. (2006) C-reactive protein gene haplotypes and risk of coronary heart disease: the Rotterdam Study. *Eur.Heart J.* 27: 1331-1337.

- Kathiresan, S., Larson, M. G., Vasan, R. S., Guo, C. Y., Gona, P., Keaney, J. F., Jr., Wilson, P. W., Newton-Cheh, C., Musone, S. L., Camargo, A. L., Drake, J. A., Levy, D., O'Donnell, C. J., Hirschhorn, J. N., & Benjamin, E. J. (2006) Contribution of clinical correlates and 13 C-reactive protein gene polymorphisms to interindividual variability in serum C-reactive protein level. *Circulation* 113: 1415-1423.
- Katz, Y. & Strunk, R. C. (1988) Synovial fibroblast-like cells synthesize seven proteins of the complement system. *Arthritis Rheum.* 31: 1365-1370.
- Keenan, R. T., Swearingen, C. J., & Yazici, Y. (2008) Erythrocyte sedimentation rate and C-reactive protein levels are poorly correlated with clinical measures of disease activity in rheumatoid arthritis, systemic lupus erythematosus and osteoarthritis patients. *Clin.Exp.Rheumatol.* 26: 814-819.
- Kelley, J., Johnson, M. R., Alarcon, G. S., Kimberly, R. P., & Edberg, J. C. (2007) Variation in the relative copy number of the TLR7 gene in patients with systemic lupus erythematosus and healthy control subjects. *Arthritis Rheum.* 56: 3375-3378.
- Kew, R. R., Hyers, T. M., & Webster, R. O. (1990) Human C-reactive protein inhibits neutrophil chemotaxis in vitro: possible implications for the adult respiratory distress syndrome. *J.Lab Clin.Med* 115: 339-345.
- Khandavilli, S. D., Ceallaigh, P. O., Lloyd, C. J., & Whitaker, R. (2009) Serum C-reactive protein as a prognostic indicator in patients with oral squamous cell carcinoma. *Oral Oncol.*
- Khoury, M. J., Beaty, T. H., & Cohen B.H. 1993, *Fundamentals of Genetic Epidemiology*, 1st edn, Oxford University Press.
- Kim, D. K., Oh, S. Y., Kwon, H. C., Lee, S., Kwon, K. A., Kim, B. G., Kim, S. G., Kim, S. H., Jang, J. S., Kim, M. C., Kim, K. H., Han, J. Y., & Kim, H. J. (2009) Clinical significances of preoperative serum interleukin-6 and C-reactive protein level in operable gastric cancer. *BMC.Cancer* 9: 155.
- Kivimaki, M., Lawlor, D. A., Smith, G. D., Eklund, C., Hurme, M., Lehtimaki, T., Viikari, J. S., & Raitakari, O. T. (2007) Variants in the CRP gene as a measure of lifelong differences in average C-reactive protein levels: the Cardiovascular Risk in Young Finns Study, 1980-2001. *Am.J.Epidemiol.* 166: 760-764.
- Knight, J. C. (2005) Regulatory polymorphisms underlying complex disease traits. *J.Mol.Med* 83: 97-109.
- Kohl, J. (2001) Anaphylatoxins and infectious and non-infectious inflammatory diseases. *Mol.Immunol.* 38: 175-187.
- Koike, Y., Miki, C., Okugawa, Y., Yokoe, T., Toiyama, Y., Tanaka, K., Inoue, Y., & Kusunoki, M. (2008) Preoperative C-reactive protein as a prognostic and therapeutic marker for colorectal cancer. *J.Surg.Oncol.* 98: 540-544.
- Kompoti, M., Drimis, S., Papadaki, A., Kotsomytis, K., Pouloupoulou, C., Gianneli, D., & Voutsinas, D. (2008) Serum C-reactive protein at admission predicts in-hospital mortality in medical patients. *Eur.J.Intern.Med* 19: 261-265.

- Kondrashov, A. S. (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum.Mutat.* 21: 12-27.
- Korb, L. C. & Ahearn, J. M. (1997) C1q binds directly and specifically to surface blebs of apoptotic human keratinocytes: complement deficiency and systemic lupus erythematosus revisited. *J.Immunol.* 158: 4525-4528.
- Kosoy, R., Nassir, R., Tian, C., White, P. A., Butler, L. M., Silva, G., Kittles, R., arcon-Riquelme, M. E., Gregersen, P. K., Belmont, J. W., De, L., V, & Seldin, M. F. (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum.Mutat.* 30: 69-78.
- Kovacs, A., Green, F., Hansson, L. O., Lundman, P., Samnegard, A., Boquist, S., Ericsson, C. G., Watkins, H., Hamsten, A., & Tornvall, P. (2005) A novel common single nucleotide polymorphism in the promoter region of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. *Atherosclerosis* 178: 193-198.
- Kozyrev, S. V. & Alarcon-Riquelme, M. E. (2007) The genetics and biology of Irf5-mediated signaling in lupus. *Autoimmunity* 40: 591-601.
- Krauss, J. C., PooH, Xue, W., Mayo-Bond, L., Todd, R. F., III, & Petty, H. R. (1994) Reconstitution of antibody-dependent phagocytosis in fibroblasts expressing Fc gamma receptor IIIB and the complement receptor type 3. *J.Immunol.* 153: 1769-1777.
- Krych-Goldberg, M. & Atkinson, J. P. (2001) Structure-function relationships of complement receptor type 1. *Immunol.Rev.* 180: 112-122.
- Krystosek, A. (1999) Preferential sites of early DNA cleavage in apoptosis and the pathway of nuclear damage. *Histochem.Cell Biol.* 111: 265-276.
- Kuller, L. H., Tracy, R. P., Shaten, J., & Meilahn, E. N. (1996) Relation of C-reactive protein and coronary heart disease in the MRFIT nested case-control study. Multiple Risk Factor Intervention Trial. *Am.J.Epidemiol.* 144: 537-547.
- Kushner, I. & Feldmann, G. (1978) Control of the acute phase response. Demonstration of C-reactive protein synthesis and secretion by hepatocytes during acute inflammation in the rabbit. *J.Exp.Med* 148: 466-477.
- Lagrand, W. K., Visser, C. A., Hermens, W. T., Niessen, H. W., Verheugt, F. W., Wolbink, G. J., & Hack, C. E. (1999) C-reactive protein as a cardiovascular risk factor: more than an epiphenomenon? *Circulation* 100: 96-102.
- Lahita, R. G. 2004, *Systemic Lupus Erythematosus*, 4th edn, Academic Press.
- Lakka, H. M., Lakka, T. A., Rankinen, T., Rice, T., Rao, D. C., Leon, A. S., Skinner, J. S., & Bouchard, C. (2006a) The TNF-alpha G-308A polymorphism is associated with C-reactive protein levels: the HERITAGE Family Study. *Vascul.Pharmacol.* 44: 377-383.
- Lakka, T. A., Rankinen, T., Rice, T., Leon, A. S., Rao, D. C., Skinner, J. S., & Bouchard, C. (2006b) Quantitative trait locus on chromosome 20q13 for plasma levels of C-reactive protein in healthy whites: the HERITAGE Family Study. *Physiol Genomics* 27: 103-107.

- Lange, L. A., Burdon, K., Langefeld, C. D., Liu, Y., Beck, S. R., Rich, S. S., Freedman, B. I., Brosnihan, K. B., Herrington, D. M., Wagenknecht, L. E., & Bowden, D. W. (2006a) Heritability and expression of C-reactive protein in type 2 diabetes in the Diabetes Heart Study. *Ann.Hum.Genet.* 70: 717-725.
- Lange, L. A., Carlson, C. S., Hindorff, L. A., Lange, E. M., Walston, J., Durda, J. P., Cushman, M., Bis, J. C., Zeng, D., Lin, D., Kuller, L. H., Nickerson, D. A., Psaty, B. M., Tracy, R. P., & Reiner, A. P. (2006b) Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. *JAMA* 296: 2703-2711.
- Latkovskis, G., Licis, N., & Kalnins, U. (2004) C-reactive protein levels and common polymorphisms of the interleukin-1 gene cluster and interleukin-6 gene in patients with coronary heart disease. *Eur.J.Immunogenet.* 31: 207-213.
- Lau, C. M., Broughton, C., Tabor, A. S., Akira, S., Flavell, R. A., Mamula, M. J., Christensen, S. R., Shlomchik, M. J., Viglianti, G. A., Rifkin, I. R., & Marshak-Rothstein, A. (2005) RNA-associated autoantigens activate B cells by combined B cell antigen receptor/Toll-like receptor 7 engagement. *J.Exp.Med* 202: 1171-1177.
- Law, S. K. & Levine, R. P. (1977) Interaction between the third complement protein and cell surface macromolecules. *Proc.Natl.Acad.Sci.U.S.A* 74: 2701-2705.
- Law, S. K., Lichtenberg, N. A., & Levine, R. P. (1979) Evidence for an ester linkage between the labile binding site of C3b and receptive surfaces. *J.Immunol.* 123: 1388-1394.
- Lee, C. C., You, N. C., Song, Y., Hsu, Y. H., Manson, J., Nathan, L., Tinker, L., & Liu, S. (2009) Relation of genetic variation in the gene coding for C-reactive protein with its plasma protein concentrations: findings from the Women's Health Initiative Observational Cohort. *Clin.Chem.* 55: 351-360.
- Lee, M. S. & Kim, Y. J. (2007) Signaling pathways downstream of pattern-recognition receptors and their cross talk. *Annu.Rev.Biochem.* 76: 447-480.
- Levinson, S. S., Miller, J. J., & Elin, R. J. (2004) Poor predictive value of high-sensitivity C-reactive protein indicates need for reassessment. *Clin.Chem.* 50: 1733-1735.
- Li, J., Pircher, P. C., Schulman, I. G., & Westin, S. K. (2005) Regulation of complement C3 expression by the bile acid receptor FXR. *J.Biol.Chem.* 280: 7427-7434.
- Li, S. P. & Goldman, N. D. (1996a) Regulation of human C-reactive protein gene expression by two synergistic IL-6 responsive elements. *Biochemistry* 35: 9060-9068.
- Li, S. P. & Goldman, N. D. (1996b) Regulation of human C-reactive protein gene expression by two synergistic IL-6 responsive elements. *Biochemistry* 35: 9060-9068.
- Libra, M., Signorelli, S. S., Bevelacqua, Y., Navolanic, P. M., Bevelacqua, V., Polesel, J., Talamini, R., Stivala, F., Mazzarino, M. C., & Malaponte, G. (2006) Analysis of G(-174)C IL-6 polymorphism and plasma concentrations of inflammatory markers in patients with type 2 diabetes and peripheral arterial disease. *J.Clin.Pathol.* 59: 211-215.
- Liuzzo, G., Biasucci, L. M., Gallimore, J. R., Grillo, R. L., Rebuzzi, A. G., Pepys, M. B., & Maseri, A. (1994) The prognostic value of C-reactive protein and serum amyloid A protein in severe unstable angina. *N.Engl.J.Med* 331: 417-424.

- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000) WinBUGS - a Bayesian modelling framework: concepts, structure and extensibility. *Statistics and Computing* 10: 325-337.
- Lunn, D. J., Whittaker, J. C., & Best, N. (2006) A Bayesian toolkit for genetic association studies. *Genet.Epidemiol.* 30: 231-247.
- MacGregor, A. J., Gallimore, J. R., Spector, T. D., & Pepys, M. B. (2004) Genetic effects on baseline values of C-reactive protein and serum amyloid A protein: a comparison of monozygotic and dizygotic twins. *Clin.Chem.* 50: 130-134.
- Mackiewicz, A., Ganapathi, M. K., Schultz, D., Brabenec, A., Weinstein, J., Kelley, M. F., & Kushner, I. (1990) Transforming growth factor beta 1 regulates production of acute-phase proteins. *Proc.Natl.Acad.Sci.U.S.A* 87: 1491-1495.
- Maller, J. B., Fagerness, J. A., Reynolds, R. C., Neale, B. M., Daly, M. J., & Seddon, J. M. (2007) Variation in complement factor 3 is associated with risk of age-related macular degeneration. *Nat.Genet.* 39: 1200-1201.
- Mancini, G., Carbonara, A. O., & Heremans, J. F. (1965) Immunochemical quantitation of antigens by single radial immunodiffusion. *Immunochemistry.* 2: 235-254.
- Manderson, A. P., Botto, M., & Walport, M. J. (2004) The role of complement in the development of systemic lupus erythematosus. *Annu.Rev.Immunol.* 22: 431-456.
- Manttari, M., Manninen, V., Palosuo, T., & Ehnholm, C. (2001) Apolipoprotein E polymorphism and C-reactive protein in dyslipidemic middle-aged men. *Atherosclerosis* 156: 237-238.
- Maranto, J., Rappaport, J., & Datta, P. K. (2008) Regulation of complement component C3 in astrocytes by IL-1beta and morphine. *J.Neuroimmune.Pharmacol.* 3: 43-51.
- Marnell, L. L., Mold, C., Volzer, M. A., Burlingame, R. W., & Du Clos, T. W. (1995) C-reactive protein binds to Fc gamma RI in transfected COS cells. *J.Immunol.* 155: 2185-2193.
- Marsik, C., Sunder-Plassmann, R., Jilma, B., Kovar, F. M., Mannhalter, C., Wagner, O., Rumpold, H., & Endler, G. (2006) The C-reactive protein (+)1444C/T alteration modulates the inflammation and coagulation response in human endotoxemia. *Clin.Chem.* 52: 1952-1957.
- Martin, N., Boomsma, D., & Machin, G. (1997) A twin-pronged attack on complex traits. *Nat.Genet.* 17: 387-392.
- Marz, W., Scharnagl, H., Hoffmann, M. M., Boehm, B. O., & Winkelmann, B. R. (2004) The apolipoprotein E polymorphism is associated with circulating C-reactive protein (the Ludwigshafen risk and cardiovascular health study). *Eur.Heart J.* 25: 2109-2119.
- Mazza, A., Bendini, M. G., Cristofori, M., Nardi, S., Leggio, M., De, C. R., Giordano, A., Cozzari, L., Giordano, G., & Cappato, R. (2009) Baseline apnoea/hypopnoea index and high-sensitivity C-reactive protein for the risk of recurrence of atrial fibrillation after successful electrical cardioversion: a predictive model based upon the multiple effects of significant variables. *Europace.*

- McLean, R. H. & Winkelstein, J. A. (1984) Genetically determined variation in the complement system: relationship to disease. *J.Pediatr.* 105: 179-188.
- McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., Fulton, R., Kucaba, T. A., Wagner-McPherson, C., Barbazuk, W. B., Gregory, S. G., Humphray, S. J., French, L., Evans, R. S., Bethel, G., Whittaker, A. *et al.* (2001) A physical map of the human genome. *Nature* 409: 934-941.
- Melzer, D., Perry, J. R., Hernandez, D., Corsi, A. M., Stevens, K., Rafferty, I., Lauretani, F., Murray, A., Gibbs, J. R., Paolisso, G., Rafiq, S., Simon-Sanchez, J., Lango, H., Scholz, S., Weedon, M. N., Arepalli, S., Rice, N., Washecka, N., Hurst, A., Britton, A. *et al.* (2008) A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS.Genet.* 4: e1000072.
- Mevorach, D., Mascarenhas, J. O., Gershov, D., & Elkon, K. B. (1998a) Complement-dependent clearance of apoptotic cells by human macrophages. *J.Exp.Med* 188: 2313-2320.
- Mevorach, D., Zhou, J. L., Song, X., & Elkon, K. B. (1998b) Systemic exposure to irradiated apoptotic cells induces autoantibody production. *J.Exp.Med* 188: 387-392.
- Miller, D. T., Zee, R. Y., Suk, D. J., Kozlowski, P., Chasman, D. I., Lazarus, R., Cook, N. R., Ridker, P. M., & Kwiatkowski, D. J. (2005) Association of common CRP gene variants with CRP levels and cardiovascular events. *Ann.Hum.Genet.* 69: 623-638.
- Miller, G. W. & Nussenzweig, V. (1975) A new complement function: solubilization of antigen-antibody aggregates. *Proc.Natl.Acad.Sci.U.S.A* 72: 418-422.
- Miyagawa, H., Yamai, M., Sakaguchi, D., Kiyohara, C., Tsukamoto, H., Kimoto, Y., Nakamura, T., Lee, J. H., Tsai, C. Y., Chiang, B. L., Shimoda, T., Harada, M., Tahira, T., Hayashi, K., & Horiuchi, T. (2008) Association of polymorphisms in complement component C3 gene with susceptibility to systemic lupus erythematosus. *Rheumatology.(Oxford)* 47: 158-164.
- Mold, C. & Du Clos, T. W. (2006) C-reactive protein increases cytokine responses to *Streptococcus pneumoniae* through interactions with Fc gamma receptors. *J.Immunol.* 176: 7598-7604.
- Mold, C., Gewurz, H., & Du Clos, T. W. (1999) Regulation of complement activation by C-reactive protein. *Immunopharmacology* 42: 23-30.
- Mold, C., Gresham, H. D., & Du Clos, T. W. (2001) Serum amyloid P component and C-reactive protein mediate phagocytosis through murine Fc gamma Rs. *J.Immunol.* 166: 1200-1205.
- Mold, C., Kingzette, M., & Gewurz, H. (1984) C-reactive protein inhibits pneumococcal activation of the alternative pathway by increasing the interaction between factor H and C3b. *J.Immunol.* 133: 882-885.
- Mold, C., Nakayama, S., Holzer, T. J., Gewurz, H., & Du Clos, T. W. (1981) C-reactive protein is protective against *Streptococcus pneumoniae* infection in mice. *J.Exp.Med* 154: 1703-1708.

- Mold, C., Rodic-Polic, B., & Du Clos, T. W. (2002a) Protection from *Streptococcus pneumoniae* infection by C-reactive protein and natural antibody requires complement but not Fc gamma receptors. *J.Immunol.* 168: 6375-6381.
- Mold, C., Rodriguez, W., Rodic-Polic, B., & Du Clos, T. W. (2002b) C-reactive protein mediates protection from lipopolysaccharide through interactions with Fc gamma R. *J.Immunol.* 169: 7019-7025.
- Moon, R., Parikh, A. A., Szabo, C., Fischer, J. E., Salzman, A. L., & Hasselgren, P. O. (1997) Complement C3 production in human intestinal epithelial cells is regulated by interleukin 1beta and tumor necrosis factor alpha. *Arch.Surg.* 132: 1289-1293.
- Morelli, A. E., Larregina, A. T., Shufesky, W. J., Zahorchak, A. F., Logar, A. J., Papworth, G. D., Wang, Z., Watkins, S. C., Falo, L. D., Jr., & Thomson, A. W. (2003) Internalization of circulating apoptotic cells by splenic marginal zone dendritic cells: dependence on complement receptors and effect on cytokine production. *Blood* 101: 611-620.
- Morita, A., Nakayama, T., Doba, N., Hinohara, S., & Soma, M. (2006) Polymorphism of the C-reactive protein (CRP) gene is related to serum CRP Level and arterial pulse wave velocity in healthy elderly Japanese. *Hypertens.Res.* 29: 323-331.
- Moser, K. L., Neas, B. R., Salmon, J. E., Yu, H., Gray-McGuire, C., Asundi, N., Bruner, G. R., Fox, J., Kelly, J., Henshall, S., Bacino, D., Dietz, M., Hogue, R., Koelsch, G., Nightingale, L., Shaver, T., Abdou, N. I., Albert, D. A., Carson, C., Petri, M.*et al.* (1998) Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. *Proc.Natl.Acad.Sci.U.S.A* 95: 14869-14874.
- Muller-Eberhard, H. J. (1986) The membrane attack complex of complement. *Annu.Rev.Immunol.* 4: 503-528.
- Muller-Eberhard, H. J. (1988) Molecular organization and function of the complement system. *Annu.Rev.Biochem.* 57: 321-347.
- Murphy, C., Beckers, J., & Ruther, U. (1995) Regulation of the human C-reactive protein gene in transgenic mice. *J.Biol.Chem.* 270: 704-708.
- Murphy, K. M., Travers, M., & Walport, M. 2008, *Janeway's Immunobiology*, 7th edn, Garland Publishing Inc..
- Murphy, T. M., Baum, L. L., & Beaman, K. D. (1991) Extrahepatic transcription of human C-reactive protein. *J.Exp.Med* 173: 495-498.
- Myllymaki, J., Saha, H., Pasternack, A., Helin, H., & Mustonen, J. (2006) High serum C3 predicts poor outcome in IgM nephropathy. *Nephron Clin.Pract.* 102: c122-c127.
- Naama, J. K., Hamilton, A. O., Yeung-Laiwah, A. C., & Whaley, K. (1984) Prevention of immune precipitation by purified classical pathway complement components. *Clin.Exp.Immunol.* 58: 486-492.
- Nadeau, J. H. (2001) Modifier genes in mice and humans. *Nat.Rev.Genet.* 2: 165-174.
- Nath, S. K., Han, S., Kim-Howard, X., Kelly, J. A., Viswanathan, P., Gilkeson, G. S., Chen, W., Zhu, C., McEver, R. P., Kimberly, R. P., arcon-Riquelme, M. E., Vyse, T. J., Li, Q. Z.,

- Wakeland, E. K., Merrill, J. T., James, J. A., Kaufman, K. M., Guthridge, J. M., & Harley, J. B. (2008) A nonsynonymous functional variant in integrin- $\alpha$ (M) (encoded by ITGAM) is associated with systemic lupus erythematosus. *Nat.Genet.* 40: 152-154.
- Nei, M. & Roychoudhury, A. K. (1993) Evolutionary relationships of human populations on a global scale. *Mol.Biol.Evol.* 10: 927-943.
- Ng, M. W., Lau, C. S., Chan, T. M., Wong, W. H., & Lau, Y. L. (2005) Polymorphisms of the toll-like receptor 9 (TLR9) gene with systemic lupus erythematosus in Chinese. *Rheumatology.(Oxford)* 44: 1456-1457.
- Nishida, N., Walz, T., & Springer, T. A. (2006) Structural transitions of complement component C3 and its activation products. *Proc.Natl.Acad.Sci.U.S.A* 103: 19737-19742.
- Nonaka, M., Azumi, K., Ji, X., Namikawa-Yamada, C., Sasaki, M., Saiga, H., Dodds, A. W., Sekine, H., Homma, M. K., Matsushita, M., Endo, Y., & Fujita, T. (1999) Opsonic complement component C3 in the solitary ascidian, *Halocynthia roretzi*. *J.Immunol.* 162: 387-391.
- O'Connell, J. R. & Weeks, D. E. (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am.J.Hum.Genet.* 63: 259-266.
- Obisesan, T. O., Leeuwenburgh, C., Phillips, T., Ferrell, R. E., Phares, D. A., Prior, S. J., & Hagberg, J. M. (2004) C-reactive protein genotypes affect baseline, but not exercise training-induced changes, in C-reactive protein levels. *Arterioscler.Thromb.Vasc.Biol.* 24: 1874-1879.
- Oksenberg, J. R., Barcellos, L. F., Cree, B. A., Baranzini, S. E., Bugawan, T. L., Khan, O., Lincoln, R. R., Swerdlin, A., Mignot, E., Lin, L., Goodin, D., Erlich, H. A., Schmidt, S., Thomson, G., Reich, D. E., Pericak-Vance, M. A., Haines, J. L., & Hauser, S. L. (2004) Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am.J.Hum.Genet.* 74: 160-167.
- Paik, J. K., Kim, O. Y., Koh, S. J., Jang, Y., Chae, J. S., Kim, J. Y., Kim, H. J., Hyun, Y. J., Cho, J. R., & Lee, J. H. (2007) Additive effect of interleukin-6 and C-reactive protein (CRP) single nucleotide polymorphism on serum CRP concentration and other cardiovascular risk factors. *Clin.Chim.Acta* 380: 68-74.
- Pancer, Z. & Cooper, M. D. (2006) The evolution of adaptive immunity. *Annu.Rev.Immunol.* 24: 497-518.
- Pankow, J. S., Folsom, A. R., Cushman, M., Borecki, I. B., Hopkins, P. N., Eckfeldt, J. H., & Tracy, R. P. (2001) Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis* 154: 681-689.
- Papadimitraki, E. D., Bertsias, G. K., & Boumpas, D. T. (2007) Toll like receptors and autoimmunity: a critical appraisal. *J.Autoimmun.* 29: 310-318.
- Patel, D. N., King, C. A., Bailey, S. R., Holt, J. W., Venkatachalam, K., Agrawal, A., Valente, A. J., & Chandrasekar, B. (2007) Interleukin-17 stimulates C-reactive protein expression in hepatocytes and smooth muscle cells via p38 MAPK and ERK1/2-dependent NF-kappaB and C/EBPbeta activation. *J.Biol.Chem.* 282: 27229-27238.

Paul, A., Ko, K. W., Li, L., Yechoor, V., McCrory, M. A., Szalai, A. J., & Chan, L. (2004) C-reactive protein accelerates the progression of atherosclerosis in apolipoprotein E-deficient mice. *Circulation* 109: 647-655.

Paulus, H. E., Ramos, B., Wong, W. K., Ahmed, A., Bulpitt, K., Park, G., Sterz, M., & Clements, P. (1999) Equivalence of the acute phase reactants C-reactive protein, plasma viscosity, and Westergren erythrocyte sedimentation rate when used to calculate American College of Rheumatology 20% improvement criteria or the Disease Activity Score in patients with early rheumatoid arthritis. Western Consortium of Practicing Rheumatologists. *J.Rheumatol.* 26: 2324-2331.

Pearson, T. A., Mensah, G. A., Alexander, R. W., Anderson, J. L., Cannon, R. O., III, Criqui, M., Fadl, Y. Y., Fortmann, S. P., Hong, Y., Myers, G. L., Rifai, N., Smith, S. C., Jr., Taubert, K., Tracy, R. P., & Vinicor, F. (2003) Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* 107: 499-511.

Peiser, L., Mukhopadhyay, S., & Gordon, S. (2002) Scavenger receptors in innate immunity. *Curr.Opin.Immunol.* 14: 123-128.

Pepys, M. B. & Hirschfield, G. M. (2003) C-reactive protein: a critical update. *J.Clin.Invest* 111: 1805-1812.

Pepys, M. B., Hirschfield, G. M., Tennent, G. A., Gallimore, J. R., Kahan, M. C., Bellotti, V., Hawkins, P. N., Myers, R. M., Smith, M. D., Polara, A., Cobb, A. J., Ley, S. V., Aquilina, J. A., Robinson, C. V., Sharif, I., Gray, G. A., Sabin, C. A., Jenvey, M. C., Kolstoe, S. E., Thompson, D.*et al.* (2006) Targeting C-reactive protein for the treatment of cardiovascular disease. *Nature* 440: 1217-1221.

Pepys, M. B., Rowe, I. F., & Baltz, M. L. (1985) C-reactive protein: binding to lipids and lipoproteins. *Int.Rev.Exp.Pathol.* 27: 83-111.

Pereira Da Silva, J. A., Elkon, K. B., Hughes, G. R., Dyck, R. F., & Pepys, M. B. (1980) C-reactive protein levels in systemic lupus erythematosus: a classification criterion? *Arthritis Rheum.* 23: 770-771.

Perissutti, S. & Tedesco, F. (1994) Effect of cytokines on the secretion of the fifth and eighth complement components by HepG2 cells. *Int.J.Clin.Lab Res.* 24: 45-48.

Perry, T. E., Muehlschlegel, J. D., Liu, K. Y., Fox, A. A., Collard, C. D., Body, S. C., & Shernan, S. K. (2009) C-Reactive protein gene variants are associated with postoperative C-reactive protein levels after coronary artery bypass surgery. *BMC.Med Genet.* 10: 38.

Persson, U., Truedsson, L., Westman, K. W., & Segelmark, M. (1999) C3 and C4 allotypes in anti-neutrophil cytoplasmic autoantibody (ANCA)-positive vasculitis. *Clin.Exp.Immunol.* 116: 379-382.

Petrilli, V., Dostert, C., Muruve, D. A., & Tschopp, J. (2007) The inflammasome: a danger sensing complex triggering innate immunity. *Curr.Opin.Immunol.* 19: 615-622.

Phillips, D. I. (1993) Twin studies in medical research: can they tell us whether diseases are genetically determined? *Lancet* 341: 1008-1009.

- Pickering, M. C. & Walport, M. J. (2000) Links between complement abnormalities and systemic lupus erythematosus. *Rheumatology*.(Oxford) 39: 133-141.
- Pietila, K. O., Harmoinen, A. P., Jokiniitty, J., & Pasternack, A. I. (1996) Serum C-reactive protein concentration in acute myocardial infarction and its relationship to mortality during 24 months of follow-up in patients under thrombolytic treatment. *Eur.Heart J.* 17: 1345-1349.
- Pisitkun, P., Deane, J. A., Difilippantonio, M. J., Tarasenko, T., Satterthwaite, A. B., & Bolland, S. (2006) Autoreactive B cell responses to RNA-related antigens due to TLR7 gene duplication. *Science* 312: 1669-1672.
- Pittoni, V. & Valesini, G. (2002) The clearance of apoptotic cells: implications for autoimmunity. *Autoimmun.Rev.* 1: 154-161.
- Poole, C. D., Conway, P., Reynolds, A., & Currie, C. J. (2008) The association between C-reactive protein and the likelihood of progression to joint replacement in people with rheumatoid arthritis: a retrospective observational study. *BMC.Musculoskelet.Disord.* 9: 146.
- Potaczek, D. P., Undas, A., Nowakowski, T., & Szczeklik, A. (2007) Association between inflammatory markers and the interleukin-6 -174 G/C polymorphism is abolished in peripheral arterial occlusive disease. *Int.Angiol.* 26: 318-323.
- Pradhan, A. D., Manson, J. E., Rossouw, J. E., Siscovick, D. S., Mouton, C. P., Rifai, N., Wallace, R. B., Jackson, R. D., Pettinger, M. B., & Ridker, P. M. (2002) Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: prospective analysis from the Women's Health Initiative observational study. *JAMA* 288: 980-987.
- Prevoo, M. L., van 't Hof, M. A., Kuper, H. H., van Leeuwen, M. A., van de Putte, L. B., & van Riel, P. L. (1995) Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum.* 38: 44-48.
- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., Gonzalez, J. R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., Marshall, C. R.*et al.* (2006) Global variation in copy number in the human genome. *Nature* 444: 444-454.
- Redondo, M. J., Fain, P. R., & Eisenbarth, G. S. (2001) Genetics of type 1A diabetes. *Recent Prog.Horm.Res.* 56: 69-89.
- Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., & Lander, E. S. (2001) Linkage disequilibrium in the human genome. *Nature* 411: 199-204.
- Reiner, A. P., Barber, M. J., Guan, Y., Ridker, P. M., Lange, L. A., Chasman, D. I., Walston, J. D., Cooper, G. M., Jenny, N. S., Rieder, M. J., Durda, J. P., Smith, J. D., Novembre, J., Tracy, R. P., Rotter, J. I., Stephens, M., Nickerson, D. A., & Krauss, R. M. (2008) Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am.J.Hum.Genet.* 82: 1193-1201.

- Reis, S., Falcao, D. A., & Isaac, L. (2006) Clinical aspects and molecular basis of primary deficiencies of complement component C3 and its regulatory proteins factor I and factor H. *Scand.J.Immunol.* 63: 155-168.
- Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316: 222-234.
- Rhodes, B., Hunnangkul, S., Morris, D. L., Hsaio, L. C., Cunningham-Graham, D. S., Nitsch, D., Whittaker, J. C., & Vyse, T. J. (2009) The heritability and genetics of complement C3 expression in UK SLE families. *Genes Immun.* 10: 525-530.
- Rhodes, B., Meek, J., Whittaker, J. C., & Vyse, T. J. (2008a) Quantification of the genetic component of basal C-reactive protein expression in SLE nuclear families. *Ann.Hum.Genet.* 72: 611-620.
- Rhodes, B., Morris, D. L., Subrahmanyam, L., Aubin, C., de Leon, C. F., Kelly, J. F., Evans, D. A., Whittaker, J. C., Oksenberg, J. R., De Jager, P. L., & Vyse, T. J. (2008b) Fine-mapping the genetic basis of CRP regulation in African Americans: a Bayesian approach. *Hum.Genet.* 123: 633-642.
- Rhodes, B. & Vyse, T. J. (2007) General aspects of the genetics of SLE. *Autoimmunity* 40: 550-559.
- Rhodes, B. & Vyse, T. J. (2008) The genetics of SLE: an update in the light of genome-wide association studies. *Rheumatology.(Oxford)*.
- Rhodes, B., Wong, A., Navarra, S. V., Villamin, C., & Vyse, T. J. (2008c) Genetic determinants of basal C-reactive protein expression in Filipino systemic lupus erythematosus families. *Genes Immun.* 9: 153-160.
- Ridker, P. M., Pare, G., Parker, A., Zee, R. Y., Danik, J. S., Buring, J. E., Kwiatkowski, D., Cook, N. R., Miletich, J. P., & Chasman, D. I. (2008) Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. *Am.J.Hum.Genet.* 82: 1185-1192.
- Ridker, P. M., Rifai, N., Clearfield, M., Downs, J. R., Weis, S. E., Miles, J. S., & Gotto, A. M., Jr. (2001) Measurement of C-reactive protein for the targeting of statin therapy in the primary prevention of acute coronary events. *N.Engl.J.Med* 344: 1959-1965.
- Ridker, P. M., Rifai, N., Pfeffer, M. A., Sacks, F., & Braunwald, E. (1999) Long-term effects of pravastatin on plasma concentration of C-reactive protein. The Cholesterol and Recurrent Events (CARE) Investigators. *Circulation* 100: 230-235.
- Risch, N. (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am.J.Hum.Genet.* 46: 222-228.
- Ritchie, R. F., Palomaki, G. E., Neveux, L. M., Navolotskaia, O., Ledue, T. B., & Craig, W. Y. (2004) Reference distributions for complement proteins C3 and C4: a practical, simple and clinically relevant approach in a large cohort. *J.Clin.Lab Anal.* 18: 1-8.
- Rodriguez, W., Mold, C., Kataranovski, M., Hutt, J., Marnell, L. L., & Du Clos, T. W. (2005) Reversal of ongoing proteinuria in autoimmune mice by treatment with C-reactive protein. *Arthritis Rheum.* 52: 642-650.

- Rodriguez, W., Mold, C., Marnell, L. L., Hutt, J., Silverman, G. J., Tran, D., & Du Clos, T. W. (2006) Prevention and reversal of nephritis in MRL/lpr mice with a single injection of C-reactive protein. *Arthritis Rheum.* 54: 325-335.
- Rogus, J., Beck, J. D., Offenbacher, S., Huttner, K., Iacoviello, L., Latella, M. C., de, G. M., Wang, H. Y., Kornman, K. S., & Duff, G. W. (2008) IL1B gene promoter haplotype pairs predict clinical levels of interleukin-1beta and C-reactive protein. *Hum.Genet.* 123: 387-398.
- Rontu, R., Ojala, P., Hervonen, A., Goebeler, S., Karhunen, P. J., Nikkila, M., Kunnas, T., Jylha, M., Eklund, C., Hurme, M., & Lehtimäki, T. (2006) Apolipoprotein E genotype is related to plasma levels of C-reactive protein and lipids and to longevity in nonagenarians. *Clin.Endocrinol.(Oxf)* 64: 265-270.
- Ropes, M. W., Rossmeisl, E., & Bauer, W. (1939) The relationship between the erythrocyte sedimentation rate and the plasma proteins. *J.Clin.Invest* 18: 791-798.
- Ross, G. D. & Lambris, J. D. (1982) Identification of a C3bi-specific membrane complement receptor that is expressed on lymphocytes, monocytes, neutrophils, and erythrocytes. *J.Exp.Med* 155: 96-110.
- Ross, G. D. & Vetvicka, V. (1993) CR3 (CD11b, CD18): a phagocyte and NK cell membrane receptor with multiple ligand specificities and functions. *Clin.Exp.Immunol.* 92: 181-184.
- Ross, P., Hall, L., Smirnov, I., & Haff, L. (1998) High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat.Biotechnol.* 16: 1347-1351.
- Ross, R. (1999) Atherosclerosis--an inflammatory disease. *N.Engl.J.Med* 340: 115-126.
- Ross, S. C. & Densen, P. (1984) Complement deficiency states and infection: epidemiology, pathogenesis and consequences of neisserial and other infections in an immune deficiency. *Medicine (Baltimore)* 63: 243-273.
- Russell, A. I., Cunninghame Graham, D. S., Shepherd, C., Robertson, C. A., Whittaker, J., Meeks, J., Powell, R. J., Isenberg, D. A., Walport, M. J., & Vyse, T. J. (2004) Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. *Hum.Mol.Genet.* 13: 137-147.
- Sabatti, C., Service, S. K., Hartikainen, A. L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C. G., Zaitlen, N. A., Varilo, T., Kaakinen, M., Sovio, U., Ruokonen, A., Laitinen, J., Jakkula, E., Coin, L., Hoggart, C., Collins, A., Turunen, H., Gabriel, S., Elliot, P.*et al.* (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat.Genet.* 41: 35-46.
- Sainz, J., Perez, E., Gomez-Lopera, S., & Jurado, M. (2008) IL1 gene cluster polymorphisms and its haplotypes may predict the risk to develop invasive pulmonary aspergillosis and modulate C-reactive protein level. *J.Clin.Immunol.* 28: 473-485.
- Samuel, J. M., Kelberman, D., Smith, A. J., Humphries, S. E., & Woo, P. (2008) Identification of a novel regulatory region in the interleukin-6 gene promoter. *Cytokine* 42: 256-264.

- Sankoh, A. J., Huque, M. F., & Dubey, S. D. (1997) Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat.Med* 16: 2529-2542.
- Saunders, C. L. & Gulliford, M. C. (2006) Heritabilities and shared environmental effects were estimated from household clustering in national health survey data. *J.Clin.Epidemiol.* 59: 1191-1198.
- Scheet, P. & Stephens, M. (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am.J.Hum.Genet.* 78: 629-644.
- Scheuner, M. T. (2001) Genetic predisposition to coronary artery disease. *Curr.Opin.Cardiol.* 16: 251-260.
- Schifferli, J. A., Ng, Y. C., Estreicher, J., & Walport, M. J. (1988) The clearance of tetanus toxoid/anti-tetanus toxoid immune complexes from the circulation of humans. Complement- and erythrocyte complement receptor 1-dependent mechanisms. *J.Immunol.* 140: 899-904.
- Schifferli, J. A., Steiger, G., Paccaud, J. P., Sjöholm, A. G., & Hauptmann, G. (1986) Difference in the biological properties of the two forms of the fourth component of human complement (C4). *Clin.Exp.Immunol.* 63: 473-477.
- Schifferli, J. A., Steiger, G., & Schapira, M. (1985) The role of C1, C1-inactivator and C4 in modulating immune precipitation. *Clin.Exp.Immunol.* 60: 605-612.
- Schifferli, J. A., Woo, P., & Peters, D. K. (1982) Complement-mediated inhibition of immune precipitation. I. Role of the classical and alternative pathways. *Clin.Exp.Immunol.* 47: 555-562.
- Schwedler, S. B., Amann, K., Wernicke, K., Krebs, A., Nauck, M., Wanner, C., Potempa, L. A., & Galle, J. (2005) Native C-reactive protein increases whereas modified C-reactive protein reduces atherosclerosis in apolipoprotein E-knockout mice. *Circulation* 112: 1016-1023.
- Seldin, M. F., Shigeta, R., Villoslada, P., Selmi, C., Tuomilehto, J., Silva, G., Belmont, J. W., Klareskog, L., & Gregersen, P. K. (2006) European population substructure: clustering of northern and southern populations. *PLoS.Genet.* 2: e143.
- Sengelov, H., Kjeldsen, L., Kroeze, W., Berger, M., & Borregaard, N. (1994) Secretory vesicles are the intracellular reservoir of complement receptor 1 in human neutrophils. *J.Immunol.* 153: 804-810.
- Sestak, A. L., Nath, S. K., Sawalha, A. H., & Harley, J. B. (2007) Current status of lupus genetics. *Arthritis Res.Ther.* 9: 210.
- Shih, P. B., Manzi, S., Shaw, P., Kenney, M., Kao, A. H., Bontempo, F., Barmada, M. M., Kammerer, C., & Kamboh, M. I. (2008) Genetic variation in C-reactive protein (CRP) gene may be associated with risk of systemic lupus erythematosus and CRP concentrations. *J.Rheumatol.* 35: 2171-2178.
- Shine, B., de Beer, F. C., & Pepys, M. B. (1981) Solid phase radioimmunoassays for human C-reactive protein. *Clin.Chim.Acta* 117: 13-23.

- Shrive, A. K., Cheetham, G. M., Holden, D., Myles, D. A., Turnell, W. G., Volanakis, J. E., Pepys, M. B., Bloomer, A. C., & Greenhough, T. J. (1996) Three dimensional structure of human C-reactive protein. *Nat.Struct.Biol.* 3: 346-354.
- Silman, A. J., MacGregor, A. J., Thomson, W., Holligan, S., Carthy, D., Farhan, A., & Ollier, W. E. (1993) Twin concordance rates for rheumatoid arthritis: results from a nationwide study. *Br.J.Rheumatol.* 32: 903-907.
- Singh, S. K., Suresh, M. V., Prayther, D. C., Moorman, J. P., Rusinol, A. E., & Agrawal, A. (2008a) C-reactive protein-bound enzymatically modified low-density lipoprotein does not transform macrophages into foam cells. *J.Immunol.* 180: 4316-4322.
- Singh, S. K., Suresh, M. V., Voleti, B., & Agrawal, A. (2008b) The connection between C-reactive protein and atherosclerosis. *Ann.Med* 40: 110-120.
- Sinosich, M. J., Teisner, B., Brandslund, I., Fisher, M., & Grudzinskas, J. G. (1982) Influence of time, temperature and coagulation on the measurement of C3, C3 split products and C4. *J.Immunol.Methods* 55: 107-114.
- Sirito, M., Lin, Q., Deng, J. M., Behringer, R. R., & Sawadogo, M. (1998) Overlapping roles and asymmetrical cross-regulation of the USF proteins in mice. *Proc.Natl.Acad.Sci.U.S.A* 95: 3758-3763.
- Sjoholm, A. G. (1975) Complement components in normal serum and plasma quantitated by electroimmunoassay. *Scand.J.Immunol.* 4: 25-30.
- Sjowall, C., Bengtsson, A. A., Sturfelt, G., & Skogh, T. (2005) Anti-CRP autoantibody levels correlate with disease activity in systemic lupus erythematosus. *Semin.Arthritis Rheum.* 35: 65.
- Sjowall, C., Eriksson, P., Almer, S., & Skogh, T. (2002) Autoantibodies to C-reactive protein is a common finding in SLE, but not in primary Sjogren's syndrome, rheumatoid arthritis or inflammatory bowel disease. *J.Autoimmun.* 19: 155-160.
- Sjowall, C. & Wettero, J. (2007) Pathogenic implications for autoantibodies against C-reactive protein and other acute phase proteins. *Clin.Chim.Acta* 378: 13-23.
- Soilu-Hanninen, M., Koskinen, J. O., Laaksonen, M., Hanninen, A., Lilius, E. M., & Waris, M. (2005) High sensitivity measurement of CRP and disease progression in multiple sclerosis. *Neurology* 65: 153-155.
- Spath, P. & Gabl, F. (1976) Critical role of the conversion of the third complement component C3 (beta 1C/beta 1A) for its immunochemical quantitation. *Clin.Chim.Acta* 73: 171-175.
- Spielman, R. S., McGinnis, R. E., & Ewens, W. J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am.J.Hum.Genet.* 52: 506-516.
- Steer, S., Lad, B., Grumley, J. A., Kingsley, G. H., & Fisher, S. A. (2005) Association of R602W in a protein tyrosine phosphatase gene with a high risk of rheumatoid arthritis in a British population: evidence for an early onset/disease severity effect. *Arthritis Rheum.* 52: 358-360.

- Stephens, M. & Donnelly, P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am.J.Hum.Genet.* 73: 1162-1169.
- Stephens, M., Smith, N. J., & Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am.J.Hum.Genet.* 68: 978-989.
- Stewart, C. A., Horton, R., Allcock, R. J., Ashurst, J. L., Atrazhev, A. M., Coggill, P., Dunham, I., Forbes, S., Halls, K., Howson, J. M., Humphray, S. J., Hunt, S., Mungall, A. J., Osoegawa, K., Palmer, S., Roberts, A. N., Rogers, J., Sims, S., Wang, Y., Wilming, L. G. *et al.* (2004) Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res.* 14: 1176-1187.
- Storm, N., rnhofer-Patel, B., van den, B. D., & Rodi, C. P. (2003) MALDI-TOF mass spectrometry-based SNP genotyping. *Methods Mol.Biol.* 212: 241-262.
- Strunk, R. C., Eidlen, D. M., & Mason, R. J. (1988) Pulmonary alveolar type II epithelial cells synthesize and secrete proteins of the classical and alternative complement pathways. *J.Clin.Invest* 81: 1419-1426.
- Suk, D. J., Chasman, D. I., Cannon, C. P., Miller, D. T., Zee, R. Y., Kozlowski, P., Kwiatkowski, D. J., & Ridker, P. M. (2006) Influence of genetic variation in the C-reactive protein gene on the inflammatory response during and after acute coronary ischemia. *Ann.Hum.Genet.* 70: 705-716.
- Suk, H. J., Ridker, P. M., Cook, N. R., & Zee, R. Y. (2005) Relation of polymorphism within the C-reactive protein gene and plasma CRP levels. *Atherosclerosis* 178: 139-145.
- Szalai, A. J., Alarcon, G. S., Calvo-Alen, J., Toloza, S. M., McCrory, M. A., Edberg, J. C., McGwin, G., Jr., Bastian, H. M., Fessler, B. J., Vila, L. M., Kimberly, R. P., & Reveille, J. D. (2005a) Systemic lupus erythematosus in a multiethnic US Cohort (LUMINA). XXX: association between C-reactive protein (CRP) gene polymorphisms and vascular events. *Rheumatology.(Oxford)* 44: 864-868.
- Szalai, A. J., McCrory, M. A., Cooper, G. S., Wu, J., & Kimberly, R. P. (2002a) Association between baseline levels of C-reactive protein (CRP) and a dinucleotide repeat polymorphism in the intron of the CRP gene. *Genes Immun.* 3: 14-19.
- Szalai, A. J., Nataf, S., Hu, X. Z., & Barnum, S. R. (2002b) Experimental allergic encephalomyelitis is inhibited in transgenic mice expressing human C-reactive protein. *J.Immunol.* 168: 5792-5797.
- Szalai, A. J., VanCott, J. L., McGhee, J. R., Volanakis, J. E., & Benjamin, W. H., Jr. (2000) Human C-reactive protein is protective against fatal *Salmonella enterica* serovar typhimurium infection in transgenic mice. *Infect.Immun.* 68: 5652-5656.
- Szalai, A. J., Weaver, C. T., McCrory, M. A., van Ginkel, F. W., Reiman, R. M., Kearney, J. F., Marion, T. N., & Volanakis, J. E. (2003) Delayed lupus onset in (NZB x NZW)F1 mice expressing a human C-reactive protein transgene. *Arthritis Rheum.* 48: 1602-1611.
- Szalai, A. J., Wu, J., Lange, E. M., McCrory, M. A., Langefeld, C. D., Williams, A., Zakharkin, S. O., George, V., Allison, D. B., Cooper, G. S., Xie, F., Fan, Z., Edberg, J. C., & Kimberly, R. P. (2005b) Single-nucleotide polymorphisms in the C-reactive protein (CRP)

gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. *J.Mol.Med.* 83: 440-447.

Szalai, A. J., Wu, J., Lange, E. M., McCrory, M. A., Langefeld, C. D., Williams, A., Zakharkin, S. O., George, V., Allison, D. B., Cooper, G. S., Xie, F., Fan, Z., Edberg, J. C., & Kimberly, R. P. (2005c) Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. *J.Mol.Med* 83: 440-447.

Tan, E. M., Cohen, A. S., Fries, J. F., Masi, A. T., McShane, D. J., Rothfield, N. F., Schaller, J. G., Talal, N., & Winchester, R. J. (1982) The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* 25: 1271-1277.

Tang, K., Opalsky, D., Abel, K., van den Boom, D., Yip, P., Del Mistro, G., Braun, A., & Cantor, C. R. (2003) Single nucleotide polymorphism analysis by MALDI-TOF MS. *Int.J.Mass.Spectrom* 226: 37-54.

Taskinen, S., Hyvonen, M., Kovanen, P. T., Meri, S., & Pentikainen, M. O. (2005) C-reactive protein binds to the 3beta-OH group of cholesterol in LDL particles. *Biochem.Biophys.Res.Comm.* 329: 1208-1216.

Taskinen, S., Kovanen, P. T., Jarva, H., Meri, S., & Pentikainen, M. O. (2002) Binding of C-reactive protein to modified low-density-lipoprotein particles: identification of cholesterol as a novel ligand for C-reactive protein. *Biochem.J.* 367: 403-412.

Taylor, K. E., Giddings, J. C., & van den Berg, C. W. (2005) C-reactive protein-induced in vitro endothelial cell activation is an artefact caused by azide and lipopolysaccharide. *Arterioscler.Thromb.Vasc.Biol.* 25: 1225-1230.

Teare, D. M. & Barrett, J. H. (2005) Genetic linkage studies. *Lancet* 366: 1036-1044.

Templeton, A. R., Maxwell, T., Posada, D., Stengard, J. H., Boerwinkle, E., & Sing, C. F. (2005) Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. *Genetics* 169: 441-453.

Tennent, G. A., Hutchinson, W. L., Kahan, M. C., Hirschfield, G. M., Gallimore, J. R., Lewin, J., Sabin, C. A., Dhillon, A. P., & Pepys, M. B. (2008) Transgenic human CRP is not pro-atherogenic, pro-atherothrombotic or pro-inflammatory in apoE<sup>-/-</sup> mice. *Atherosclerosis* 196: 248-255.

Terry, C. F., Loukaci, V., & Green, F. R. (2000) Cooperative influence of genetic polymorphisms on interleukin 6 transcriptional regulation. *J.Biol.Chem.* 275: 18138-18144.

The Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437: 69-87.

The International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945.

Thompson, D., Pepys, M. B., & Wood, S. P. (1999) The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure.* 7: 169-177.

- Thompson, S. G., Kienast, J., Pyke, S. D., Haverkate, F., & van de Loo, J. C. (1995) Hemostatic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris. European Concerted Action on Thrombosis and Disabilities Angina Pectoris Study Group. *N.Engl.J.Med* 332: 635-641.
- Tian, C., Gregersen, P. K., & Seldin, M. F. (2008a) Accounting for ancestry: population substructure and genome-wide association studies. *Hum.Mol.Genet.* 17: R143-R150.
- Tian, C., Kosoy, R., Lee, A., Ransom, M., Belmont, J. W., Gregersen, P. K., & Seldin, M. F. (2008b) Analysis of East Asia genetic substructure using genome-wide SNP arrays. *PLoS.ONE.* 3: e3862.
- Tian, C., Plenge, R. M., Ransom, M., Lee, A., Villoslada, P., Selmi, C., Klareskog, L., Pulver, A. E., Qi, L., Gregersen, P. K., & Seldin, M. F. (2008c) Analysis and application of European genetic substructure using 300 K SNP information. *PLoS.Genet.* 4: e4.
- Tilg, H., Vannier, E., Vachino, G., Dinarello, C. A., & Mier, J. W. (1993) Antiinflammatory properties of hepatic acute phase proteins: preferential induction of interleukin 1 (IL-1) receptor antagonist over IL-1 beta synthesis by human peripheral blood mononuclear cells. *J.Exp.Med* 178: 1629-1636.
- Tillett, W. S. & Francis, T. (1930) Serological reactions in pneumonia with non-protein fraction of pneumococcus. *J.Exp.Med* 52: 561-571.
- Toloza, S. M., Uribe, A. G., McGwin, G., Jr., Alarcon, G. S., Fessler, B. J., Bastian, H. M., Vila, L. M., Wu, R., Shoenfeld, Y., Roseman, J. M., & Reveille, J. D. (2004) Systemic lupus erythematosus in a multiethnic US cohort (LUMINA). XXIII. Baseline predictors of vascular events. *Arthritis Rheum.* 50: 3947-3957.
- Trion, A., de Maat, M. P., Jukema, J. W., van der, L. A., Maas, M. C., Offerman, E. H., Havekes, L. M., Szalai, A. J., Princen, H. M., & Emeis, J. J. (2005) No effect of C-reactive protein on early atherosclerosis development in apolipoprotein E\*3-leiden/human C-reactive protein transgenic mice. *Arterioscler.Thromb.Vasc.Biol.* 25: 1635-1640.
- Tsai, C. J., Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M., & Nussinov, R. (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J.Mol.Biol.* 383: 281-291.
- Turner, M. W. & Hamvas, R. M. (2000) Mannose-binding lectin: structure, function, genetics and disease associations. *Rev.Immunogenet.* 2: 305-322.
- Ueda, S., Ikeda, U., Yamamoto, K., Takahashi, M., Nishinaga, M., Nago, N., & Shimada, K. (1996) C-reactive protein as a predictor of cardiac rupture after acute myocardial infarction. *Am.Heart J.* 131: 857-860.
- Ueki, A., Sai, T., Oka, H., Tabata, M., Hosokawa, K., & Mochizuki, Y. (1987) Biosynthesis and secretion of the third component of complement by human endothelial cells in vitro. *Immunology* 61: 11-14.
- Valentijn, R. M., van, O. H., Hazevoet, H. M., Hermans, J., Cats, A., Daha, M. R., & van ES, L. A. (1985) The value of complement and immune complex determinations in monitoring disease activity in patients with systemic lupus erythematosus. *Arthritis Rheum.* 28: 904-913.

- van der Helm-van Mil, A., Detert, J., le, C. S., Filer, A., Bastian, H., Burmester, G. R., Huizinga, T. W., & Raza, K. (2008) Validation of a prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: moving toward individualized treatment decision-making. *Arthritis Rheum.* 58: 2241-2247.
- van der Helm-van Mil, A., le, C. S., van, D. H., Breedveld, F. C., Toes, R. E., & Huizinga, T. W. (2007) A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: how to guide individual treatment decisions. *Arthritis Rheum.* 56: 433-440.
- van der Vieren, Le, T. H., Wood, C. L., Moore, P. F., St, J. T., Staunton, D. E., & Gallatin, W. M. (1995) A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-3. *Immunity.* 3: 683-690.
- van Lookeren, C. M., Wiesmann, C., & Brown, E. J. (2007) Macrophage complement receptors and pathogen clearance. *Cell Microbiol.* 9: 2095-2102.
- van Rossum, E. F. & Lamberts, S. W. (2004) Polymorphisms in the glucocorticoid receptor gene and their associations with metabolic parameters and body composition. *Recent Prog.Horm.Res.* 59: 333-357.
- Venugopal, S. K., Devaraj, S., & Jialal, I. (2005) Macrophage conditioned medium induces the expression of C-reactive protein in human aortic endothelial cells: potential for paracrine/autocrine effects. *Am.J.Pathol.* 166: 1265-1271.
- Verbovetski, I., Bychkov, H., Trahtemberg, U., Shapira, I., Hareuveni, M., Ben-Tal, O., Kutikov, I., Gill, O., & Mevorach, D. (2002) Opsonization of apoptotic cells by autologous iC3b facilitates clearance by immature dendritic cells, down-regulates DR and CD86, and up-regulates CC chemokine receptor 7. *J.Exp.Med* 196: 1553-1561.
- Verzilli, C., Shah, T., Casas, J. P., Chapman, J., Sandhu, M., Debenham, S. L., Boekholdt, M. S., Khaw, K. T., Wareham, N. J., Judson, R., Benjamin, E. J., Kathiresan, S., Larson, M. G., Rong, J., Sofat, R., Humphries, S. E., Smeeth, L., Cavalleri, G., Whittaker, J. C., & Hingorani, A. D. (2008) Bayesian meta-analysis of genetic association studies with different sets of markers. *Am.J.Hum.Genet.* 82: 859-872.
- Vickers, M. A., Green, F. R., Terry, C., Mayosi, B. M., Julier, C., Lathrop, M., Ratcliffe, P. J., Watkins, H. C., & Keavney, B. (2002) Genotype at a promoter polymorphism of the interleukin-6 gene is associated with baseline levels of plasma C-reactive protein. *Cardiovasc.Res.* 53: 1029-1034.
- Viglianti, G. A., Lau, C. M., Hanley, T. M., Miko, B. A., Shlomchik, M. J., & Marshak-Rothstein, A. (2003) Activation of autoreactive B cells by CpG dsDNA. *Immunity.* 19: 837-847.
- Vigushin, D. M., Pepys, M. B., & Hawkins, P. N. (1993) Metabolic and scintigraphic studies of radioiodinated human C-reactive protein in health and disease. *J.Clin.Invest* 91: 1351-1357.
- Visser, P. M., Hill, W. G., & Wray, N. R. (2008) Heritability in the genomics era--concepts and misconceptions. *Nat.Rev.Genet.* 9: 255-266.

- Vivier, E., Tomasello, E., Baratin, M., Walzer, T., & Ugolini, S. (2008) Functions of natural killer cells. *Nat.Immunol.* 9: 503-510.
- Vladutiu, A. O. & Winiarski, B. M. (1976) Complement C3 in serum and plasma, as measured by radial immunodiffusion with four commercial kits. *Clin.Chem.* 22: 267-269.
- Volanakis, J. E. (1982) Complement activation by C-reactive protein complexes. *Ann.N.Y.Acad.Sci.* 389: 235-250.
- Volanakis, J. E. & Kaplan, M. H. (1971) Specificity of C-reactive protein for choline phosphate residues of pneumococcal C-polysaccharide. *Proc.Soc.Exp.Biol.Med* 136: 612-614.
- Volanakis, J. E. & Wirtz, K. W. (1979) Interaction of C-reactive protein with artificial phosphatidylcholine bilayers. *Nature* 281: 155-157.
- Voll, R. E., Herrmann, M., Roth, E. A., Stach, C., Kalden, J. R., & Girkontaite, I. (1997) Immunosuppressive effects of apoptotic cells. *Nature* 390: 350-351.
- Walport, M. J. (2001) Complement. First of two parts. *N.Engl.J.Med* 344: 1058-1066.
- Walport, M. J., Davies, K. A., & Botto, M. (1998) C1q and systemic lupus erythematosus. *Immunobiology* 199: 265-285.
- Walston, J. D., Fallin, M. D., Cushman, M., Lange, L., Psaty, B., Jenny, N., Browner, W., Tracy, R., Durda, P., & Reiner, A. (2007) IL-6 gene variation is associated with IL-6 and C-reactive protein levels but not cardiovascular outcomes in the Cardiovascular Health Study. *Hum.Genet.* 122: 485-494.
- Wang, C. H., Li, S. H., Weisel, R. D., Fedak, P. W., Dumont, A. S., Szmítko, P., Li, R. K., Mickle, D. A., & Verma, S. (2003) C-reactive protein upregulates angiotensin type 1 receptors in vascular smooth muscle. *Circulation* 107: 1783-1790.
- Wang, Q., Hunt, S. C., Xu, Q., Chen, Y. E., Province, M. A., Eckfeldt, J. H., Pankow, J. S., & Song, Q. (2006) Association study of CRP gene polymorphisms with serum CRP level and cardiovascular risk in the NHLBI Family Heart Study. *Am.J.Physiol Heart Circ.Physiol* 291: H2752-H2757.
- Wang, Q., Zhu, X., Xu, Q., Ding, X., Chen, Y. E., & Song, Q. (2005) Effect of C-reactive protein on gene expression in vascular endothelial cells. *Am.J.Physiol Heart Circ.Physiol* 288: H1539-H1545.
- Warren, H. B., Pantazis, P., & Davies, P. F. (1987) The third component of complement is transcribed and secreted by cultured human endothelial cells. *Am.J.Pathol.* 129: 9-13.
- Wassel Fyr, C. L., Kanaya, A. M., Cummings, S. R., Reich, D., Hsueh, W. C., Reiner, A. P., Harris, T. B., Moffett, S., Li, R., Ding, J., Miljkovic-Gacic, I., & Ziv, E. (2007) Genetic admixture, adipocytokines, and adiposity in Black Americans: the Health, Aging, and Body Composition study. *Hum.Genet.* 121: 615-624.
- Webster, R. O., Heuertz, R., Xia, D., & Samols, D. (1994) Attenuation of complement-mediated acute lung injury in rabbits and transgenic mice by C-reactive protein. *Chest* 105: 101S.

- Weeks, D. E. & Lathrop, G. M. (1995) Polygenic disease: methods for mapping complex disease traits. *Trends Genet.* 11: 513-519.
- Weis, J. J., Tedder, T. F., & Fearon, D. T. (1984) Identification of a 145,000 Mr membrane protein as the C3d receptor (CR2) of human B lymphocytes. *Proc.Natl.Acad.Sci.U.S.A* 81: 881-885.
- Weiser, J. N. & Pan, N. (1998) Adaptation of *Haemophilus influenzae* to acquired and innate humoral immunity based on phase variation of lipopolysaccharide. *Mol.Microbiol.* 30: 767-775.
- Welch, T. R., Beischel, L., & Kleesattel, A. (1990) Functional consequences of the genetic polymorphism of the third component of complement. *J.Pediatr.* 116: S92-S97.
- Welch, T. R., Beischel, L. S., & Witte, D. P. (1993) Differential expression of complement C3 and C4 in the human kidney. *J.Clin.Invest* 92: 1451-1458.
- Wessel, J., Moratorio, G., Rao, F., Mahata, M., Zhang, L., Greene, W., Rana, B. K., Kennedy, B. P., Khandrika, S., Huang, P., Lillie, E. O., Shih, P. A., Smith, D. W., Wen, G., Hamilton, B. A., Ziegler, M. G., Witztum, J. L., Schork, N. J., Schmid-Schonbein, G. W., & O'Connor, D. T. (2007) C-reactive protein, an 'intermediate phenotype' for inflammation: human twin studies reveal heritability, association with blood pressure and the metabolic syndrome, and the influence of common polymorphism at catecholaminergic/beta-adrenergic pathway loci. *J.Hypertens.* 25: 329-343.
- Westergren, A. (1921) Studies of the suspension stability of blood in pulmonary tuberculosis. *Acta.Med.Scan* 54: 247-282.
- Westfall, P. H. & Young, S. S. 1993, *Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment*. Wiley, New York.
- Whicher, J. T., Price, C. P., & Spencer, K. (1983) Immunonephelometric and immunoturbidimetric assays for proteins. *Crit Rev.Clin.Lab Sci.* 18: 213-260.
- Wieme, R. J. & Demeulenaere, L. (1967) Genetically determined electrophoretic variant of the human complement component C'3. *Nature* 214: 1042-1043.
- Wiesmann, C., Katschke, K. J., Yin, J., Helmy, K. Y., Steffek, M., Fairbrother, W. J., McCallum, S. A., Embuscado, L., DeForge, L., Hass, P. E., & van Lookeren, C. M. (2006) Structure of C3b in complex with CRIg gives insights into regulation of complement activation. *Nature* 444: 217-220.
- Williams, R. C., Jr., Harmon, M. E., Burlingame, R., & Du Clos, T. W. (2005) Studies of serum C-reactive protein in systemic lupus erythematosus. *J.Rheumatol.* 32: 454-461.
- Wilson, D. R., Juan, T. S., Wilde, M. D., Fey, G. H., & Darlington, G. J. (1990) A 58-base-pair region of the human C3 gene confers synergistic inducibility by interleukin-1 and interleukin-6. *Mol.Cell Biol.* 10: 6181-6191.
- Wong, L. Y., Leung, R. Y., Ong, K. L., & Cheung, B. M. (2007) Plasma levels of fibrinogen and C-reactive protein are related to interleukin-6 gene -572C>G polymorphism in subjects with and without hypertension. *J.Hum.Hypertens.* 21: 875-882.

- Wordsworth, P. (1995) Genes and arthritis. *Br.Med Bull.* 51: 249-266.
- Wurzner, R., Orren, A., & Lachmann, P. J. (1992) Inherited deficiencies of the terminal components of human complement. *Immunodeficiency Rev.* 3: 123-147.
- Wuyts, B., Hetet, G., Grandchamp, B., & Delanghe, J. R. (2002) Novel haptoglobin insertion/deletion polymorphism is associated with the lipid profile and C-reactive protein (CRP) concentration. *Clin.Chem.Lab Med* 40: 469-474.
- Xavier, R. J. & Rioux, J. D. (2008) Genome-wide association studies: a new window into immune-mediated diseases. *Nat.Rev.Immunol.* 8: 631-643.
- Xia, D. & Samols, D. (1997) Transgenic mice expressing rabbit C-reactive protein are resistant to endotoxemia. *Proc.Natl.Acad.Sci.U.S.A* 94: 2575-2580.
- Xiong, Y., Cao, C., Makarova, A., Hyman, B., & Zhang, L. (2006) Mac-1 promotes FcγRIIA-dependent cell spreading and migration on immune complexes. *Biochemistry* 45: 8721-8731.
- Yanbaeva, D. G., Dentener, M. A., Spruit, M. A., Houwing-Duistermaat, J. J., Kotz, D., Passos, V. L., & Wouters, E. F. (2009) IL6 and CRP haplotypes are associated with COPD risk and systemic inflammation: a case-control study. *BMC.Med Genet.* 10: 23.
- Yang, Y., Chung, E. K., Wu, Y. L., Savelli, S. L., Nagaraja, H. N., Zhou, B., Hebert, M., Jones, K. N., Shu, Y., Kitzmiller, K., Blanchong, C. A., McBride, K. L., Higgins, G. C., Rennebohm, R. M., Rice, R. R., Hackshaw, K. V., Roubey, R. A., Grossman, J. M., Tsao, B. P., Birmingham, D. J. *et al.* (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am.J.Hum.Genet.* 80: 1037-1054.
- Yasojima, K., Schwab, C., McGeer, E. G., & McGeer, P. L. (2000) Human neurons generate C-reactive protein and amyloid P: upregulation in Alzheimer's disease. *Brain Res.* 887: 80-89.
- Yates, J. R., Sepp, T., Matharu, B. K., Khan, J. C., Thurlby, D. A., Shahid, H., Clayton, D. G., Hayward, C., Morgan, J., Wright, A. F., Armbricht, A. M., Dhillon, B., Deary, I. J., Redmond, E., Bird, A. C., & Moore, A. T. (2007) Complement C3 variant and the risk of age-related macular degeneration. *N.Engl.J.Med* 357: 553-561.
- Young, D. P., Kushner, I., & Samols, D. (2008) Binding of C/EBPβ to the C-reactive protein (CRP) promoter in Hep3B cells is associated with transcription of CRP mRNA. *J.Immunol.* 181: 2420-2427.
- Zacho, J., Tybjaerg-Hansen, A., Jensen, J. S., Grande, P., Sillesen, H., & Nordestgaard, B. G. (2008) Genetically elevated C-reactive protein and ischemic vascular disease. *N.Engl.J.Med* 359: 1897-1908.
- Zee, R. Y. & Ridker, P. M. (2002) Polymorphism in the human C-reactive protein (CRP) gene, plasma concentrations of CRP, and the risk of future arterial thrombosis. *Atherosclerosis* 162: 217-219.

- Zhang, D., Sun, M., Samols, D., & Kushner, I. (1996) STAT3 participates in transcriptional activation of the C-reactive protein gene by interleukin-6. *J.Biol.Chem.* 271: 9503-9509.
- Zhang, Y. X., Cliff, W. J., Schoefl, G. I., & Higgins, G. (1999) Coronary C-reactive protein distribution: its relation to development of atherosclerosis. *Atherosclerosis* 145: 375-379.
- Zhang, Y. Y., Gottardo, L., Mlynarski, W., Frazier, W., Nolan, D., Duffy, J., Marescotti, M. C., Gervino, E. V., Johnstone, M. T., Mantzoros, C. S., Avogaro, A., & Doria, A. (2007) Genetic variability at the leptin receptor (LEPR) locus is a determinant of plasma fibrinogen and C-reactive protein levels. *Atherosclerosis* 191: 121-127.
- Zhao, Y. X., Andoh, A., Shimada, M., Takaya, H., Hata, K., Fujiyama, Y., & Bamda, T. (2000) Secretion of complement components of the alternative pathway (C3 and factor B) by the human alveolar type II epithelial cell line A549. *Int.J.Mol.Med* 5: 415-419.
- Zlonis, M. (1993) The mystique of the erythrocyte sedimentation rate. A reappraisal of one of the oldest laboratory tests still in use. *Clin.Lab Med* 13: 787-800.
- Zondervan, K. T. & Cardon, L. R. (2004) The complex interplay among factors that influence allelic association. *Nat.Rev.Genet.* 5: 89-100.
- Zouki, C., Beauchamp, M., Baron, C., & Filep, J. G. (1997) Prevention of In vitro neutrophil adhesion to endothelial cells through shedding of L-selectin by C-reactive protein and peptides derived from C-reactive protein. *J.Clin.Invest* 100: 522-529.
- Zwaka, T. P., Hombach, V., & Torzewski, J. (2001) C-reactive protein-mediated low density lipoprotein uptake by macrophages: implications for atherosclerosis. *Circulation* 103: 1194-1197.

## **APPENDICES**

## A – Sequenom reagents and reaction conditions

### *PCR reagents (HME protocol)*

Using 2.5ng template DNA dissolved in 2 $\mu$ L TE buffer.

Reagents per reaction for a 3-6plex PCR making a total reaction volume of 5 $\mu$ L

| Reagent                                       | Volume/reaction |
|-----------------------------------------------|-----------------|
| Nanopure H <sub>2</sub> O                     | 1.930 $\mu$ L   |
| 10x PCR Buffer (inc. 15mM MgCl <sub>2</sub> ) | 0.500 $\mu$ L   |
| 25mM MgCl <sub>2</sub>                        | 0.200 $\mu$ L   |
| 10mM dNTPs                                    | 0.100 $\mu$ L   |
| PCR Primer Mix (100nM each)                   | 0.250 $\mu$ L   |
| HotstarTaq DNA polymerase(5U/ $\mu$ L )       | 0.020 $\mu$ L   |

### *PCR reagents (iPLEX and iPLEX gold protocols)*

Using 10ng template DNA dissolved in 2 $\mu$ L TE buffer.

Reagents per reaction making a total reaction volume of 5 $\mu$ L

| Reagent                                       | Volume/reaction |
|-----------------------------------------------|-----------------|
| Nanopure H <sub>2</sub> O                     | 0.850 $\mu$ L   |
| 10x PCR Buffer (inc. 15mM MgCl <sub>2</sub> ) | 0.625 $\mu$ L   |
| 25mM MgCl <sub>2</sub>                        | 0.325 $\mu$ L   |
| 25mM dNTPs                                    | 0.100 $\mu$ L   |
| PCR Primer Mix (500nM each)                   | 1.000 $\mu$ L   |
| HotstarTaq DNA polymerase (5U/ $\mu$ L )*     | 0.100 $\mu$ L   |

\*increased to 1U/reaction for plexes >27.

*PCR step thermocycler conditions (all protocols)*

1. 95°C for 15 mins
2. 95°C for 20 seconds
3. 65°C for 30 seconds
4. 72°C for 1 minute
5. Steps 2 to 4 repeated 4 times
6. 95°C for 20 seconds
7. 58°C for 30 seconds
8. 72°C for 1 minute
9. Steps 6 to 8 repeated 4 times
10. 95°C for 20 seconds
11. 53°C for 30 seconds
12. 72°C for 1 minute
13. Steps 10 to 12 repeated 37 times
14. 72°C for 3 minutes

*SAP reagents (all protocols)*

Added to 5µL PCR reaction mix to form a 7µL SAP reaction.

| Reagent                             | Volume/reaction |
|-------------------------------------|-----------------|
| Nanopure H <sub>2</sub> O           | 1.530µL         |
| 10xSAP buffer                       | 0.170µL         |
| Shrimp Alkaline Phosphatase (1U/µL) | 0.300µL         |

### *SAP step thermocycler conditions (all protocols)*

1. 37°C for 20 minutes
2. 85°C for 5 minutes
3. 4°C forever

### *Hme Reaction Reagents*

Primer concentration altered according to plexing level. Total mix for 384 plate given. 2µL added to 7 µL SAP reaction mix to form a 9µL reaction.

|                           | 3-plex | 4-plex | 5-plex | 6-plex |
|---------------------------|--------|--------|--------|--------|
| Nanopure H <sub>2</sub> O | 925.3  | 898.9  | 872.5  | 846.1  |
| hME termination mix       | 105.9  | 105.9  | 105.9  | 105.9  |
| Thermosequenase (32U/µL)  | 9.6    | 9.6    | 9.6    | 9.6    |
| Primer Mix (100µM each)   | 79.2   | 105.6  | 132    | 158.4  |

### *Hme Reaction thermocycler conditions*

1. 94°C for 2 minutes
2. 94°C for 5 seconds
3. 52°C for 5 seconds
4. 72°C for 5 seconds
5. Steps 2 to 4 repeated 55 times
6. 4°C forever

### *iPLEX Reaction Reagents*

Added to 7µL SAP reaction mix to form a 9µL reaction.

| Reagent                   | Volume/reaction |
|---------------------------|-----------------|
| Nanopure H <sub>2</sub> O | 0.755 µL        |
| iPLEX Buffer (10x)        | 0.200 µL        |
| iPLEX termination mix     | 0.200 µL        |
| Primer Mix                | 0.804 µL        |
| iPLEX enzyme              | 0.041 µL        |

### *iPLEX Reaction thermocycler conditions*

7. 94°C for 30 seconds
8. 94°C for 5 seconds
9. 52°C for 5 seconds
10. 80°C for 5 seconds
11. Steps 3 to 4 repeated 5 times
12. Steps 2 to 5 repeated 40 times
13. 72°C for 3 minutes
14. 4°C forever

## **B – Radial immunodiffusion reagents.**

### *Complement fixation test buffer*

CFT buffer reconstituted from tablets (Oxoid Ltd, Cambridge, UK). A 1% solution contains:

|                   |          |
|-------------------|----------|
| Barbitone         | 0.575g/L |
| NaCl              | 8.5g/L   |
| MgCl <sub>2</sub> | 0.168g/L |
| CaCl <sub>2</sub> | 0.028g/L |
| Barbitone soluble | 0.185g/L |

### *Phosphate Buffered Saline*

PBS was reconstituted from tablets (Sigma-Aldrich, Gillingham, Dorset, UK) and pH adjusted using small aliquots of HCl. The reconstituted solution contained:

|                  |         |
|------------------|---------|
| Phosphate buffer | 0.01M   |
| KCl              | 0.0027M |
| NaCl             | 0.137M  |

### *Coomassie stain*

Coomassie stain was made by dissolving Coomassie Brilliant blue R250 (Sigma-Aldrich, Gillingham, Dorset, UK) using the following reagents:

|                               |       |
|-------------------------------|-------|
| Coomassie Brilliant blue R250 | 2.5g  |
| Ethanol                       | 225mL |
| ddH <sub>2</sub> O            | 225mL |
| Glacial Acetic acid           | 50mL  |

Destain solution contained ethanol, ddH<sub>2</sub>O and acetic acid only in the same ratio.

## C – UK and Filipino Recruitment Questionnaire

Project Title: **Genetic Susceptibility in Systemic Lupus Erythematosus**

Principal Investigator: T. Vyse

**Full Name:** \_\_\_\_\_

**Date of Birth (day/mo/yr)** \_\_\_\_\_ **Family #** \_\_\_\_\_ **Ind** \_\_\_\_\_  
Office use only

If you have read the Information Sheet and agree to participate in this study, please *sign* below.

**Signed:** \_\_\_\_\_ **Date:** \_\_\_\_\_

Please circle your **GRANDPARENTS** ethnic group: 1) Afro-Caribbean 2) Caucasian  
3) Asian (Indian/Pakistani) 4) Asian (Oriental) 5) Other (please specify)

When were you diagnosed having SLE? Month \_\_\_\_\_ Year \_\_\_\_\_ WEIGHT st  
lb \_\_\_\_\_

kg \_\_\_\_\_  
When did you first notice any symptoms? Month \_\_\_\_\_ Year \_\_\_\_\_

Do you smoke? YES NO HEIGHT m.....

If so, how many per day? .....

**Please answer each of the following questions carefully. If a question is unclear or you don't know the answer please write "unsure" in the space provided. Please feel free to call us for clarification.**

**Please circle your correct answer**

**1. Have you *ever* suffered from any type arterial disease:**

|                                |     |    |
|--------------------------------|-----|----|
| a) Stroke                      | YES | NO |
| b) TIA (minor stroke)          | YES | NO |
| c) Heart attack                | YES | NO |
| d) Angina (pain in your chest) | YES | NO |

Name

Family #

Individual #

**2. Lupus can sometimes affect your kidneys**

- |    |                                                               |     |    |
|----|---------------------------------------------------------------|-----|----|
| a) | Have you ever been told that Lupus has affected your kidneys? | YES | NO |
| b) | Have you had a renal (kidney) biopsy?                         | YES | NO |
|    | Date.....Hospital.....                                        |     |    |
| c) | Have you had any significant protein in your urine?           | YES | NO |

**3. Have you *ever* suffered from thrombosis (blood clots):**

- |    |                                                                       |       |    |
|----|-----------------------------------------------------------------------|-------|----|
| a) | Deep vein (leg) thrombosis                                            | YES   | NO |
|    | If YES, how many times?                                               | ..... |    |
| b) | Pulmonary (lung) embolism                                             | YES   | NO |
| c) | Have you <b>ever</b> had a thrombosis in any other part of your body? | YES   | NO |

|                                                                             |     |
|-----------------------------------------------------------------------------|-----|
| <b>4. Have you <i>ever</i> had a low platelet count (thrombocytopenia)?</b> | YES |
| NO                                                                          |     |

**5. Have you *ever* been diagnosed with:**

- |    |              |     |    |
|----|--------------|-----|----|
| a) | Pleurisy     | YES | NO |
| b) | Pericarditis | YES | NO |

If YES please provide details:

|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| <b>6. Have you <i>ever</i> been diagnosed with Raynaud's phenomenon</b> | YES |
| NO                                                                      |     |

(colour changes to your fingers and/or toes in cold weather)?

|                                                                |     |
|----------------------------------------------------------------|-----|
| <b>7. Have you <i>ever</i> had a DEXA (bone density) scan?</b> | YES |
| NO                                                             |     |

If so, please give Date and Hospital

|      |          |              |
|------|----------|--------------|
| Name | Family # | Individual # |
|------|----------|--------------|

**8. Have you *ever* had/suffered :**

|                                              |     |    |
|----------------------------------------------|-----|----|
| a) Serious depression (requiring medication) | YES | NO |
| b) Seizures (fits)                           | YES | NO |
| c) Hallucinations                            | YES | NO |

**9. Have you *ever* had a spontaneous miscarriage?** YES  
NO

If YES, how many? .....

**10. Do you have any children?**

If YES, please give their year of birth:

**11. Have you *ever* been diagnosed with Sjögren's syndrome** YES  
NO

(prolonged dry eyes and/or dry mouth) ?

**12. Have you more than once suffered oral (mouth) ulceration?** YES  
NO

**13. Have you *ever* suffered from arthritis in the hands, wrists or feet?** YES  
NO

If YES please provide details:

**14. Have you *ever* suffered from skin rashes?** YES  
NO

Are these skin rashes worsened or triggered by exposure to sunlight? YES NO

If YES please provide details:

**15. Have you even been admitted into hospital in the past five years due to lupus?**

Please provide details: YES NO

|      |          |              |
|------|----------|--------------|
| Name | Family # | Individual # |
|------|----------|--------------|

**Previous Medical History**

**16. Have you *ever* been diagnosed with any of the following:**

|                                                                           |     |    |
|---------------------------------------------------------------------------|-----|----|
| Systemic lupus erythematosus                                              | YES | NO |
| Rheumatoid arthritis                                                      | YES | NO |
| Diabetes mellitus                                                         | YES | NO |
| If YES, AGE of onset.....? Insulin therapy.....?                          |     |    |
| Thyroid diseases (Graves' disease, Hashimoto's thyroiditis)               | YES | NO |
| Multiple sclerosis                                                        | YES | NO |
| Optic neuritis                                                            | YES | NO |
| Inflammatory bowel disease (Crohn's disease or ulcerative colitis)        | YES | NO |
| Liver disease (chronic autoimmune hepatitis<br>primary biliary cirrhosis) | YES | NO |

**17. Do you have any other medical conditions?** YES

NO

Please provide details:

**18. Do you take any medications regularly?** YES

NO

Please provide details:

*For Lupus sufferer only:* Please write: the name of your hospital Lupus doctor(s) and your GP(s)

|                  |                 |          |
|------------------|-----------------|----------|
| Lupus doctor(s): | GP:             |          |
| Hospital:        | Hosp. Record No | Surgery: |
| Town             |                 | Address: |

## **D - Publications Resulting from Work Presented in Thesis**

**Rhodes B**, Meek J, Whittaker JC and Vyse TJ. (2008) Quantification of the genetic component of basal C-reactive protein expression in SLE nuclear families. *Ann Hum Gen.* 72: 611-620.

**Rhodes B**, Wong A, Navarra S, Villamin C and Vyse TJ. (2008) Genetic determinants of basal C-reactive protein expression in Filipino Systemic Lupus Erythematosus families. *Genes Immun.* 9: 153-160.

**Rhodes B**, Morris DL, Subrahmanyam L, Aubin C, Mendes de Leon CF, Kelly JF, Evans DA, Whittaker JC, Oksenberg JR, De Jager PL and Vyse TJ. (2008) Fine-mapping the genetic basis of CRP regulation in African-Americans: a Bayesian approach. *Hum Genet.* 123: 633-642.

Hunnangkul S\*, Nitsch D\*, **Rhodes B\***, Chadha S, Robertson CA, Pessôa-Lopes P, Norsworthy PJ, Fernando MMA, Mackworth-Young CP, Isenberg DA, Whittaker JC and Vyse TJ. (2008) Familial clustering of non-nuclear autoantibodies, C3 and C4 complement components in systemic lupus erythematosus. *Arthritis Rheum.* 58: 1116-1124. \*Equal first authors.

**Rhodes B**, Hunnangkul S, Morris DL, Hsaio L, Nitsch D, Whittaker JC and Vyse TJ. (2009) The Heritability and Genetics of Complement C3 expression in UK SLE Families. *Genes Immun.* 10: 525-530