

Indoles as gut microbial pharmaco-metabolites:  
Targets, mechanisms and consequences for obesity  
and cardiometabolic diseases

Thesis submitted by  
**Ana Luisa Fernandes Neves Soares**

For the degree of Doctor in Philosophy of Imperial College London  
and the Diploma of Imperial College London (DIC)

Supervisors  
Dr. Marc-Emmanuel Dumas  
Prof. Jeremy K. Nicholson

Computational and Systems Medicine  
Department of Surgery and Cancer  
Faculty of Medicine  
Imperial College London  
2017



The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence.

Researchers are free to copy, distribute, or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform, or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Copyright © 2017



*To my family.*



# Abstract

The gut microbiota play a pivotal role in the onset and development of obesity and cardiometabolic disorders. Obese subjects exhibit significant gut microbial changes, including enrichment for bacterial genes associated with the degradation of tryptophan, which is converted by gut bacteria into indoles. Indoles act as signalling molecules in the microbial-mammalian crosstalk; however, their human targets and mechanisms are not fully understood. In particular, the human kinome remains an unexplored repertoire. The overall hypothesis of this work is that indoles bind human kinases, impact signalling pathways and interfere with cellular processes relevant for obesity and cardiometabolic diseases, thus acting as chemical messengers in the microbiome-kinome crosstalk. Gut microbial degradation of tryptophan converges to the production of indoxyl sulfate and indole-3-acetate. Using  $^1\text{H}$  NMR profiling, indoxyl sulfate was found to be positively correlated with adiposity and plasma levels of leptin; its effect was further confirmed *in vitro*, where it increased both lipid accumulation and leptin production. A high-throughput kinase screening revealed that indoxyl sulfate has kinome-wide effects and was subjected to a network analysis to connect its pro-adipogenic effect. Kinase screening and molecular docking analysis suggest that another indolic compound, indole-3-acetate, inhibits VEGFR-2 - a kinase-linked receptor essential for angiogenesis and involved in several cardiometabolic disorders (*e.g.* atherosclerosis, diabetes, and hypertension). This inhibition was validated *in vitro* by demonstrating that indole-3-acetate reduces the phosphorylation of VEGFR-2 and downstream AKT signalling in endothelial cells. Consistently, indole-3-acetate inhibited endothelial cell migration and tube formation, which are required for angiogenesis *in vitro*, and decreased vascular network formation in an *in vivo* angiogenesis model. Altogether, this study unravels a unique microbiome-kinome crosstalk by demonstrating that specific gut microbiota-generated indoles can impact adipogenesis (indoxyl sulfate) and angiogenesis (indole-3-acetate), two key processes in obesity and cardiometabolic disorders.



# Acknowledgements

First and foremost, I would like to thank Dr. Marc E. Dumas and Prof. Jeremy K. Nicholson for maintaining a steady presence at all stages of my PhD, while providing guidance and trust to allow me to pursue my research interests. Thank you for your enthusiastic attitude about science and your persevering optimism about my work.

I would also like to thank Dr. Julien Chilloux for his mentorship, by teaching technical skills and always providing a supportive environment for open interaction and scientific reflection. My thanks to Dr. Lesley Hoyles for advising and providing support on technical aspects and for the fruitful scientific discussions.

My deepest gratitude to the PhD students that closely shared this journey with me: Andrea, Delfim, Magali, Badrin, Joram and Claire. We discussed science and co-authored papers, we found unexpected collaborations, we shared an endless number of tea-stained mugs – and, in our free time, we started a social start-up. Your energy, altruism and support made my PhD a very enjoyable journey. Thank you for all that, and for being there whenever life-outside-the-lab decided to throw some lemons. Many other people contributed to this work, including Dr. Antony Dona, Dr. Michael Osbourne, Prof. Nigel Gooderham, Prof. Sara Rankin, Dr. Antonis Myridakis and Dr. Laura Martinez-Gili.

Last but not least, a special thanks to the GABBA Fellowship Program and to the Portuguese Governmental Agency for Science and Technology (Fundacao para a Ciencia e Tecnologia, SFRH/BD/52036/2012), for giving me the chance to embrace this opportunity.



# Statement of Originality and Contributions of Others

I certify that this thesis, and the research to which it refers, are the product of my own work, conducted during my PhD years. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged, in accordance with the standard referencing practices of the discipline. Contributions of others are described below.

- Dr. Lesley Hoyles performed the BLASTP searches, provided scientific support for the design of the computational strategy used in Chapter 3 and provided constructive feedback on this thesis.
- Mrs. Andrea Rodriguez-Martinez also contributed to the design of the computational strategy and co-wrote the code used to generate the predictions (Chapter 3).
- Collaborators from the MetaHIT consortium performed phenotypic, biochemical and dietary assessments shown in Chapter 4, as described by Le Chatelier et al., 2013.
- DiscoverX and Kinexus provided high-throughput kinome screening services (Chapters 5 and 6).
- Mr. Arinbjorn Kolbeisson generated data and Professor Robert Glen provided supervision on the molecular docking models shown in Chapter 6.
- Mrs. Sophie Thayanithy contributed to the optimisation and validation of experimental conditions for phosphorylation assays and tube formation assays shown in Chapter 7.
- Dr. Suet Ping-Wong performed the subcutaneous injections in the *in vivo* validation study described in Chapter 8. Professor Sara Rankin provided access to the animal facilities for this experiment and scientific support on the design of the experiment and interpretation of its results.



# Table of Contents

|                                                            |    |
|------------------------------------------------------------|----|
| Abstract .....                                             | 7  |
| Acknowledgements .....                                     | 9  |
| Statement of Originality and Contributions of Others ..... | 11 |
| Table of Contents.....                                     | 13 |
| Table of Figures .....                                     | 21 |
| List of Tables .....                                       | 23 |
| List of Equations .....                                    | 25 |
| List of Abbreviations.....                                 | 27 |

## Chapter 1

### Introduction

|                                                                                                                                      |    |
|--------------------------------------------------------------------------------------------------------------------------------------|----|
| 1. Obesity and cardiometabolic diseases .....                                                                                        | 36 |
| 1.1 Obesity .....                                                                                                                    | 36 |
| 1.1.1 Clinical assessment of obesity .....                                                                                           | 37 |
| 1.1.2 Obesity, insulin resistance and metabolic syndrome .....                                                                       | 38 |
| 1.1.3 Basic mechanisms of obesity and energy homeostasis .....                                                                       | 39 |
| 1.1.4 The multifactorial aetiology of obesity .....                                                                                  | 42 |
| 1.2 Cardiovascular diseases .....                                                                                                    | 43 |
| 1.2.1 Clinical assessment of cardiovascular diseases .....                                                                           | 43 |
| 1.2.2 Basic mechanisms of cardiovascular diseases .....                                                                              | 44 |
| 1.2.3 The multifactorial aetiology of cardiovascular diseases .....                                                                  | 47 |
| 2. The gut microbiota.....                                                                                                           | 49 |
| 2.1 Composition of the gut microbiota.....                                                                                           | 49 |
| 2.2 Host and environmental factors affecting gut microbial ecology.....                                                              | 52 |
| 2.3 Gut microbial changes in obesity and cardiovascular disease .....                                                                | 54 |
| 2.4 Functions of the gut microbiota in health.....                                                                                   | 57 |
| 2.5 The microbial-mammalian signalling axis .....                                                                                    | 60 |
| 2.5.1 Gut microbial metabolites as signalling molecules.....                                                                         | 60 |
| 2.5.2 Types of receptors involved in cell signalling.....                                                                            | 60 |
| 2.5.3 The human kinome as an unexplored target repertoire .....                                                                      | 63 |
| 2.6 Indoles as a key family of gut microbial metabolites .....                                                                       | 64 |
| 2.6.1 Gut bacterial production and transport of indoles.....                                                                         | 64 |
| 2.6.2 Relevance of indoles to obesity and cardiovascular health .....                                                                | 65 |
| 3. A combined metabonomic and cell-based strategy to investigate the interactions between obesity and gut microbial metabolites..... | 67 |
| 4. Scope of the thesis: hypothesis and aims.....                                                                                     | 69 |

## Chapter 2

### General Methodology

|        |                                                                            |    |
|--------|----------------------------------------------------------------------------|----|
| 1.     | Cell lines and cell-based assays .....                                     | 74 |
| 1.1.   | 3T3-L1 .....                                                               | 74 |
| 1.1.1. | Cell viability .....                                                       | 75 |
| 1.1.2. | Lipid staining (Oil Red O) .....                                           | 75 |
| 1.2.   | Human Umbilical Vein Endothelial Cells (HUVEC) .....                       | 76 |
| 1.2.1. | Cell viability .....                                                       | 76 |
| 1.2.2. | Phosphorylation assays .....                                               | 76 |
| 1.2.3. | NO quantification .....                                                    | 78 |
| 1.2.4. | Cell migration assay .....                                                 | 79 |
| 1.2.5. | Tube formation assay .....                                                 | 80 |
| 2.     | <i>In vivo</i> mouse study: Matrigel subcutaneous plug .....               | 81 |
| 3.     | <sup>1</sup> H Nuclear Magnetic Resonance Spectroscopy .....               | 81 |
| 3.1.   | Principles of <sup>1</sup> H NMR Spectroscopy .....                        | 82 |
| 3.2.   | Monodimensional and multidimensional <sup>1</sup> H NMR Spectroscopy ..... | 85 |
| 3.3.   | Unsupervised multivariate statistical analysis .....                       | 86 |
| 3.4.   | Supervised multivariate statistical analysis .....                         | 87 |
| 4.     | Kinase screening .....                                                     | 90 |
| 5.     | Molecular docking .....                                                    | 91 |
| 6.     | Basic Local Alignment Search Tool (BLAST) .....                            | 92 |
| 7.     | Networks and graph theory .....                                            | 94 |
| 7.1.   | Graphs and networks .....                                                  | 94 |
| 7.2.   | Network optimisation and centrality measures .....                         | 95 |

## Chapter 3

### Overview of microbial production of indoles and prediction of species involved

|      |                                                                           |     |
|------|---------------------------------------------------------------------------|-----|
| 1.   | Introduction .....                                                        | 99  |
| 2.   | Materials and Methods .....                                               | 101 |
| 2.1. | Tryptophan reaction network .....                                         | 101 |
| 2.2. | Protein Basic Local Alignment Search Tool (BLASTP) .....                  | 101 |
| 2.3. | Subnetwork analysis .....                                                 | 102 |
| 3.   | Results .....                                                             | 103 |
| 3.1. | Contribution of prokaryotic metabolism to the production of indoles ..... | 103 |
| 3.2. | BLASTP as a tool to predict gut microbial species .....                   | 105 |
| 3.3. | Prokaryotic communities involved in the production of indole .....        | 107 |

|      |                                                                                                                            |     |
|------|----------------------------------------------------------------------------------------------------------------------------|-----|
| 3.4. | Prokaryotic communities involved in the production of indole-3-acetate .....                                               | 109 |
| 4.   | Discussion .....                                                                                                           | 111 |
| 4.1. | Gut prokaryotic metabolism controls the majority of the reactions leading to the production of indoles .....               | 111 |
| 4.2. | Indole is exclusively produced by prokaryotic metabolism: Proteobacteria and Bacteroidetes are the main contributors ..... | 112 |
| 4.3. | Proteobacteria and Firmicutes are the highest predicted contributors to the production of indole-3-acetate.....            | 113 |
| 5.   | Conclusion .....                                                                                                           | 115 |

## Chapter 4

### Contribution of indoles to the metabolic signature of human obesity

|      |                                                                                                                |     |
|------|----------------------------------------------------------------------------------------------------------------|-----|
| 1.   | Introduction.....                                                                                              | 119 |
| 2.   | Materials and Methods.....                                                                                     | 121 |
| 2.1. | Study participants.....                                                                                        | 121 |
| 2.2. | Phenotyping, biochemical and dietary assessment .....                                                          | 121 |
| 2.3. | Sample preparation for metabolic profiling .....                                                               | 122 |
| 2.4. | Spectra acquisition and pre-processing.....                                                                    | 122 |
| 2.5. | Metabolite assignment.....                                                                                     | 123 |
| 2.6. | Statistical analysis .....                                                                                     | 123 |
| 3.   | Results .....                                                                                                  | 124 |
| 3.1. | MetaHit cohort characterisation .....                                                                          | 124 |
| 3.2. | Identifying the optimal measure to investigate the association between obesity and gut microbial changes ..... | 126 |
| 3.3. | Investigating the metabolic profiles of patients with increased adiposity by 1H NMR ....                       | 128 |
| 3.4. | Exploring the relationship between obesity and indoxyl sulfate .....                                           | 133 |
| 3.5. | Assessing the effect of confounders, mediators and co-variables .....                                          | 134 |
| 4.   | Discussion .....                                                                                               | 137 |
| 4.1. | Whole body fat (%Fat) exhibits the most significant correlation with leptin and bacterial gene count .....     | 137 |
| 4.2. | %Fat is positively correlated with indoxyl sulfate .....                                                       | 138 |
| 4.3. | Levels of indoxyl sulfate are not correlated with tryptophan dietary intake .....                              | 139 |
| 5.   | Conclusion .....                                                                                               | 141 |

## Chapter 5

Validation of the pro-adipogenic effect of indoxyl sulfate *in vitro*

|      |                                                                                                            |     |
|------|------------------------------------------------------------------------------------------------------------|-----|
| 1.   | Introduction.....                                                                                          | 145 |
| 2.   | Materials and Methods.....                                                                                 | 150 |
| 2.1. | Cell culture.....                                                                                          | 150 |
| 2.2. | 3T3-L1 differentiation .....                                                                               | 150 |
| 2.3. | Cell viability assay .....                                                                                 | 150 |
| 2.4. | Lipogenesis assay .....                                                                                    | 151 |
| 2.5. | Adipocytokines quantification .....                                                                        | 151 |
| 2.6. | Kinome screening.....                                                                                      | 152 |
| 2.7. | Network analysis .....                                                                                     | 152 |
| 2.8. | Statistical analysis.....                                                                                  | 153 |
| 3.   | Results.....                                                                                               | 154 |
| 3.1. | Effect of indoxyl sulfate on 3T3-L1 viability.....                                                         | 154 |
| 3.2. | Effect of indoxyl sulfate on lipid accumulation .....                                                      | 155 |
| 3.3. | Effect of indoxyl sulfate on adipocytokines secretion .....                                                | 156 |
| 3.4. | Investigating the targets of indoxyl sulfate in the human kinome .....                                     | 157 |
| 3.5. | Network prediction of targets of indoxyl sulfate involved in lipogenesis-related signalling pathways ..... | 160 |
| 4.   | Discussion.....                                                                                            | 162 |
| 4.1. | Indoxyl sulfate does not affect 3T3-L1 cell viability .....                                                | 162 |
| 4.2. | Indoxyl sulfate promotes lipid accumulation in 3T3-L1.....                                                 | 163 |
| 4.3. | Advantages and limitations of cell-based assays in the study of adipocyte metabolism and obesity .....     | 164 |
| 4.4. | Network analysis predicts that indoxyl sulfate has a multi-kinase pro-adipogenic effect .....              | 164 |
| 5.   | Conclusion .....                                                                                           | 166 |

## Chapter 6

Exploring the effect of indole-3-acetate in the human kinome

|      |                                                       |     |
|------|-------------------------------------------------------|-----|
| 1.   | Introduction.....                                     | 169 |
| 2.   | Material and Methods .....                            | 171 |
| 2.1. | Kinase screening .....                                | 171 |
| 2.2. | Molecular docking.....                                | 171 |
| 2.3. | Cell culture.....                                     | 172 |
| 2.4. | Total and phosphorylated VEGFR-2 quantification ..... | 172 |
| 2.5. | AKT phosphorylation assay .....                       | 173 |
| 2.6. | NO quantification assay .....                         | 173 |
| 2.7. | Statistical analysis.....                             | 173 |

|      |                                                                                                        |     |
|------|--------------------------------------------------------------------------------------------------------|-----|
| 3.   | Results .....                                                                                          | 174 |
| 3.1. | Exploring the targets of indole-3-acetate in the human kinome .....                                    | 174 |
| 3.2. | Investigating the interaction between indole-3-acetate and VEGFR-2 by molecular docking analysis ..... | 175 |
| 3.3. | Optimisation of experimental conditions for phosphorylation assays .....                               | 178 |
| 3.4. | Effect of indole-3-acetate on VEGFR-2 phosphorylation .....                                            | 180 |
| 3.5. | Effect of indole-3-acetate on VEGFR-2 downstream signalling .....                                      | 181 |
| 4.   | Discussion .....                                                                                       | 183 |
| 4.1. | High-throughput kinase screening identified indole-3-acetate as a VEGFR-2 inhibitor.....               | 183 |
| 4.2. | Molecular docking suggests indole-3-acetate inhibits VEGFR-2 through an ATP-competitive mechanism..... | 184 |
| 4.3. | Indole-3-acetate inhibits VEGFR-2 phosphorylation in HUVEC .....                                       | 185 |
| 4.4. | Indole-3-acetate supresses VEGFA-induced AKT phosphorylation and NO production .....                   | 186 |
| 5.   | Conclusion .....                                                                                       | 187 |

## Chapter 7

### Effect of indole-3-acetate on angiogenesis *in vitro*

|        |                                                                      |     |
|--------|----------------------------------------------------------------------|-----|
| 1.     | Introduction.....                                                    | 191 |
| 2.     | Materials and Methods.....                                           | 195 |
| 2.1.   | Cell culture.....                                                    | 195 |
| 2.2.   | Cell migration assay .....                                           | 195 |
| 2.3.   | Tube formation assay .....                                           | 195 |
| 2.4.   | Statistical analysis .....                                           | 196 |
| 3.     | Results .....                                                        | 197 |
| 3.1.   | Effect of indole-3-acetate on cell migration.....                    | 197 |
| 3.2.   | Effect of indole-3-acetate on tube formation .....                   | 200 |
| 3.2.1. | Exploratory approach using unsupervised multivariate methods .....   | 201 |
| 3.2.2. | Univariate analysis of the constitutive elements of the network..... | 203 |
| 4.     | Discussion .....                                                     | 205 |
| 4.1.   | Indole-3-acetate inhibits cell migration in vitro .....              | 205 |
| 4.2.   | Indole-3-acetate inhibits tube formation in vitro .....              | 206 |
| 4.3.   | Angiogenesis assays in vitro: advantages and pitfalls .....          | 208 |
| 5.     | Conclusion .....                                                     | 209 |

## Chapter 8

### Effect of indole-3-acetate on angiogenesis *in vivo*

|      |                                                                |     |
|------|----------------------------------------------------------------|-----|
| 1.   | Introduction.....                                              | 213 |
| 2.   | Materials and Methods.....                                     | 215 |
| 2.1. | Animal studies.....                                            | 215 |
| 2.2. | Matrigel in vivo plug assay .....                              | 215 |
| 2.3. | Immunohistochemistry .....                                     | 216 |
| 2.4. | Imaging protocol.....                                          | 216 |
| 2.5. | Statistical analysis .....                                     | 218 |
| 3.   | Results.....                                                   | 218 |
| 3.1. | Effect of indole-3-acetate on plug weight .....                | 219 |
| 3.2. | Effect of indole-3-acetate on the vascular area.....           | 220 |
| 4.   | Discussion .....                                               | 223 |
| 4.1. | Indole-3-acetate reduces plug weight.....                      | 223 |
| 4.2. | Indole-3-acetate reduces the area of the vascular network..... | 223 |
| 5.   | Conclusion .....                                               | 225 |

## Chapter 9

### General Discussion

|      |                                                                                                                  |     |
|------|------------------------------------------------------------------------------------------------------------------|-----|
| 1.   | Key findings .....                                                                                               | 231 |
| 1.1. | Gut microbial degradation of tryptophan converges to the production of indoxyl sulfate and indole-3-acetate..... | 231 |
| 1.2. | Indoxyl sulfate exhibits pro-adipogenic properties through a multiple-kinase effect .....                        | 232 |
| 1.3. | Indole-3-acetate exhibits anti-angiogenic properties through a putative VEGFR-2 inhibition .....                 | 233 |
| 2.   | Experimental strengths and future work.....                                                                      | 234 |
| 3.   | Consequences for obesity and cardiometabolic diseases .....                                                      | 237 |
| 3.1. | A novel contribution of the gut microbiota to human obesity.....                                                 | 237 |
| 3.2. | Impact of inhibition of angiogenesis by gut microbial metabolites in human health and disease .....              | 238 |
| 3.3. | The gut microbial pharmacopeia as a potential therapeutic tool.....                                              | 239 |
| 3.4. | Gut microbial metabolites in risk stratification and personalised medicine.....                                  | 240 |
| 4.   | Conclusion .....                                                                                                 | 241 |

|                                            |     |
|--------------------------------------------|-----|
| List of published papers .....             | 245 |
| References.....                            | 247 |
| Appendix A: Published papers.....          | 276 |
| Appendix B: Reproduction permissions ..... | 285 |



# Table of Figures

|                                                                                                                                                                                      |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 1. Overview of glucose, free fatty acid and amino acid metabolism.....                                                                                                        | 44  |
| Figure 2. Cardiovascular complications associated with obesity and metabolic disorders. ....                                                                                         | 48  |
| Figure 3. Trilaminar vascular structure.....                                                                                                                                         | 49  |
| Figure 4. Longitudinal variation of the microbial composition in human distal gut.....                                                                                               | 55  |
| Figure 5. Functions of the healthy gut microbiota. ....                                                                                                                              | 62  |
| Figure 6. The microbial-mammalian signalling axis .....                                                                                                                              | 64  |
| Figure 7. General reaction catalysed by kinases. ....                                                                                                                                | 66  |
| Figure 8. Generation of indoles from tryptophan by gut bacteria.....                                                                                                                 | 69  |
| Figure 9. Thesis structure. ....                                                                                                                                                     | 74  |
| Figure 10. Principle of the MSD Phosphoprotein Assay. ....                                                                                                                           | 82  |
| Figure 11. Griess Reaction. ....                                                                                                                                                     | 79  |
| Figure 12. Cell migration assay: quantitative analysis. ....                                                                                                                         | 80  |
| Figure 13. Principles of $^1\text{H}$ NMR Spectroscopy. ....                                                                                                                         | 83  |
| Figure 14. Kinase screening by KINOMEScan technology. ....                                                                                                                           | 90  |
| Figure 15. Example of a simple directed graph. ....                                                                                                                                  | 95  |
| Figure 16. Overview of the production of indoles from tryptophan. ....                                                                                                               | 103 |
| Figure 17. Prediction of prokaryotic species involved in the production of indoles. ....                                                                                             | 110 |
| Figure 18. Contribution of the gut microbiota to the production of indole. ....                                                                                                      | 108 |
| Figure 19. Contribution of the gut microbiota to the production of indole-3-acetate.....                                                                                             | 110 |
| Figure 20. Correlation matrix between obesity parameter and other metabolic variables in the MetaHit study.....                                                                      | 125 |
| Figure 21. %Fat exhibits the strongest anti-correlation with bacterial gene count.....                                                                                               | 127 |
| Figure 22. A typical urine $^1\text{H}$ NMR spectrum. ....                                                                                                                           | 130 |
| Figure 23. Metabonomic signature of increased adiposity: Principal Component Analysis (PCA), Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) and skyline plot ..... | 132 |
| Figure 24. Confirmation of the assignment of indoxyl sulfate by a spike-in experiment.. ....                                                                                         | 133 |
| Figure 25. Correlations between indoxyl sulfate and other obesity-related parameters.....                                                                                            | 135 |
| Figure 26. Total correlations between indoxyl sulfate and dietary compounds. ....                                                                                                    | 136 |
| Figure 27. Schematic representation of the several cell types involved in obesity development <i>in vivo</i> .....                                                                   | 146 |
| Figure 28. Regulation of lipogenesis by extracellular factors.....                                                                                                                   | 152 |
| Figure 29. Effect of indoxyl sulfate on cell viability. ....                                                                                                                         | 154 |
| Figure 30. Indoxyl sulfate promotes lipogenesis in vitro in 3T3-L1. ....                                                                                                             | 155 |
| Figure 31. Effect of indoxyl sulfate on adipocytokine secretion in 3T3-L1. ....                                                                                                      | 156 |
| Figure 32. Indoxyl sulfate exhibits kinome-wide effects in a single-dose screening. ....                                                                                             | 158 |
| Figure 33. Indoxyl sulfate does not counteract the anti-adipogenic effect of PDGF-DD.....                                                                                            | 160 |
| Figure 34. Prediction of relevant signalling pathways and metabolic nodes involved in the pro-adipogenic effect of indoxyl sulfate. ....                                             | 161 |
| Figure 35. Indole-3-acetate binds VEGFR-2 and functionally inhibits its phosphorylation activity in a high-throughput kinase screening. ....                                         | 174 |

|                                                                                                    |     |
|----------------------------------------------------------------------------------------------------|-----|
| Figure 36. Indole-3-acetate binds to the ATP-binding pocket in molecular docking analysis. ....    | 176 |
| Figure 37. Indole-3-acetate prevents ATP binding in the ATP-binding pocket of VEGFR-2. ....        | 177 |
| Figure 38. Indole-3-acetate affects HUVEC viability at 1000 $\mu$ M. ....                          | 178 |
| Figure 39. Kinetics of total and Phospho(Tyr1175)-VEGFR-2 expression after stimulation. ....       | 179 |
| Figure 40. Effect of indole-3-acetate on VEGFR-2 phosphorylation. ....                             | 180 |
| Figure 41. Effect of indole-3-acetate on VEGFR-2 downstream signalling effectors. ....             | 182 |
| Figure 42. Hallmarks of angiogenesis. ....                                                         | 192 |
| Figure 43. Role of VEGF <sub>A</sub> /VEGFR-2 in key steps of angiogenesis. ....                   | 193 |
| Figure 44. Assessment of HUVEC migrating activity <i>in vitro</i> . ....                           | 197 |
| Figure 45. Indole-3-acetate inhibits cell migration in HUVEC. ....                                 | 199 |
| Figure 46. Assessment of tube formation by HUVEC <i>in vitro</i> . ....                            | 200 |
| Figure 47. Unsupervised multivariate analysis by PCA and O2PLS-DA. ....                            | 202 |
| Figure 48. Indole-3-acetate inhibits tube formation in HUVEC. ....                                 | 204 |
| Figure 49. Microscopic imaging protocol for the quantification of the CD31 <sup>+</sup> area. .... | 217 |
| Figure 50. Matrigel subcutaneous plug assay (control plug). ....                                   | 219 |
| Figure 51. Plug weight is lower in plugs treated with indole-3-acetate. ....                       | 220 |
| Figure 52. Macroscopic and microscopic overview of the Matrigel plugs. ....                        | 221 |
| Figure 53. Indole-3-acetate inhibits angiogenesis in a subcutaneous Matrigel plug model. ....      | 222 |

# List of Tables

|                                                                                                              |     |
|--------------------------------------------------------------------------------------------------------------|-----|
| Table 1. Overview of different types of Basic Local Alignment Search Tools.....                              | 93  |
| Table 2. List of the metabolic reactions involved in tryptophan metabolism. ....                             | 104 |
| Table 3. Metabolic summary description of the MetaHit study population.....                                  | 124 |
| Table 4. Spearman correlations between the percentage of body fat (%Fat) and other metabolic parameters..... | 126 |
| Table 5. Spearman correlations between obesity measures and gene count.....                                  | 128 |
| Table 6. List of most abundant metabolites assigned and respective chemical shift.....                       | 129 |
| Table 7. High binding affinity kinases for indoxyl sulfate, ranked by <i>IC</i> <sub>50</sub> .....          | 158 |
| Table 8. Betweenness ranks for the nodes of the shortest-paths subnetwork.....                               | 162 |
| Table 9. Best docking solutions for ATP and indole-3-acetate .....                                           | 176 |
| Table 10. Best docking solutions for ATP after previous docking of indole-3-acetate in VEGFR-2 .....         | 177 |



# List of Equations

|                                                                                                |     |
|------------------------------------------------------------------------------------------------|-----|
| Equation 1. Quantification of cell migrating activity. ....                                    | 80  |
| Equation 2. Boltzmann distribution.....                                                        | 83  |
| Equation 3. Energy distribution between high- and low-energy statuses.....                     | 84  |
| Equation 4. Construction of the Principal Component Analysis matrix.....                       | 91  |
| Equation 5. Quantification of primary screen binding interactions in the kinome screening..... | 91  |
| Equation 6. Betweenness centrality of a given node.....                                        | 96  |
| Equation 7. Calculation of the path-values.....                                                | 102 |
| Equation 8. Calculation of the individual bacterial score values.....                          | 102 |



# List of Abbreviations

|                |                                                           |
|----------------|-----------------------------------------------------------|
| 1D             | One-dimensional                                           |
| 2D             | Two-dimensional                                           |
| 4-AAP          | 4-Aminoantipyrine                                         |
| ADP            | Adenosine diphosphate                                     |
| AGC            | Protein kinase A, G and C families                        |
| AHR            | Aryl Hydrocarbon Receptor                                 |
| AKT            | Protein Kinase B                                          |
| ANOVA          | Analysis of Variance                                      |
| AMPK           | Adenosine Monophosphate-activated Protein Kinase          |
| API            | Application Programming Interface                         |
| Asn            | Asparagine                                                |
| Asp            | Aspartate                                                 |
| ATP            | Adenosine Triphosphate                                    |
| BH             | Benjamini-Hochberg                                        |
| BLAST          | Basic Local Alignment Search Tool                         |
| BLASTN         | Nucleotide-Nucleotide BLAST                               |
| BLASTP         | Protein-Protein BLAST                                     |
| BLASTX         | Nucleotide 6-Frame Translation Protein BLAST              |
| BMI            | Body Mass Index                                           |
| BMP            | Bone Morphogenetic Protein                                |
| C/EBP $\alpha$ | CCAAT-enhancer-binding protein $\alpha$ (C/EBP $\alpha$ ) |
| CAD            | Coronary Artery Disease                                   |
| CAM            | Chick Chorioallantoic Membrane                            |
| CAMK           | Calcium and Calmodulin-regulated Kinases                  |
| cAMP           | Cyclic Adenosine Monophosphate                            |
| CD31           | Cluster of Differentiation 31                             |
| CDC42          | Cell Division Control Protein 42 Homolog                  |
| CERB           | cAMP Response Element-Binding Protein                     |
| cGMP           | Cyclic Guanosine Monophosphate                            |
| CK1            | Cell Kinase 1                                             |
| c-KIT          | KIT Proto-Oncogene Receptor Tyrosine Kinase               |
| CLK3           | CDC Like Kinase 3                                         |
| COSY           | Correlation Spectroscopy                                  |
| CRP            | C-reactive Protein                                        |
| CutC           | Choline Trimethylamine-lyase                              |
| CVD            | Cardiovascular Disease                                    |
| DAG            | Diacylglycerol                                            |
| DALY           | Disability-Adjusted Life Year                             |

|                               |                                                                         |
|-------------------------------|-------------------------------------------------------------------------|
| DAP                           | Dihydroxyacetone Phosphate                                              |
| DAPI                          | 4-6-Diamidino-2-Phenylindole                                            |
| DEXA                          | Dual Energy X Ray Absorptiometry                                        |
| DMA                           | Dimethylamine                                                           |
| DMEM                          | Dulbecco' s Modified Eagle' s medium                                    |
| DNA                           | Deoxyribonucleic Acid                                                   |
| EBM                           | Endothelial Basal Medium                                                |
| <i>EC<sub>50</sub></i>        | Half Maximal Effective Concentration                                    |
| EGM                           | Endothelial Growth Medium                                               |
| EHS                           | Engelbreth-Holm-Swarm                                                   |
| ELISA                         | Enzyme-Linked Immunosorbent Assay                                       |
| eNOS                          | Endothelial Nitric Oxide Synthase                                       |
| ER                            | Endoplasmic reticulum                                                   |
| ERK                           | Extracellular Signal-Regulated Protein Kinase                           |
| ESPA                          | Sodium- <i>N</i> -ethyl- <i>N</i> -(3-sulfopropyl)- <i>m</i> -anisidine |
| FADH2                         | Flavin Adenine Dinucleotide (reduced)                                   |
| FDR                           | False Discovery Rate                                                    |
| FFA                           | Free Fatty Acid                                                         |
| FFAR2                         | Free Fatty Acid Receptor 2                                              |
| FFAR3                         | Free Fatty Acid Receptor 3                                              |
| FFM                           | Fat-Free Mass                                                           |
| FGF                           | Fibroblast Growth Factor                                                |
| FGFR                          | Fibroblast Growth Factor Receptor                                       |
| FIAF                          | Fasting-Induced Adipose Factor                                          |
| FID                           | Free Induction Decay                                                    |
| FXR                           | Farnesoid X Receptor                                                    |
| G-1-P                         | Glycerol-1-Phosphate                                                    |
| GPCR                          | G-Coupled Protein Receptor                                              |
| GK                            | Glycerol Kinase                                                         |
| GLK                           | Germinal Center Kinase-like Kinase                                      |
| GLUT2                         | Glucose Transporter 2                                                   |
| GPO                           | Glycerol Phosphate Oxidase                                              |
| GTP                           | Guanosine Triphosphate                                                  |
| GWAS                          | Genome Wide Association Study                                           |
| H&E                           | Haematoxylin and eosin                                                  |
| H <sub>2</sub> O <sub>2</sub> | Hydrogen Peroxide                                                       |
| HDL                           | High-Density Lipoprotein                                                |
| HMDB                          | Human Metabolome Database                                               |
| HMP                           | Human Microbiome Project                                                |
| HOMA-IR                       | Homeostatic Model Assessment for Insulin Resistance                     |
| HSL                           | Hormone-Sensitive Lipase                                                |
| HSQC                          | Heteronuclear Single Quantum Coherence Spectroscopy                     |
| HUVEC                         | Human Vascular Endothelial Cells                                        |

|                         |                                                            |
|-------------------------|------------------------------------------------------------|
| I3A                     | Indole-3-Acetate                                           |
| IBMX                    | 3-Isobutyl-1-methylxanthine                                |
| <i>IC</i> <sub>50</sub> | Half-Maximal Inhibitory Concentration                      |
| IDF                     | International Diabetes Federation                          |
| IgA                     | Immunoglobulin A                                           |
| IgG1                    | Immunoglobulin G, subclass G1                              |
| IgG2                    | Immunoglobulin G, subclass G2                              |
| IKK                     | Inhibitor of Kappa B-Kinase                                |
| IL-10                   | Interleukin-10                                             |
| IL-1 $\beta$            | Interleukin-1 $\beta$                                      |
| IL-6                    | Interleukin-6                                              |
| Ile                     | Isoleucine                                                 |
| IP3                     | Inositol 1,4,5-Triphosphate                                |
| IRS                     | Insulin Receptor Substrate                                 |
| IS                      | Indoxyl sulfate                                            |
| <i>J</i> -Res           | <i>J</i> -Resolved                                         |
| <i>KD</i>               | Dissociation Constant                                      |
| JNK                     | c-Jun <i>N</i> -terminal Kinase                            |
| KEGG                    | Kyoto Encyclopaedia of Genes and Genomes                   |
| LC-MS                   | Liquid Chromatography - Mass Spectrometry                  |
| LDL                     | Low-Density Lipoprotein                                    |
| LEF/TCF                 | T-Cell Factor/Lymphoid Enhancer-Binding Factor             |
| Leu                     | Leucine                                                    |
| LIMK1                   | LIM Domain Kinase 1                                        |
| LPS                     | Lipopolysaccharide                                         |
| LRP                     | Lipoprotein Receptor-Related Protein Receptor              |
| MAPK                    | Mitogen-Activated Protein Kinases                          |
| MAPK14                  | Mitogen-Activated Protein Kinase Kinase 14                 |
| MAPK2K6                 | Dual Specificity Mitogen-Activated Protein Kinase Kinase 6 |
| MARKCS                  | Myristoylated Alanine-Rich Protein Kinase C Substrate      |
| MCP-1                   | Monocyte Chemoattractant Protein-1                         |
| MMP                     | Matrix Metalloproteinase                                   |
| MRCKA                   | CDC42 Binding Protein Kinase $\alpha$                      |
| MRCKB                   | CDC42 Binding Protein Kinase $\beta$                       |
| MSD                     | Meso Scale Discovery                                       |
| MYLK                    | Myosin Light Chain Kinase                                  |
| MYLK12B                 | Myosin Light Chain Kinase 12B                              |
| MYLK3                   | Myosin Light Chain Kinase 3                                |
| NADH                    | Nicotinamide Adenine Dinucleotide (reduced)                |
| NMR                     | Nuclear Magnetic Resonance                                 |
| NO                      | Nitric Oxide                                               |
| OPLS                    | Orthogonal Partial Least Squares                           |

|              |                                                                  |
|--------------|------------------------------------------------------------------|
| OPLS-DA      | Orthogonal Partial Least Squares - Discriminant Analysis         |
| OSC          | Orthogonal Signal Correlation                                    |
| PAG          | Phenylacetylglutamine                                            |
| PC           | Principal Component                                              |
| PCA          | Principal Component Analysis                                     |
| PCR          | Polymerase Chain Reaction                                        |
| PDB          | Protein Data Bank                                                |
| PDGF-DD      | Platelet-Derived Growth Factor-DD                                |
| PDGFR        | Platelet-Derived Growth Factor Receptor                          |
| PECAM-1      | Platelet Endothelial Cell Adhesion Molecule-1                    |
| Phe          | Phenylalanine                                                    |
| PIP2         | Phosphatidylinositol 4,5 Bisphosphate                            |
| PI3K         | Phosphoinositide-3-Kinase                                        |
| PIGF         | Placental Growth Factor                                          |
| PLCB2        | 1-Phosphatidylinositol-4,5-Bisphosphate Phosphodiesterase Beta-2 |
| PLS          | Partial Least Squares                                            |
| PLS-DA       | Partial Least Square - Discriminant Analysis                     |
| PMSF         | Phenylmethylsulfonyl Fluoride                                    |
| PPARg        | Peroxisome Proliferator-Activated Receptor g                     |
| PRKCA        | Protein Kinase C $\alpha$                                        |
| PSI-BLAST    | Position-Specific Iterative BLAST                                |
| RCA          | Right Coronary Artery                                            |
| RTK          | Receptor Tyrosine Kinases                                        |
| SCFA         | Short-Chain Fatty Acid                                           |
| SHN2         | Schnurri-2                                                       |
| SRC          | Proto-Oncogene Tyrosine-Protein Kinase Src                       |
| STE          | STE Kinase Group                                                 |
| STK32C       | Serine/Threonine Kinase 32C                                      |
| STOCSY       | Statistical Total Correlation Spectroscopy                       |
| STORM        | Subset Optimisation by Reference Matching                        |
| TAG          | Triacylglycerol                                                  |
| TBBLASTN     | Protein-Nucleotide 6-Frame Translation BLAST                     |
| TCA          | Tricarboxylic Acid Cycle                                         |
| TG           | Triglycerides                                                    |
| TGF- $\beta$ | Transforming Growth Factor- $\beta$                              |
| TK           | Tyrosine Kinase                                                  |
| TKL          | Tyrosine Kinase-like                                             |
| TMA          | Trimethylamine                                                   |
| TMAO         | Trimethylamine- <i>N</i> -oxide                                  |
| TnaA         | Tryptophanase                                                    |
| TNF          | Tumour Necrosis Factor                                           |
| TReg         | T Regulatory                                                     |

|                   |                                                          |
|-------------------|----------------------------------------------------------|
| TSP               | Trimethylsilylpropanoic acid                             |
| Tyr               | Tyrosine                                                 |
| Val               | Valine                                                   |
| VEGF <sub>A</sub> | Vascular Endothelial Growth Factor <sub>A</sub>          |
| VEGFR-1           | Vascular Endothelial Growth Factor Receptor-1            |
| VEGFR-2           | Vascular Endothelial Growth Factor Receptor-2            |
| VEGFR-3           | Vascular Endothelial Growth Factor Receptor-3            |
| VLDL              | Very Low Density Lipoprotein                             |
| WHO               | World Health Organisation                                |
| ZAK               | Sterile Alpha Motif and Leucine Zipper-Containing Kinase |







## Chapter 1

# Introduction

Obesity is a well-established risk factor for various medical conditions, including type 2 diabetes, dyslipidaemia and arterial hypertension (Gelber, et al., 2007; Shamaï, et al., 2011; Speliotes, et al., 2010); it is also an independent risk factor for cardiovascular disease (CVD), the leading cause of mortality worldwide (WHO, 2016). Due to the close relationship between obesity and its co-morbidities, these conditions are often referred to collectively as cardiometabolic disorders.

A growing body of evidence supports that the gut microbiota play a critical role on the onset and development of cardiometabolic disorders. This complex microbial community communicates with its mammalian host *via* a range of signalling molecules, that can lately impact obesity and CVD (Delzenne and Cani, 2011; Xu and Gordon, 2003). In this chapter, the basic mechanisms and aetiology of obesity and CVD will be revisited, as well as the concept of gut microbial metabolites - in particular the indoles - as signalling molecules that, by binding human targets and activating signalling pathways, influence human health and disease.

# 1. Obesity and cardiometabolic diseases

## 1.1 Obesity

Obesity is one of the most important public health problems of the 21<sup>st</sup> century. In 2014, worldwide prevalence of obesity and overweight was, respectively, 13% and 39%, comprising over 1.9 billion individuals (WHO, 2016). Although initially considered a public health problem of high-income countries only, the prevalence of obesity has also steadily increased in low and middle-income countries, assuming pandemic proportions (WHO, 2016).

Obesity is a well-established risk factor for type 2 diabetes, dyslipidaemia, arterial hypertension, musculoskeletal disorders and cancer (Anandacoomarasamy, et al., 2008; Chen, et al., 2012; Dougan, et al., 2015; Gelber, et al., 2007; Genkinger, et al., 2011; Ma, et al., 2013; Shamai, et al., 2011; Speliotes, et al., 2010; Wang and Xu, 2014). Obesity also contributes greatly to cardiovascular-related mortality, both as an independent risk factor and due to its association with other cardiovascular risk factors (Hubert, et al., 1983).

The disability associated with obesity affects the individual's quality of life and predisposes them to increased frailty (Blaum, et al., 2005). The social and economic impact of obesity has also steadily increased in the last few decades. In Europe, obesity is now estimated to account for more than 12 million years lost due to ill-health (disability-adjusted life years (DALYs)) (Ezzati M, 2004) and to cost more than 59 billion euros per year (OECD, 2012). As a leading cause of disability, morbidity, mortality and healthcare costs, strategies to reduce obesity rates are considered a priority for health care organisations worldwide (Cuschieri and Mamo, 2016).

### 1.1.1 Clinical assessment of obesity

Obesity is defined as an excessive accumulation of fat mass, as consequence of the imbalance between energy intake and energy expenditure. In obese patients, fat mass is mostly compartmentalised in two body locations: the abdominal region (central fat mass) and the limbs (peripheral fat mass). Central fat mass, comprising visceral adipose tissue and abdominal subcutaneous fat, has a positive association with CVD (Despres, et al., 2008; Phillips and Prins, 2008). Peripheral fat mass, especially gluteofemoral fat, might be less atherogenic, due to a lower fatty acid turnover and a different hormonal production profile (Ferreira, et al., 2004; Frayn, 2002).

Several measurements have been proposed to assess obesity. The Body Mass Index (BMI), defined as the ratio between weight (in kg) and the square of height (in m<sup>2</sup>) (Keys, et al., 1972), is one of the most extensively used. According to the World Health Organisation, individuals with a BMI of 30 kg/m<sup>2</sup> or higher are classified as obese, while those with values between 25 kg/m<sup>2</sup> and 29.9 kg/m<sup>2</sup> are considered overweight (WHO, 2016). BMI measurements are simple, inexpensive and non-invasive; however, their sensitivity and predictive power have been questioned (Pories, et al., 2010; Romero-Corral, et al., 2008). It is important to note that BMI is a surrogate marker based on a mathematical estimation, not a direct measurement of the individual's fat mass (Shah and Braverman, 2012). Consequently, it incorporates the contribution of other body components such as muscle mass and bone mass. As an example, high BMI can be observed in individual with muscular hypertrophy; while patients with sarcopenic obesity (increased adiposity in the presence of decreased lean mass) may exhibit normal BMI values (Donini, et al., 2013). Some authors propose that obesity would be better described by central obesity anthropometric measures (*e.g.* waist circumference, waist-to-hip ratio and waist-to-height ratio) (Wells and Fewtrell, 2006).

Recently, more advanced techniques have allowed the direct quantification of both the individual's fat mass (FM) and the fat-free mass (FFM) (*e.g.* densitometry, dual energy X ray absorptiometry (DEXA)). These methods provide direct

quantification of FM, and do not rely on predictive models based on anthropometric measurements, thus being more accurate than BMI (Wells and Fewtrell, 2006).

### **1.1.2 Obesity, insulin resistance and metabolic syndrome**

Obesity is a major risk factor for several chronic metabolic complications, including type 2 diabetes, dyslipidaemia and arterial hypertension (Alberti, et al., 2005). Each 1-unit increase in an individual's BMI value increases by 8% the risk of incident arterial hypertension over 14.5 years (Gelber, et al., 2007). Obesity is also an independent risk factor for type 2 diabetes; it is estimated that the risk increases between 4.5 and 9% for every kilogram of weight gained (Ford, et al., 1997). Higher BMI values are associated with a deleterious lipid profile, characterised by increased levels of triglycerides and reduced levels of high-density lipoprotein (HDL) cholesterol (Shamai, et al., 2011). The epidemiological clustering of obesity with other metabolic disorders led to the introduction of the concept of "metabolic syndrome", a predictive marker for the risk of type 2 diabetes and CVD (Despres, et al., 2008; Kim and Reaven, 2004). According to the latest guidelines of the International Diabetes Federation (IDF), metabolic syndrome is diagnosed in the presence of central obesity ( $\text{BMI} \geq 30 \text{ kg/m}^2$  or, if below, increased waist circumference, defined by ethnicity specific values), plus one of the following criteria: raised plasma glucose, raised systolic blood pressure, raised total cholesterol or reduced HDL cholesterol (IDF, 2005).

Obesity plays a central role in the onset of the metabolic syndrome: the accumulation of adipose tissue triggers local inflammatory responses and the release of cytokines that desensitise insulin signalling pathways in peripheral tissues, leading to insulin resistance (Bonder, et al., 2016). Consequently, the plasma concentration of glucose increases, and pancreatic  $\beta$ -cells respond by increasing the production of insulin (hyperinsulinemia) (Benson, 2016). Eventually, the inability of the pancreatic  $\beta$ -cells to produce sufficient insulin leads to the transition from insulin resistance to type 2 diabetes (Jeffery, et al., 2012). Insulin resistance also reduces the uptake of circulating lipids and increases lipolysis, increasing the concentration of circulating lipids and ultimately promoting dyslipidaemia (Atarashi, et al., 2013; Sender, et al., 2016).

### 1.1.3 Basic mechanisms of obesity and energy homeostasis

Obesity develops as a result of energy intake in excess of energy expenditure; as a consequence, excess energy is stored in adipocytes in the form of triglycerides (TG). For most organisms, the main energy source is glucose; however, free fatty acids (FFA) and amino acids can also be used as alternative energy sources.

Dietary carbohydrates are digested into monosaccharides (mainly glucose), which are transported from the intestine to various tissues for energy production (Towle, et al., 1997). In most tissues, the major pathway for adenosine triphosphate (ATP) production is the metabolism of glucose into pyruvate (*glycolysis*) (Han, et al., 2016; Pilkis, et al., 1988). In mitochondria-rich tissues, pyruvate enters the mitochondrial matrix and is converted into acetyl-CoA, which is then incorporated in the tricarboxylic acid cycle (TCA) (Figure 1). The TCA cycle generates not only guanosine triphosphate (GTP), an energy equivalent of ATP, but also reduced nicotinamide adenine dinucleotide (NADH) and reduced flavin adenine dinucleotide (FADH<sub>2</sub>) that act as electron carriers for the process of electron transport chain-oxidative phosphorylation, thus generating ATP (Han, et al., 2016).

In case of energy surplus, excessive glucose is metabolised into Acetyl-CoA, which is then used for FFA synthesis *via* lipogenesis (Kersten, 2001; Towle, et al., 1997) (Figure 1). FFA are esterified with glycerol, producing triglycerides that are stored in adipose tissue, liver and peripheral tissues (Towle, et al., 1997). Excess glucose can also be stored in the liver in the form of glycogen, a secondary energy source (*glycogenesis*) (Pilkis, et al., 1988).

Energy storage is tightly controlled by glucoregulatory hormones, amongst which insulin is the main one (Giugliano, et al., 2008). In state of energy excess, insulin secretion increases and promotes glucose uptake by the cell, lipogenesis and glycogenesis, thus favouring energy storage (Giugliano, et al., 2008). Conversely, during fasting, there is a combined hormonal effect to increase glucose availability: insulin secretion is reduced and the production of glucagon, a hormone with opposite effects to insulin, increases (Aronoff, et al., 2004; Giugliano, et al., 2008). Glucagon

promotes the degradation of glycogen (glycogenolysis) in liver and muscle (Han, et al., 2016).

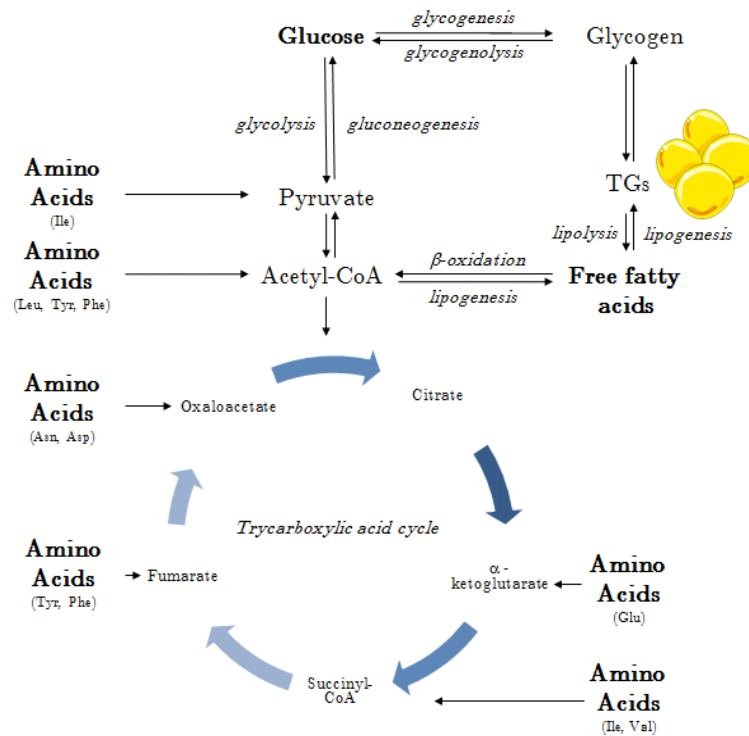


Figure 1. Overview of glucose, free fatty acid and amino acid metabolism.

The three pathways share common metabolites that converge into the tricarboxylic acid (TCA) cycle. Glucose, amino acid and free fatty acid metabolic pathways share several common metabolites that converge into the TCA cycle (Rodwell, et al., 2015). Acetyl-CoA, pyruvate and other intermediary metabolites activate the TCA cycle, producing reduced coenzymes (i.e. NADH, FADH<sub>2</sub>). These coenzymes are further re-oxidised by oxidative phosphorylation, producing ATP that is finally used as an intracellular energy source. Asn: Asparagine, Asp: Aspartate, Ile: Isoleucine, Leu: Leucine, Phe: Phenylalanine, TG: triglycerides; Tyr: Tyrosine, Val: Valine). Designed based on information from (Rodwell, et al., 2015).

In muscle, glycogenolysis provides an immediate source of glucose-6-phosphate for glycolysis, leading to the production of pyruvate (Han, et al., 2016). In hepatocytes, the main purpose of glycogenolysis is the release of glucose into the systemic circulation to be used by other cells. In this case, hepatic glucose-6-phosphatase removes the

phosphate group, and free glucose exits the cell via glucose transporter 2 (GLUT2) facilitated diffusion channels (van Schaftingen and Gerin, 2002). During fasting, glucagon also increases circulating glucose by increasing lipolysis, thus producing FFA that undergo  $\beta$ -oxidation and produce acetyl-CoA (Miles and Jensen, 1993; Oscar, 1991).

As mentioned above, to avoid low glycaemic levels amino acids can also be used as secondary energy sources. Dietary protein breakdown by digestive enzymes generates peptides that enter enterocytes *via* peptide transporters; inside the enterocytes, peptides are converted into amino acids that finally enter systemic circulation (Rodwell, et al., 2015). Following nitrogen removal, most of the carbon atoms present in amino acids skeletons are degraded into TCA cycle intermediates, pyruvate and acetyl-CoA (Jungas, et al., 1992). Glucogenic amino acids (and other non-carbohydrate carbon substrates, *e.g.* pyruvate, glycerol) can also generate glucose *via gluconeogenesis*, a metabolic pathway taking place in liver and, to a lesser extent, in renal cortex (Jungas, et al., 1992) (Figure 1).

Glucose, amino acid and free fatty acid metabolic pathways share several common metabolites that converge into the TCA cycle (Rodwell, et al., 2015). Acetyl-CoA, pyruvate and other intermediary metabolites activate the TCA cycle (Figure 1), producing reduced coenzymes (i.e. NADH, FADH<sub>2</sub>). These coenzymes are further re-oxidised by oxidative phosphorylation, producing ATP that is finally used as an intracellular energy source.

Changes in glucose, FFA or amino acid metabolism promote significant disruption to physiological energy homeostasis. Excess glucose availability induces a deregulation of insulin action, in which cells fail to respond to this hormone and do not uptake glucose efficiently (*insulin resistance*) (Muoio and Newgard, 2008). As consequence, plasma levels of glucose increase (hyperglycaemia) and the energy surplus is stored in the form of TG (Unger, et al., 2010).

#### 1.1.4 The multifactorial aetiology of obesity

Although there is growing knowledge of the metabolic pathways involved in obesity, its causes are not completely understood. Most evidence suggests obesity has a multi-factorial aetiology, comprising genetic, environmental and epigenetic factors (Bleich, et al., 2008).

To date, more than 50 genetic variants associated with obesity have been identified in genome-wide association studies (GWAS) (Herrera, et al., 2011; Leon-Mimila, et al., 2013), including genes involved in the TCA cycle, glycolysis and inflammatory signalling (Cheung, et al., 2006). The obese (*ob*) gene and its product leptin are involved in the regulation of both appetite and energy expenditure in the mouse (Zhang, et al., 1994), but the frequency of this mutation is extremely rare in human populations (“syndromic obesity”) (Pi-Sunyer, et al., 1999). *FTO*, the first gene associated with non-syndromic obesity in humans, interacts through a noncoding region with the promoter of two homeobox genes (*IRX3* and *IRX5*) which modulate obesity and body mass composition (Gerken, et al., 2007; Wahlen, et al., 2008). However, the combined contribution of all genetic variants only accounts for a small proportion of body weight variation; also, while the prevalence of obesity has steadily increased over the last three decades, genetic background has not changed greatly (Herrera, et al., 2011; Leon-Mimila, et al., 2013; Speliotes, et al., 2010). Altogether, these facts suggest that the growing prevalence of obesity is more likely to be explained by environmental changes and lifestyle factors rather than host genetics.

Sedentary habits and high-carbohydrate, high-fat diets are known to promote a disproportion between energy intake and expenditure, resulting in increased body FM (WHO, 2016). Environmental stimuli can also trigger epigenetic mechanisms (*e.g.* histone acetylation, DNA methylation) that regulate the expression of genes involved in energy homeostasis (Musri and Parrizas, 2012). Interactions between environmental cues and epigenetic mechanisms associated with increased adiposity have been described for various genes, including *MCR4*, *FTO*, *PPAR $\gamma$*  and *POMC* (Choy, et al., 2010; Gerken, et al., 2007; Plagemann, et al., 2009; Widiker, et al., 2010).

Recently, the gut microbiota - a complex microbial community consisting of trillions of non-pathogenic commensal organisms living in a symbiotic relationship with its human host - have been suggested to play a pivotal role in obesity and its related metabolic co-morbidities (Delzenne and Cani, 2011). This perspective will be further explored in *Section 2.3. Gut microbial changes in obesity and cardiovascular disease*.

## **1.2 Cardiovascular diseases**

Obesity is also an independent risk factor for CVD, a broad group of pathologies comprising vascular retinopathy, vascular nephropathy, peripheral artery disease, arterial hypertension, stroke and coronary artery disease (CAD) (Hubert, et al., 1983) (Figure 2). In 2013, CVD was the leading cause of death worldwide with a total of 17.9 million deaths. CAD accounted for 80% and 75% of deaths in males and females, respectively (Brestoff and Artis, 2013).

### **1.2.1 Clinical assessment of cardiovascular diseases**

CVD share a common mechanism: atherosclerosis (Guyton and Hall, 2005). The visualisation and characterisation of atherosclerosis in living patients is now possible by the use of several techniques. Amongst them, the echographic assessment of the intima media thickness assessment is particularly useful to detect the presence of atherosclerosis and predict the risk of acute events in high-risk patients (Al-Waiz, et al., 1987; Dolphin, et al., 1997). Coronary atherosclerosis can be further characterised by angiography (Craciun and Balskus, 2012), computed tomography (Al-Waiz, et al., 1992) or resonance magnetic imaging (Al-Waiz, et al., 1987).

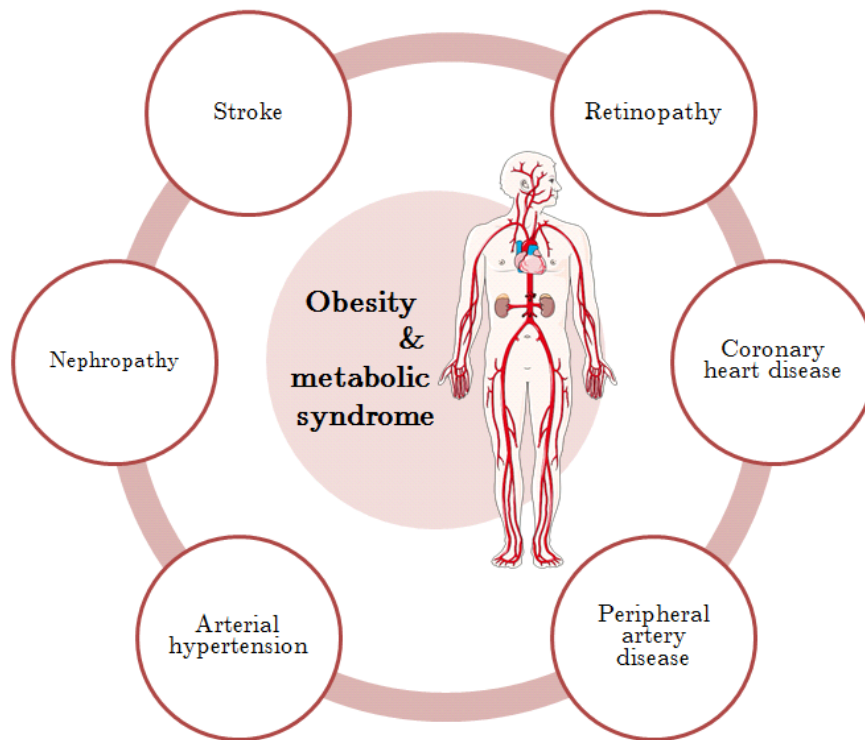


Figure 2. Cardiovascular complications associated with obesity and metabolic disorders. Both through a positive association with cardiometabolic risk factors and as an independent risk factor, obesity is associated with deleterious cardiovascular diseases. Designed using Servier Medical Art elements.

### 1.2.2 Basic mechanisms of cardiovascular diseases: the role of atherosclerosis

Atherosclerosis is a pathological process of the vascular endothelium, a monolayer of vascular endothelial cells that lines the entire circulatory system (Guyton and Hall, 2005) (Figure 3). This pathological process is the common underlying cause of several cardiovascular diseases, including stroke and coronary heart disease.

The atherosclerotic process starts with the adhesion of leukocytes to the endothelial monolayer, followed by migration into the intima layer and, finally, maturation into macrophages (Libby, et al., 2011; Lusis, 2000). Macrophages are able to uptake lipid, leading to the formation of the initial atherosclerotic lesions - the “*fatty streaks*” (Guyton and Hall, 2005). The combined effect of several mechanisms

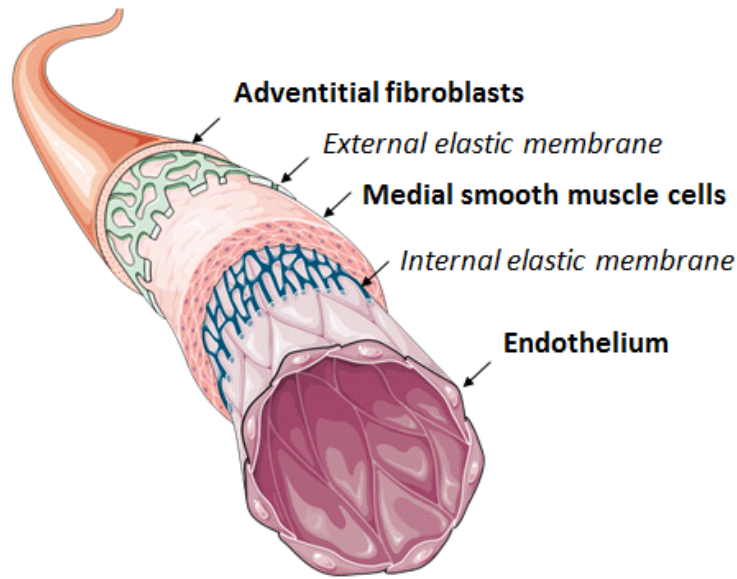


Figure 3. Trilaminar vascular structure.

Arterial and venous vessels comprise three layers, at a macroscopic level. The innermost layer (also called *intima*) includes the endothelium, a monolayer of vascular endothelial cells. The middle layer (also called *media*) contains mostly smooth muscle cells; it is separated from the other layers by the internal and external elastic membranes. The outermost layer (also called *adventitia*) consists of fibroblasts and collagen fibers. Designed using Servier Medical Art elements.

(*i.e.* migration of smooth muscle cells, increased synthesis of extracellular components, inflammation and oxidative stress) contributes to the progression of the lesion (Forstermann, et al., 2017; Hansson, 2005; Libby, et al., 2002). The last step of the natural history of atherosclerosis is thrombosis (Lusis, 2000). When the plaque ruptures, the tissue factors in its interior are exposed to the coagulation components present in the blood, triggering the formation of a thrombus that occludes the vascular lumen (Lusis, 2000). The blood flow is progressively limited by the increasing size of the atherosclerotic plaque, leading to stable angina, unstable angina and, if the blood flow is critically compromised, myocardial infarction and death (Guyton and Hall, 2005).

Atherosclerosis is closely linked to angiogenesis, the mechanism of formation of new blood vessels (Barger and Beeuwkes, 1990). However, the association is complex and likely dependent on the stage of the disease (Khurana, et al., 2005). Adventitial angiogenesis allows atherosclerotic plaques to develop beyond a critical thickness point,

by supplying them with oxygen and nutrients (Barger and Beeuwkes, 1990). In line with these results, associations between neovascularisation and the extent of the atherosclerotic plaque have been found in both primate and human studies (Kamat, et al., 1987; Williams, et al., 1988; Zhang, et al., 1993). In later stages of the disease, an increased neovascular network renders plaques more fragile and prone to rupture, leading to plaque destabilisation, luminal occlusion and, finally, acute coronary syndromes (Barger and Beeuwkes, 1990). A study involving 269 patients with advanced atherosclerotic lesions found a strong correlation between microvessel formation and plaque rupture (Moreno, et al., 2004). Moulton *et al* (1999) demonstrated that angiogenesis inhibitors (endostatin and TNP-470) are able to reduce the area of the plaque by 85% and 70%, respectively. Nevertheless, the authors cautiously point out that this effect may be due to non-specific effects of the compounds used (Moulton, et al., 1999). Interestingly, preclinical studies suggest that, after myocardial infarction, angiogenesis could actually have a beneficial effect by enhancing post-ischemic restorative revascularisation. (Aicher, et al., 2005; Melo, et al., 2004). However, clinical trials have failed to demonstrate this beneficial effect and most of the current evidence globally supports that angiogenesis has a pro-atherogenic effect (Celletti, et al., 2001; Heeschen, et al., 2001).

It is important to note that obesity also contributes to several basic mechanisms of atherosclerosis, thus explaining the epidemiological clustering of both conditions. First, by inducing profound hemodynamic changes, obesity indirectly increases metabolic demand, cardiac workload and ventricular pressure and increased blood pressure (Collis, et al., 2001; Ku, et al., 1994). The latter is associated with shear stress and oxidative stress, which promote endothelial dysfunction and contribute to the progression of the atherosclerotic plaque (Dharmashankar and Widlansky, 2010). Adipose tissue is not an inert energy storage tissue; conversely, it has an active role as the largest endocrine organ in humans, synthesising a range of products that impact the development and progression of CAD (Coelho, et al., 2013). Obese patients exhibit increased levels of pro-inflammatory cytokines (tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) and interleukin-6 (IL-6)); IL-6 also increases the hepatic production of C-reactive protein (CRP), which enhances the uptake of low-density lipoprotein (LDL) cholesterol by macrophages in the atherosclerotic plaque (Mehta, et al., 2007). Finally, obesity creates

a pro-angiogenic environment (Miyazawa-Hoshimoto, et al., 2003). The expansion of the FM requires a simultaneous increase of the supporting vascular network, to provide an adequate supply of oxygen and nutrients (Cao, 2007). Therefore, adipose tissue expresses several growth factors that promote the formation of new blood vessels (angiogenesis), including fibroblast growth factor (FGF), placental growth factor (PIGF) and vascular endothelial growth factor (VEGF), the main regulator of angiogenesis and vascular endothelial cell function (Cao, 2007). Consistently, in human studies, BMI and visceral fat area measurements are positively correlated with VEGF concentrations (Miyazawa-Hoshimoto, et al., 2003), thus creating a pro-angiogenic – and consequently, pro-atherogenic – drive in these patients.

### **1.2.3 The multifactorial aetiology of cardiovascular diseases**

Most evidence suggests that atherosclerosis has a multi-factorial aetiology, comprising both environmental and genetic factors. As previously mentioned, obesity is an independent risk factor for cardiovascular mortality (Engeland, et al., 2003); interestingly, children with higher BMI exhibit an increased risk of CVD in adulthood, highlighting the impact of early life programming on long-term cardiovascular outcomes (Baker, et al., 2007). Diabetes is also associated with a 2- to 3-fold increase in the likelihood of developing CVD (Fox, et al., 2004). Similarly, both systolic and diastolic blood pressure have a continuous and positive association with CVD outcomes (Stamler, et al., 1993). In individual aged between 50 and 70 years, an increment of 20 mm Hg in systolic blood pressure was shown to double the risk of CVD (Lewington, et al., 2002). There is also a strong relationship between serum total cholesterol and cardiovascular risk (Stamler, et al., 1993). While, LDL cholesterol, the principal lipoprotein transporting cholesterol in the blood, is positively associated with CVD (Kannel, et al., 1979), HDL cholesterol levels are inversely correlated with coronary risk (Gordon, et al., 1989).

Several environmental factors have also been implicated in CVD outcomes. A high consumption of processed meats increases CVD-related mortality, possibly due to increased salt intake (Gauch, 2003). Conversely, decreasing the consumption of saturated fat reduces this risk by 17% (Tzoulaki, et al., 2014). Moderate physical

activity (150 minutes of moderate physical activity each week) was shown to reduce the risk of CAD (Engin, 2017). Smokers have an increased risk of myocardial infarction or sudden death; importantly, the risk is proportional to the number of cigarettes smoked daily (Doyle, et al., 1962). Exposure to air pollutants, especially particulate matters, is also linked to high values of *intima media* thickness and an increased risk acute CVD events (Eckers, et al., 2016).

To date, the genetic contribution to atherosclerosis seems to be mostly related to genes associated with lipid metabolism. *APOE* polymorphisms have been linked to atherosclerosis and hypercholesterolemia (Borrel, et al., 2017). *APOE* deficiency in mice results in a profound susceptibility to atherosclerosis; however, the same effect is not observed in humans (Borrel, et al., 2017). Common variants of *LIPC*, *APOC3/A4/A5* and *Apo(a)* loci influence, respectively, the levels of HDL, triglycerides and lipoprotein(a) levels (Cohen, et al., 1999; Kinexus, 2017; Lustig, 2003). Both *LIPC*, and *APOC3/A4/A5* loci seem to influence plasma lipoprotein(a) size, a known risk marker for CAD (Lustig, 2003). However, the estimated heritability of CAD is only between 40% and 60% (Lusis, 2000). As for all complex diseases, the phenotype of atherosclerosis is likely a product of many genes with small effects, modified by the environment and epigenetic effects.

Recently, the gut microbiota was also proposed to contribute to the pathogenesis of CVD (Jonsson and Backhed, 2017). This perspective will be further explored in *Section 2.3. Gut microbial changes in obesity and cardiovascular disease*.

## 2. The gut microbiota

A growing body of evidence supports that the gut microbiota, a complex microbial community inhabiting the human gut, plays a pivotal role in the development of obesity, metabolic diseases and CVD (Delzenne and Cani, 2011; Xu and Gordon, 2003).

This section describes contributors to the gut microbiota, as well as changes observed in the composition in the context of obesity and CVD. The contribution of the healthy gut microbiota in relation to host barrier function, immunomodulation, metabolism and signalling is also discussed. Basic concepts of cell signal sensing and signal transduction are revisited, and exemplified with several cases of gut microbial metabolites that act as ligands for G-coupled protein receptors (GPCR) and nuclear receptors. Finally, the human kinome (*i.e.* the complete set of protein kinases encoded by the human genome) is introduced as an unexplored target for gut microbial metabolites.

### 2.1 Composition of the gut microbiota

For millions of years, humans have co-evolved in a symbiotic relationship with their microbial communities – termed the microbiota. Although many anatomical regions (*e.g.* skin, vaginal mucosa, respiratory system, oral cavity) harbour specific microbial populations, the digestive tract is by far the most populated system of the human body, home to the gut microbiota (Kim, et al., 2013).

The gut microbiota consist of approximately 100 trillion of microbes, a number of the same order than the number of human cells (Sender, et al., 2016). The gut microbiome is the collective designation for the more than 10 million microbial genes coded by the gut microbiota (Backhed, et al., 2005; Li, et al., 2014). At the kingdom level, the gut microbiota are predominately bacteria; however, archaea, fungi and viruses are also present (Lozupone, et al., 2012). At the phylum level, more than 50 different phyla of bacteria have been identified; amongst those, only four are dominant:

Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria (Arumugam, et al., 2011). The relative proportion between phyla (in particular, the Firmicutes:Bacteroidetes ratio) was initially suggested to be especially relevant for human health; however, it was demonstrated that the gut Bacteroidetes/Firmicutes balance may not be a significant attribute of an individual’s gut microbiome; conversely, individuals carry a ‘personal equilibrium’ among Bacteroidetes and a group of temporally variable Firmicutes (Lloyd-Price, et al., 2017) (*Further details in Section 2.3. Gut microbial changes in obesity and cardiovascular disease*).

At the genus level, cluster analysis revealed the existence of three distinct bacterial profiles, designated enterotypes (Arumugam, et al., 2011). The identification of each enterotype is based on the predominance of one of three main genera: *Bacteroides* (phylum Bacteroidetes), *Prevotella* (phylum Bacteroidetes) or *Ruminococcus* (phylum Firmicutes) (Arumugam, et al., 2011; Knights, et al., 2014).

Although the classification into enterotypes provides an appealing framework to describe gut bacterial variation, the collapse of the continuous metagenomic bacterial variability into three discrete clusters remains a matter of debate (Jeffery, et al., 2012; Knights, et al., 2014). As the abundance of dominant taxa is a continuous gradient, predictive models based on the taxon-relative abundances have been suggested to be more effective as disease biomarkers than enterotype-based models (Knights, et al., 2014). Also, enterotype-based models should be used carefully in cases in which specific bacteria (or bacterial genes) are associated with health or disease, as they may mask within-cluster variation of individual strains. In these cases, the use of single-effect models is likely to be more powerful as disease biomarkers (Knights, et al., 2014).

The gut microbiota have also been profiled by quantitative metagenomics regarding their microbial genes, which seems to be an indicator for human health (*Further details in Section 2.3. Gut microbial changes in obesity and cardiovascular disease*). Several proxies and diversity indexes have been used, including the total bacterial gene count, microbial gene abundance (*i.e.* the number of different types of genes in a given sample, rarefaction at 6 or 7 M genes), the Chao1 index (based upon the number of rare classes found in a sample) and the Shannon index (that accounts

for both abundance and evenness of the species present (Le Chatelier, et al., 2013; Santacruz, et al., 2010) (Tuomisto, 2010).

Finally, the composition of the gut microbiota varies across different segments and niches of the digestive tract (*latitudinal variation*). While the stomach and small intestine are sparsely populated, the luminal contents of the colon harbour a complex and densely populated bacterial ecosystem with up to  $10^{12}$  cells/gram of intestinal content (Gill, et al., 2006). Intestinal pH and oxygen gradients influence the spatial distribution of the microbiota. Colon pH is mildly acidic in the proximal colon and more neutral distally; several *Firmicutes* species are more tolerant to acidic pH, thus exhibiting a competitive advantage in these conditions (Flint, et al., 2012). The colonic lumen is highly anaerobic and most colonic bacteria are strict anaerobes, mostly due to the consumption of available oxygen by facultative anaerobes (Flint, et al., 2012). In the colon, bacterial composition also varies between the mucosa and the crypts (*longitudinal variation*) (Swidsinski, et al., 2005) (Figure 4). The spatial relationship between the gut microbiota and the colonic epithelium may explain the differential responses (*e.g.* inflammation) promoted by specific bacterial groups (Swidsinski, et al., 2005).

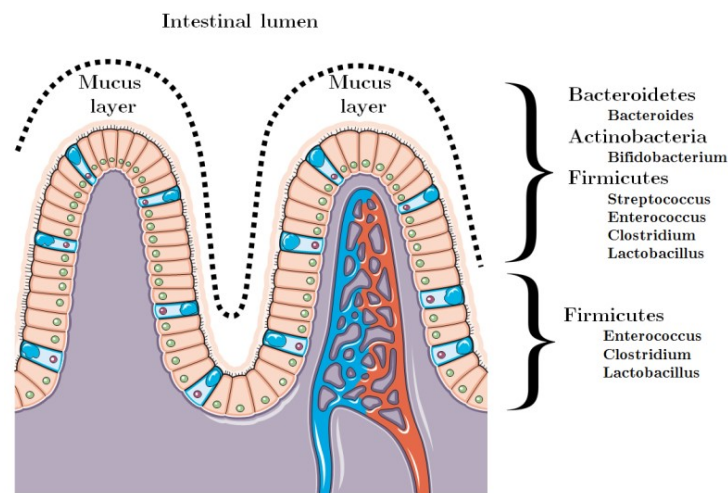


Figure 4. Longitudinal variation of the microbial composition in human distal gut. Based on information taken from (Swidsinski, et al., 2005) and designed using Servier Medical Art graphic elements.

## 2.2 Host and environmental factors affecting gut microbial ecology

The composition of the gut microbiota is a complex and polygenic trait, influenced by both intrinsic and extrinsic factors (*e.g.* diet, lifestyle) (Benson, et al., 2010; Goodrich, et al., 2014). Furthermore, it experiences changes as a function of the host's age (O' Toole and Jeffery, 2015). In this section, the impact of these factors across the life course will be described.

### *Genetic factors*

Initial evidence suggesting a host genetic effect comes from studies showing that family members often share similar gut microbiota patterns (Tims, et al., 2013; Turnbaugh, et al., 2009). Although these similarities could be an effect of a shared environment, the effect of host genetic factors was reinforced in twin studies showing that microbiomes of monozygotic twins share more similarities than the ones of dizygotic twins (Benson, 2016; Bonder, et al., 2016; Goodrich, et al., 2016; Goodrich, et al., 2014). Specific gene-microbiota interactions were also identified: *Nod2*, for instance, was shown to be essential for the development of the gut microbiota (Rehman, et al., 2011); other genes were specifically associated with the abundance of certain genera, such as *FUT2* and Bifidobacteria (Wacklin, et al., 2011).

### *Early life*

The effect of environmental factors in the gut microbial ecology starts *in utero*. The identification of bacteria in the placenta and amniotic fluid of healthy newborns has challenged the widely accepted dogma that the placenta keeps fetuses sterile during pregnancy (Rodriguez, et al., 2015). Bacterial transmission during pregnancy was unequivocally demonstrated by the oral administration of a genetically labelled *Enterococcus* species to pregnant mice, followed by its identification in the amniotic fluid and in the first stool of the offspring, called meconium (Jimenez, et al., 2005). Human meconium also harbours a microbial ecosystem (Jimenez, et al., 2008), mostly dominated by Proteobacteria and Actinobacteria (Eckburg, et al., 2005). During the

first years of life the diversity of the gut microbiota increases, due to the emergence and subsequent dominance of Firmicutes and Bacteroidetes (Eckburg, et al., 2005). Several environmental factors are known to influence the development of the gut microbiota during childhood, including the mode of delivery (*i.e.* eutocic delivery *versus* caesarean section) (Jakobsson, et al., 2014; Salminen, et al., 2004) and breastfeeding (Fallani, et al., 2010). By 2-5 years of age, the ecology of the gut microbiota is fully developed, resembling that observed in adults both in terms of composition and diversity (Koenig, et al., 2011).

### *Adult life*

Although the gut microbial ecosystem is relatively stable during adulthood, diet changes its composition (Faith, et al., 2013; Rodriguez, et al., 2015). Antibiotic treatment, even for short periods, can destabilise microbial ecology and affect metabolic programming, impacting long-term health (Cox, et al., 2014). Long-term dietary interventions are associated with the frequency of certain enterotypes: for example, carbohydrate-rich diets are associated with the *Prevotella* enterotype, whilst diets rich in animal protein and fat are linked to a higher prevalence of *Bacteroides* (Wu, et al., 2011). Dietary supplements (*e.g.* sweeteners, prebiotics) and drugs (*e.g.* metformin) are also able to modify the gut microbiota, modulating both its ecology and functions (Forslund, et al., 2015; Suez, et al., 2014).

### *Advanced age*

The ageing process also results in gradual changes in the gut microbiota over time (O' Toole and Jeffery, 2015). Subjects aged 65 years or older show different phyla proportions, with increased contribution of *Bacteroides* spp. (Claesson, et al., 2011) and reduced abundance of *Prevotella* (Jeffery, et al., 2016). It is important to note, however, that microbiota of elderly subjects exhibit extreme variability that can be explained by external factors (*e.g.* diet, exercise, medication and cohabitation patterns) (Claesson, et al., 2012). Frail elderly subjects exhibit a loss of diversity in the core microbiota groups and increased abundance of members of the order *Clostridiales* (Biagi, et al., 2010; Jeffery, et al., 2016). Further studies will shed light on whether

these changes are a consequence of the natural decline of general health, or if they are associated with specific co-morbidities in elderly subjects (Rodriguez, et al., 2015).

## 2.3 Gut microbial changes in obesity and cardiovascular disease

Over the last decades, studies have highlighted the critical role of the gut microbiota in the onset and development of obesity and CVD (Delzenne and Cani, 2011; Jonsson and Backhed, 2017; Xu and Gordon, 2003).

### *Gut microbial changes in obesity*

One of the first studies in the field demonstrated that transplanting gnotobiotic mice with the gut microbiota from an obese subject induced the development of the obese phenotype (Turnbaugh, et al., 2006). Similarly, transferring the gut microbiota of mice that underwent bariatric surgery into gnotobiotic mice also decreased body weight and fat mass (Liou, et al., 2013).

Genetically obese mice (*ob/ob*) show an increase in Firmicutes and a decrease in Bacteroidetes, a phenotype that seems to be associated with more efficient energy extraction from dietary compounds (Turnbaugh, et al., 2006). These results were further replicated in various mouse models of obesity, most of them reporting an association between obesity, increased levels of Firmicutes and decreased levels of Bacteroidetes (Hildebrandt, et al., 2009; Murphy, et al., 2010; Turnbaugh, et al., 2008). In humans, obesity was also associated with a lower Bacteroidetes/Firmicutes ratio (Ley, et al., 2006). Cani *et al* (2007) demonstrated that a high-fat diet also profoundly affects the gut microbiota, decreasing the abundance of *Bacteroides*-like mouse intestinal bacteria, the [*Eubacterium*] *rectale-Blautia coccoides* group and *Bifidobacterium* spp. (Cani, et al., 2007). The less abundant phylum Proteobacteria is also increased in overweight patients (Santacruz, et al., 2010).

A critical aspect of the microbial changes observed in obesity is an increase on the permeability of the gut barrier. Various mouse models of obesity are characterised by increased gut permeability and changes in the expression of tight-junction proteins

(*e.g.* claudin, occludin and zonula occludens 1) (Brun, et al., 2007; Cani, et al., 2008; Cani, et al., 2009; de La Serre, et al., 2010; Turner, 2009). As a consequence of increased intestinal permeability, obese patients exhibit a 2- to 5-fold increase in plasma levels of the microbiota-derived lipopolysaccharide (LPS), a component of the bacterial wall of Gram-negative bacteria (such as *Escherichia*, *Klebsiella*, *Salmonella*, *Serratia* spp.) with pro-inflammatory properties (Cani, et al., 2007; Cani, et al., 2007). The increase in the levels of LPS (which contributes to metabolic endotoxemia) dysregulates the inflammatory status of the host and triggers both weight gain and type 2 diabetes (Cani, et al., 2007). The association between high-fat diet, LPS levels, obesity and type 2 diabetes has been further validated in several human studies (Amar, et al., 2008; Erridge, et al., 2007) (Pussinen, et al., 2011) and functional and compositional changes were also observed in the gut metagenomes of type 2 diabetes women, including enrichment of Clostridiales (Karlsson, et al., 2013).

Functional analyses revealed that the microbiome from lean and obese twins express differences on the transcription of more than 300 enzymes (Ridaura, et al., 2013). These data highlight that the relationship between the gut microbiota and metabolic parameters is complex and incorporates not only the impact of different taxa, but also different metabolic functions and resulting metabolites (Cani, 2014). Addressing this topic, some studies examined the impact of gene richness (*i.e.* the number of genes in the gut microbiome), a proxy marker for the diversity of microbial metabolic functions. Le Chatelier et al (2013) demonstrated that lower bacterial gene abundance (*i.e.* gene count) was associated with increased adiposity and deleterious metabolic outcomes (*e.g.* insulin resistance and inflammation) (Le Chatelier, et al., 2013). Interestingly, dietary intervention was partially able to positively affect microbial gene richness (Cotillard, et al., 2013).

Individuals with high gene count have a higher abundance of microbial anti-oxidative stress genes and a higher genomic potential to produce short-chain fatty acids, which have been associated with metabolic beneficial outcomes (Le Chatelier, et al., 2013). Conversely, individuals with increased adiposity and low gene count exhibit a higher genomic potential for  $\beta$ -glucuronide degradation and metabolism of aromatic amino acids (*e.g.* tryptophan) (Le Chatelier, et al., 2013). Upon bacterial metabolism,

tryptophan is degraded into a range of compounds designated indoles (*Further details in the production of indoles and their impact in cardiometabolic health are provided in Section 2.5. The microbial mammalian signalling axis and Section 3. Indoles as a key family of gut microbial metabolites*).

#### *Gut microbial changes in cardiovascular disease*

Both local and distant gastrointestinal infections can promote a pro-inflammatory response, thus aggravating plaque development. There is a well-established association between periodontal disease and CVD (Trevisan and Dorn, 2010). Oral or intravenous administration of oral bacteria (*i.e. Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans*) promotes an increase in the size of atherosclerotic lesions in mouse models (Hayashi, et al., 2011).

Patients with symptomatic atherosclerosis also harbour changes in their gut microbiome (Karlsson, et al., 2012). Using shotgun sequencing, Karlsson et al (2012) demonstrated that patients with stenotic carotid plaques have an enrichment of the genus *Collinsella*, while health controls were enriched in *Roseburia* and *Eubacterium* (Karlsson, et al., 2012). Further functional analyses revealed that patients with atherosclerosis had higher levels of genes encoding peptidoglycan and a depletion of phytoene dehydrogenase (Karlsson, et al., 2012). Patients with CAD exhibit an increased abundance of *Bacteroides*, *Clostridium*, and *Lactobacillales* (Emoto, et al., 2017; Emoto, et al., 2016)

Finally, the gut microbiota can influence atherosclerosis *via* the production of specific compounds that exhibit either beneficial (e.g. short-chain fatty acids (SCFA)) or harmful effects (e.g. trimethylamine-*N*-oxide (TMAO)). The mechanisms through which SCFA benefit cardiometabolic health are unclear but might involve systemic anti-inflammatory effects. TMAO promotes vascular inflammation, enhances platelet activation and thrombosis risk and its levels are an independent predictor of the presence of diffuse atherosclerotic lesions (Seldin, et al., 2016; Senthong, et al., 2016; Zhu, et al., 2016). The effects of TMAO are also thought to be, as least partially,

mediated by an inhibition of reverse cholesterol transport (Koeth, et al., 2014; Warrier, et al., 2015).

## 2.4 Functions of the gut microbiota in health

To better understand the impact of gut microbial changes in obesity and CVD, it is important to revisit the basic functions of the ‘healthy’ gut microbiota. Having evolved in a symbiotic relationship with its mammalian host, this microbial community performs a range of functions with wide-ranging effects on human physiology (Figure 5).

### *Barrier function*

The gut microbiota play a developmental role in early life, stimulating the formation of intestinal microvilli (Lloyd-Price, et al., 2017). During life, the gut microbiota are involved in the maintenance of the gut barrier, a layer of epithelial intestinal cells connected by junctional complexes (tight junctions, adherens junctions and desmosomes) (Groschwitz and Hogan, 2009) (Figure 5). While *Bacteroides thetaiotaomicron* induces the expression of proteins involved in the maintenance of desmosomes (Lutgendorff, et al., 2008), components of the bacterial cell wall were shown to contribute to the maintenance of tight junctions (Cario, et al., 2007). Unsurprisingly, germ-free mice exhibit marked changes in gut barrier structure, including reduced intestinal surface area and lower epithelial regeneration (Banasaz, et al., 2002; Gordon and Bruckner-Kardoss, 1961).

### *Immunomodulation*

Immunomodulation is another important role of the gut microbiota (Figure 5). Gnotobiotic mice show an impaired development of immune organs (e.g. spleen and lymph nodes) (Bauer, et al., 1963) and lower levels of serum immunoglobulins (IgG1 and IgG2) (Benveniste, et al., 1971). In fact, the gut microbiota are involved in the development of intestinal and systemic immune systems throughout the host’s life,

including the maturation of precursor lymphoid cells and IgA plasmocytes and the production of circulating cytokines (Brestoff and Artis, 2013; Franchi, et al., 2012; Geuking, et al., 2011; Rivollier, et al., 2012; Suzuki, et al., 2010). Gut microbiota, in particular *Clostridia* strains, promote the accumulation of T regulatory cells (TReg) (Atarashi, et al., 2013; Atarashi, et al., 2011).

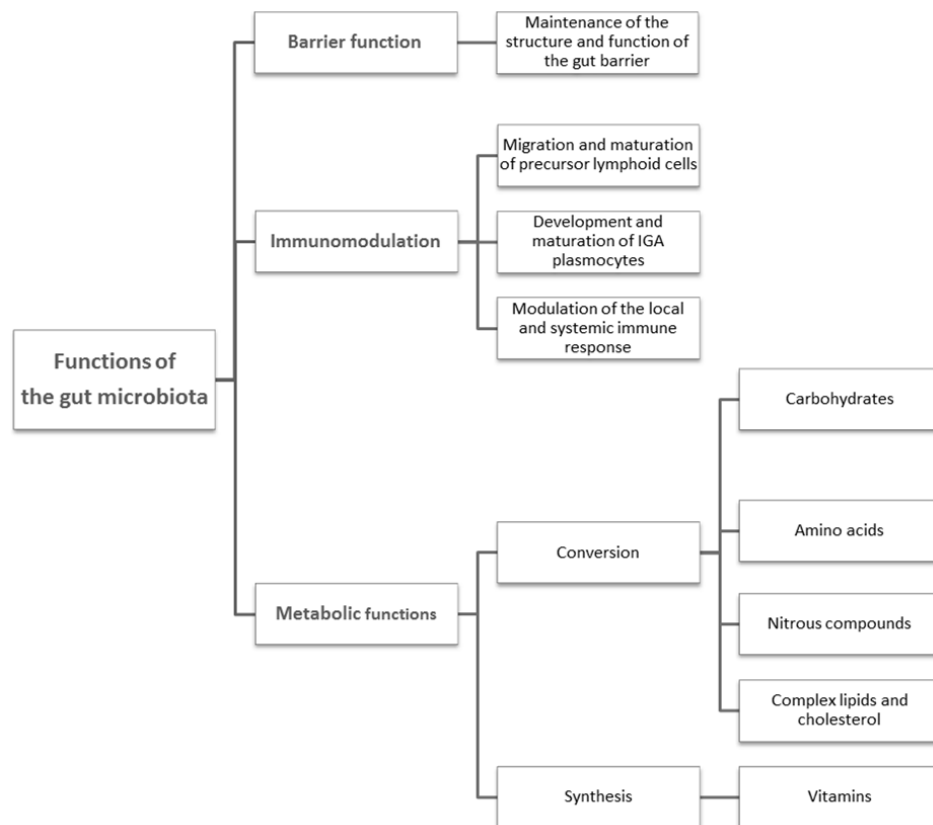


Figure 5. Functions of the healthy gut microbiota. Amongst the known functions of the gut microbiota, three aspects are particularly important: gut barrier homeostasis, immunomodulation and metabolic functions.

### *Metabolism of dietary compounds*

The gut microbiota is also involved in the metabolism of various dietary compounds (Figure 5). Amongst these, products of the metabolism of carbohydrates, amino acids, bile acids and choline seem to be particularly relevant for obesity and cardiometabolic health (Chilloux, et al., 2016). Therefore, its production will be described in further detail.

The saccharolytic fermentation of indigestible fibre by the gut microbiota produces a range of SCFA (*e.g.* formate, acetate, propionate, butyrate, isobutyrate, valerate, isovalerate). The majority of the butyrate fraction is produced by *Faecalibacterium prausnitzii*, *Eubacterium rectale* and *Eubacterium hallii* (Seldin, et al., 2016). Propionate can be produced by a range of bacteria, including *Negativicutes*, *Verrucomicrobia*, *Bacteroidetes* and *Lachnospiraceae* (Hoyles, et al., 2017).

Likewise, the gut microbiota is also able to metabolise bile acids: primary bile acids (cholic acid and chenodeoxycholic acid) synthesised by the liver enter the enterohepatic circulation where they are deconjugated, hydroxylated, oxidized and epimerized by the gut bacteria (*e.g.* *Lactobacillus* spp.) to form secondary bile acids (Hoyles, et al., 2017). Similarly, gut microbial species also degrade macronutrients; in particular, bacteria carrying operons encoding cutC and cutD are able to degrade choline into trimethylamine (TMA), which is further detoxified into TMAO by hepatic flavin-monooxygenase (FMO3) (Al-Waiz, et al., 1987; Al-Waiz, et al., 1992; Craciun and Balskus, 2012; Dolphin, et al., 1997). Gut archaea also play a role in choline metabolism, as genes for utilising TMA were found in Methanomassiliicoccales (Borrel, et al., 2017).

Finally, the gut microbiota also contributes to the degradation of amino acids. Tyrosine and phenylalanine can be degraded into 4-cresol, *via* 4-hydroxyphenylacetate, by *Clostridioides* (formerly *Clostridium*) *difficile* and certain *Lactobacillus* strains (Santacruz, et al., 2010; Senthong, et al., 2016). Finally, tryptophan can be converted into indoles by tryptophanase-containing bacteria (Sasaki-Imamura, et al., 2011), leading to the production of indoles.

Gut microbial metabolites can be used as energetic and nutritional sources or, importantly, act as signalling molecules in the microbial mammalian crosstalk. This signalling function is particularly important in the context of obesity and cardiometabolic diseases and will be explored in further detail in the next section.

## 2.5 The microbial-mammalian signalling axis

### 2.5.1 Gut microbial metabolites as signalling molecules

The concept of the “*microbial-mammalian signalling axis*” proposes that the gut microbiota communicate with its human host *via* microbial metabolites, *i.e.* signalling molecules that bind human targets and initiate cellular responses (Figure 6).

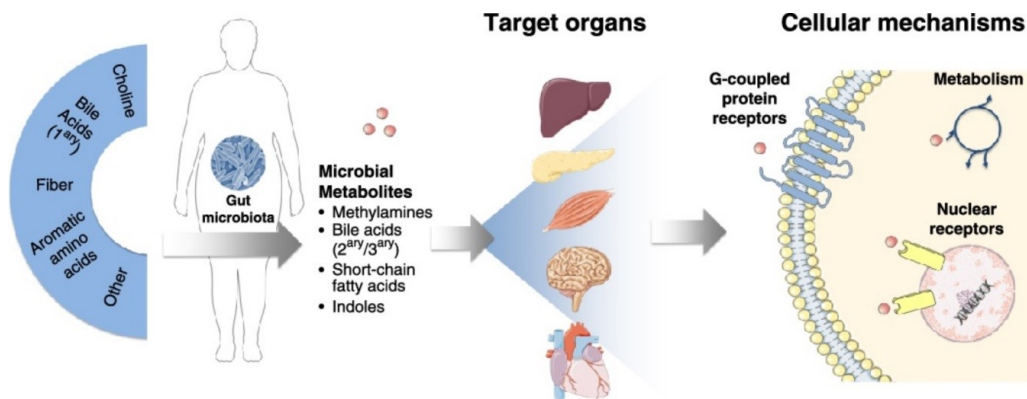


Figure 6. The microbial-mammalian signalling axis. (Chilloux, et al., 2016)

(Reprinted with permission, Licence Agreement in Appendix)

### 2.5.2 Types of receptors involved in cell signalling

Cell signalling is a key biological process through which extracellular signals are converted into cellular responses. This basic mechanism controls cell growth, proliferation and, importantly, cell metabolism (Gerhard, 2008). Cell signalling involves three steps: 1) a receptor molecule that perceives the extracellular stimulus, 2) a transducer and 3) an effector (Tuteja, 2009).

Receptors can be classified in four types, according to their mechanisms: ligand-gated ion channels, (GPCR, nuclear receptors and kinase-linked receptors).

Ligand-gated ion channels belong to a family of transmembrane ion channel proteins that incorporate a ligand-binding site (Valbuena and Lerma, 2016). Upon

ligand binding, the channel opens and ions (*e.g.*  $\text{Na}^+$ ,  $\text{Ca}^{2+}$  and/or  $\text{Cl}^-$ ) flow across the cell membrane inducing changes in the electrical properties of the cell within milliseconds (Valbuena and Lerma, 2016). GPCR are transmembrane proteins coupled to guanine nucleotide-binding proteins (G-proteins).

G-proteins are heterotrimeric complexes consisting of  $\text{G}\alpha$ ,  $\text{G}\beta$  and  $\text{G}\gamma$  subunits (Qin, et al., 2011). When a GPCR recognises a given ligand, the receptor undergoes conformational changes;  $\text{G}\alpha$  binds a molecule of GTP and separates from the remaining subunits, exposing sites on the free  $\text{G}\beta$ - $\text{G}\gamma$  complex for interaction with other molecules and activating the G-protein (Qin, et al., 2011). The activated G-protein can interact directly with an ion channel or activate/inhibit an enzyme that promotes the release of second messengers (*e.g.* cyclic adenosine monophosphate (cAMP), cyclic guanosine monophosphate (cGMP), inositol triphosphate (IP3), diacylglycerol (DAG)), that link to a number of different effectors (*e.g.* phospholipase C and D, adenylate cyclase (Valbuena and Lerma, 2016; Yang and Xia, 2006)).

Nuclear receptors are intracellular receptors consisting of a binding-domain and a DNA-binding domain. Upon ligand binding, the receptor undergoes conformational changes and enters the nucleus, where it interacts directly with DNA and alters gene expression, ultimately affecting protein synthesis (in hours).

Finally, kinase-linked receptors comprise a ligand-binding extracellular domain coupled to intracellular kinase domain (Cohen, 2002). Upon ligand binding, the kinase domain catalyses the transfer of phosphate moiety from high-energy donors (*e.g.* ATP) to target molecules (Figure 7). Kinases catalyse the reaction by orienting both the substrate and the phosphoryl group within their active site, thus increasing the rate of the reaction. Moreover, they commonly use positively charged amino acid residues (*e.g.* serine, threonine), which further stabilise the transition by interacting with the phosphate groups (Cohen, 2002). Subsequently, adapter molecules bind to the phosphorylated residues and initiate downstream signalling cascades of cytoplasmic molecules (Hubbard, 2004).

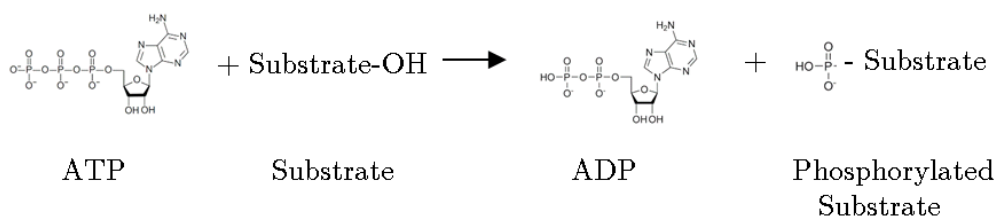


Figure 7. General reaction catalysed by kinases.

The major signalling pathways are the *cAMP-dependent pathway*, the *IP3/DAG pathway* and the *MAPK/ERK pathway* (Gerhard, 2008). In humans, the *cAMP-dependent pathway* works by activating protein kinase A; cellular responses are cell-specific, depending on the cAMP-dependent protein kinase. The *IP3/DAG pathway* starts with the cleavage of phosphatidylinositol 4,5 bisphosphate (PIP2) into DAG and IP3. IP3 diffuses through the cytosol and binds calcium channels, increasing the cytosolic concentration of calcium and promoting a cascade of changes in intracellular activity. Calcium and DAG activate the protein kinase C (AKT). Finally, the *MAPK/ERK pathway* usually links cell surface receptors to intracellular responses leading to the binding of growth factors through a range of complex protein components and interactions (Orton, et al., 2005).

These signalling cascades often interact with each other, thus allowing coordinated cellular responses (Papin, et al., 2005). At the molecular level, these responses involve changes in gene transcription or translation, as well as post-translational protein changes. Such responses are the basic mechanism controlling cell growth, proliferation, metabolism and many other processes (Gerhard, 2008; Hubbard, 2004). To date, most research has examined the role of gut microbial metabolites as ligands for GPCR and nuclear receptors.

### *Gut microbial metabolites as GPCR ligands*

SCFA act as ligands for several GPCR receptors, including Free Fatty Acid Receptor 2 (FFAR2) and Free Fatty Acid Receptor 3 (FFAR3); the activation of these receptors inhibits cAMP production, respectively, *via* G $\alpha$ i and G $\alpha$ i/q (Bindels, et al., 2013). FFAR2 activation suppresses the accumulation of adipose tissue and decreases insulin sensitivity in peripheral tissues, while increasing it in liver and muscle (Kimura, et al., 2013). Conversely, FFAR3 activation increases leptin secretion, a hormone with anorexigenic properties (Xiong, et al., 2004). Butyrate, a SCFA, is also able to bind the receptor GPR109A, initiating anti-inflammatory responses (reduced production of TNF- $\alpha$  and monocyte chemoattractant protein-1 (MCP1)) (Digby, et al., 2012; Thangaraju, et al., 2009).

### *Gut microbial metabolites as nuclear receptor ligands*

Gut microbial metabolites were also shown to bind nuclear receptors. Bile acids bind the nuclear receptor Farnesoid X Receptor (FXR), which is involved in both glucose and lipid homeostasis (*e.g.* gluconeogenesis, HDL-cholesterol, very low density lipoproteins (VLDL) and TG) (Thomas, et al., 2008). Bile acids metabolised by the gut microbiota, such as lithocholic acid, are able to activate the Pregnane X Receptor (PXR) that protects against liver toxicity (Tuomisto, 2010). Altogether, these examples demonstrate how the identification of the pharmacological targets can provide a deeper understanding of the mechanistic basis underlying the association between the gut microbiota and cardiometabolic outcomes.

## **2.5.3 The human kinome as an unexplored target repertoire**

Over the past two decades, kinases have become one of the most intensively studied protein classes in the target and drug discovery, with 46 drugs approved by the U.S. Food and Drug Administration (FDA) (Flint, et al., 2012; Goodrich, et al., 2016). Consisting of 518 kinases, the human kinome represents a broad repertoire of signal transduction triggers for human homeostasis and metabolism (Cohen, 2002; Manning, et al., 2002). According to PhosphoNET, the world's largest repository of

known and predicted information on human phosphorylation sites, the human kinome has 144,424 experimentally-confirmed phosphosites, amongst which 56.0% are serine phosphosites (Kinexus, 2017). However, in the context of gut microbial targets, this class of receptors remains mostly unexplored.

## **2.6 Indoles as a key family of gut microbial metabolites**

Indoles are gut microbial metabolites produced from dietary tryptophan and consistently reported in human biofluids. In this section, we address in further detail their bacterial production, as well as the gaps in knowledge on their human targets and relevance for obesity and cardiometabolic disorders, therefore providing the rationale for this study.

### **2.6.1 Gut bacterial production and transport of indoles**

Indoles are heterocyclic compounds produced from dietary tryptophan upon microbial degradation (Figure 8). The production of indoles from tryptophan depends on gut bacteria; consistently, their production is reduced by the use of antibiotics (Sun et al., 2013) or modulated by probiotics (Xu, Hu, & Wang, 2002). Indoles are absorbed in the colon, thus entering systemic circulation. Whilst indole seems to diffuse across cell membranes without the aid of a specific transporter (Sender, et al., 2016), the absorption for other indolic compounds has not been clearly described.

Although several studies evaluated the impact of single species in the production of indoles, the contribution of the gut microbiota, as a whole complex ecosystem, remains unclear. To date, no studies have systematically addressed the relative contribution of mammalian and/or microbial metabolism to the production of these compounds.

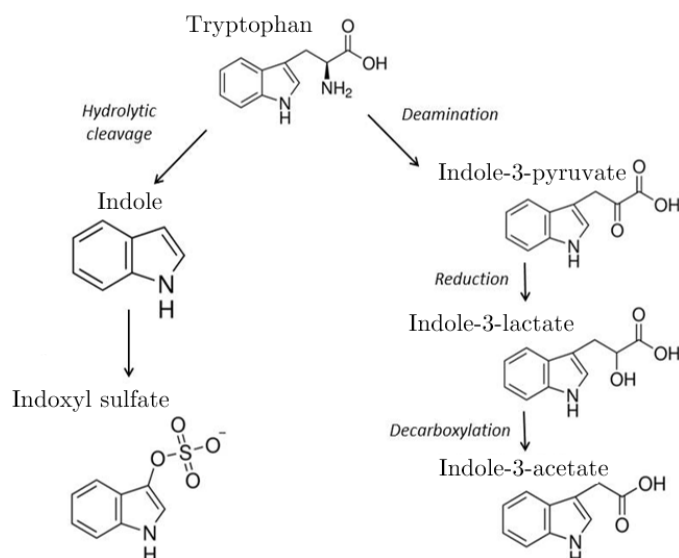


Figure 8. Generation of indoles from tryptophan by gut bacteria.

Tryptophan can be directly converted to indole through the activity of bacterial tryptophanase (present in the major member of the gut microbiome *Bacteroides thetaiotaomicron*, but also *Proteus vulgaris* and *Escherichia coli*, amongst others) (DeMoss and Moser, 1969). Indole can be conjugated in the liver to form indoxyl sulfate (DeMoss and Moser, 1969). *Clostridium* spp. and *Lactobacillus* spp. are able to deaminate tryptophan, producing indole-3-pyruvate, indole-3-lactate, indole-3-acetate and 3-methylindole (Jensen, et al., 1995) (Wikoff, et al., 2009).

### 2.6.2 Relevance of indoles to obesity and cardiovascular health

Indoxyl sulfate has been associated with cardiovascular mortality (Barreto et al., 2009) and various deleterious vascular outcomes. In endothelial cells, it induces oxidative stress by modifying the balance between pro- and anti-oxidant mechanisms, stimulating the release of endothelial micro-particles and diminishing endothelial-healing ability (Faure et al., 2006); it also aggravates cardiac fibrosis and cardiomyocyte hypertrophy by increasing oxidative stress (Yisireyili et al., 2013). Indole-3-propionate is involved in inflammatory mechanisms and in the maintenance of intestinal barrier integrity (Cheng, Shah, & Gonzalez, 2012; Venkatesh et al., 2014; Zelante et al., 2013). Most pro-inflammatory and pro-oxidative effects of indoles in vascular endothelial cells are thought to be mediated by the aryl hydrocarbon receptor (AHR) (Ito, et al., 2016; Jin, et al., 2014). This ligand-activated transcription factor forms a nuclear heterodimer with the AHR nuclear translocator protein to activate genes involved in the signalling of nuclear factor-erythroid 2-related factor-2 (Nrf2), p53 (TRP53), retinoblastoma (RB1), and NF-κB (Jin, et al., 2014).

In 2013, it was demonstrated for the first time that obese patients show higher numbers of genes associated with the degradation of tryptophan, suggesting that circulating indoles may also be increased in obesity (Le Chatelier, et al., 2013); however, to date, no study in adults has demonstrated this association. Recently, a study in a paediatric cohort firstly reported for that obese children present a higher production of indoxyl sulfate, as well as a higher abundance of genes involved in the production of this metabolite (Zhang, et al., 2015).

The putative targets and signalling pathways mediating the effect(s) of indoles in adipocytes and their mechanism of action on energy metabolism remain unclear. AHR expression is progressively decreased and eventually lost during adipocyte differentiation (Shimba, et al., 2003), making it an unlikely modulator of lipid accumulation. The identification of these targets is, therefore, critical to understand the molecular crosstalk underlying the putative impact of indoles in obesity and CVD.

### **3. A combined metabonomic and cell-based strategy to investigate the interactions between obesity and gut microbial metabolites**

A range of high-throughput technologies (*i.e.* genomics, transcriptomics, proteomics) have been used to provide an integrated perspective on human health and disease (Tyers and Mann, 2003; Wang, et al., 2009). However, since these methods focus on the impact of individual genes or proteins, they explain only a small proportion of the etiopathology of complex phenotypes such as obesity and cardiometabolic disorders.

The study of the metabolites contained in human biofluids (collectively designated as the human metabolome) has emerged as a more promising alternative (Nicholson, et al., 2008; Wishart, et al., 2007). The human metabolome provides a global picture of human physiology, where gene expression, protein translation, epigenetics and environmental stimuli converge into an altered metabolic signature or metabotype (Gavaghan, et al., 2000). Furthermore, exploratory metabolic phenotyping captures systems-level information, reflecting the effect of genetic and environmental cues on the microbial-mammalian symbiotic system (Holmes, et al., 2011; Holmes, et al., 2008; Nicholson, et al., 2002; Nicholson, et al., 2005). It is important to note that the metabolic response of living systems is not constant over time; conversely, it is affected by both external factors and genetic modification. The quantitative measurement of this dynamic multiparametric response has been coined metabonomics (Lindon, et al., 2003; Lindon, et al., 2004; Nicholson, et al., 2002; Nicholson, et al., 1999).

The two most important analytical approaches in metabonomics are mass spectrometry (MS) and nuclear magnetic resonance spectroscopy ( $^1\text{H}$  NMR) (Markley, et al., 2017; Nicholson, et al., 2011).  $^1\text{H}$  NMR spectroscopy has several advantages, mostly related to its high reproducibility and easy sample preparation (Dona, et al., 2014; Dumas, et al., 2006; Keun, et al., 2002; Nicholson, et al., 1995). This analytical method produces high-density data that can be further analysed by multivariate statistical analysis and pattern recognition methods (Lindon and Nicholson, 2008; Nicholson, et al., 2002; Nicholson, et al., 1999; Smolinska, et al., 2012).

Both MS and  $^1\text{H}$  NMR can be run in either untargeted or targeted mode. Untargeted profiling encompasses the quantification of as many metabolites as possible in a single biological sample; therefore, this mode does not require an *a priori* hypothesis and is frequently used to discover novel markers or metabolic associations. The main disadvantage of this mode includes eventual difficulties in the identification and quantification of a given metabolite, as the method is not optimised for individual compounds (Tzoulaki, et al., 2014). Conversely, targeted mode requires a previously defined hypothesis and can be optimized to quantify metabolites of interest. Semi-targeted analysis is frequently used to perform the measurement of a class or group of metabolites (*e.g.* bile acids, phenolics) (Tzoulaki, et al., 2014).

Metabolic phenotyping of obese and lean subjects has been used to highlight the contribution of mammalian metabolites involved in energy metabolism processes, such as the TCA cycle and lipid and amino acid homeostasis (Elliott, et al., 2015). The role of the microbial-mammalian interaction was also emphasised by various  $^1\text{H}$  NMR-based studies focusing on the development of obesity and cardiometabolic disorders (Calvani, et al., 2010; Dumas, et al., 2017; Elliott, et al., 2015; Nicholson, et al., 1984). In particular, the Intercontinental Study in Macronutrients in Arterial Pressure showed that BMI is linked to five gut microbial-host co-metabolic pathways, including choline degradation, protein colonic putrefaction and the production and excretion of hippurate (Elliott, et al., 2015). Indoles, particularly indoxyl sulfate, have been consistently detected in metabonomic studies (Garcia-Perez, et al., 2012; Heinzmann, et al., 2012; Zhang, et al., 2015).

It is important to note that metabonomic data are particularly useful to identify significant associations between metabolites and genetic, environmental or microbial cues, but they are not sufficient to generate causal inferences. Nevertheless, the associations found in metabolic profiling can be particularly useful for leveraging further experimental designs, thus moving from correlations to causation. With respect to microbial-mammalian signalling, the establishment of causality requires the demonstration of a direct effect either *in vitro* and/or *in vivo*, complemented by the clarification of the targets and signalling pathways involved.

Therefore, in this project,  $^1\text{H}$  NMR-based metabolic profiling was combined with high-throughput kinase screening, cell-based and molecular studies, and *in vivo* validations, to provide a comprehensive assessment of the role of indoles in obesity and cardiometabolic disorders.

## 4. Scope of the thesis: hypothesis and aims

The gut microbiota plays a pivotal role in obesity and cardiometabolic diseases. Obese subjects exhibit significant gut microbial changes, including higher numbers of microbial genes associated with the degradation of tryptophan, which is converted by tryptophanase-containing bacteria into indoles (Le Chatelier, et al., 2013). A recent study in a paediatric cohort reported a positive association between obesity and indoxyl sulfate levels (Zhang, et al., 2015); however, no studies have validated this finding. Furthermore, the human targets, signalling pathways and cellular responses putatively involved in this effect remain unexplored.

The overall hypothesis of this work is that gut microbial metabolites - and in particular indoles - bind to human kinases, impact signalling pathways and interfere with cellular processes relevant to obesity and cardiometabolic disorders, thus acting as chemical messengers in the microbiome-kinome crosstalk.

The specific aims of this project included:

1. To systematically describe the contribution of the gut microbiota to the production of indoles (Chapter 3);
2. To characterise the contribution of indoles to the metabolic signature of obesity (Chapter 4);
3. To identify novel pharmacological targets for indoles in the human kinome (Chapter 5 and Chapter 6);
4. To validate the role of indoles *in vitro* and *in vivo* (Chapter 5, Chapter 7 and 8).

An overview of the structure of this thesis is provided in Figure 9.

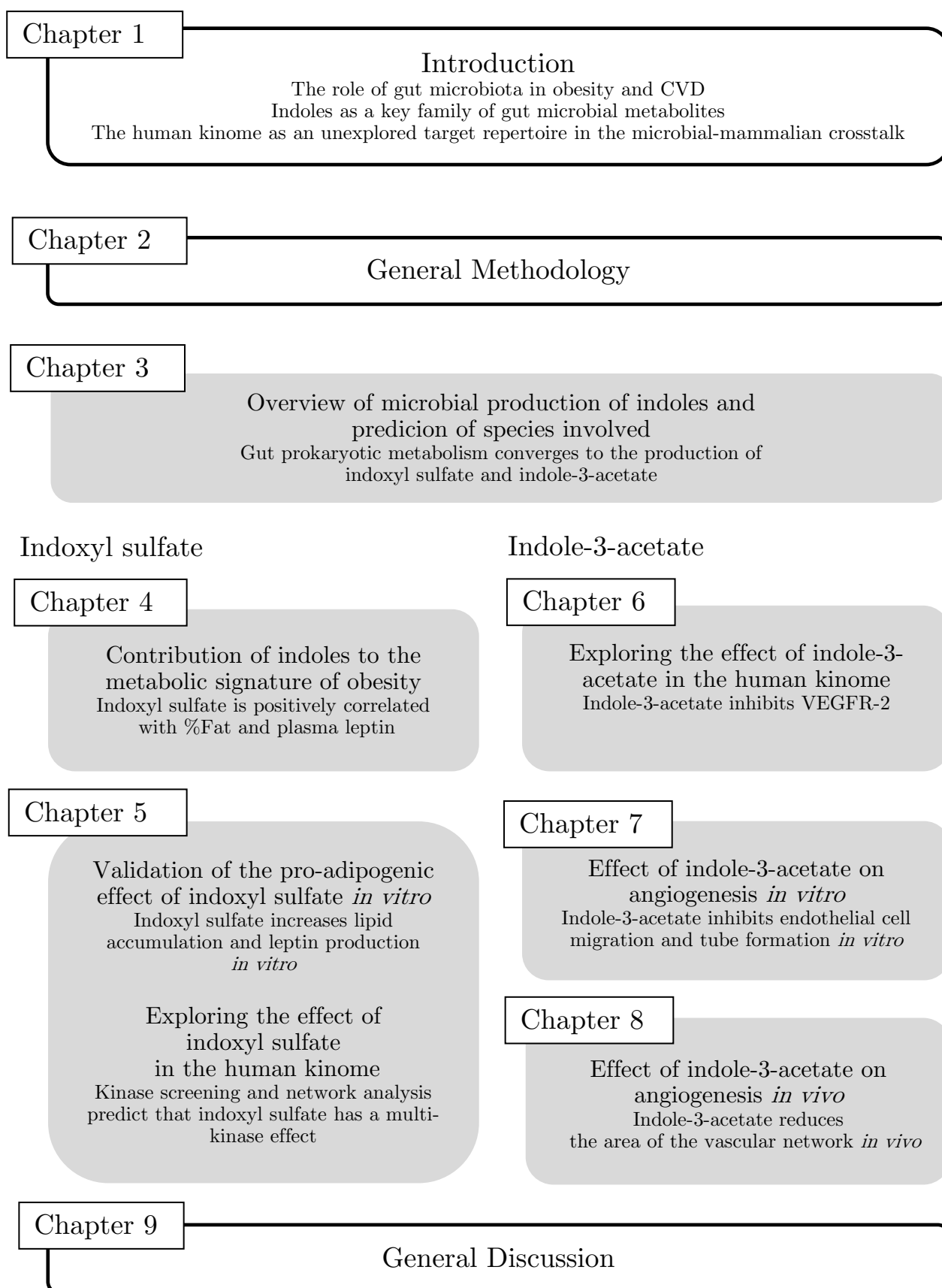


Figure 9. Thesis structure.





## Chapter 2

# General Methodology

Several methodologies were applied in order to address the specific aims of this project. In order to systematically describe the contribution of the gut microbiota to the production of indoles (Chapter 3), a combined approach using network analysis and the Basic Local Alignment Search Tool (BLAST) was used. The contribution of indoles to the metabolic signature of obesity and cardiometabolic disorders (Chapter 3) was investigated using a  $^1\text{H}$  NMR-based metabonomic approach. The potential targets of indoles in the human kinome were subsequently screened by a high-throughput kinase screening using KdSELECT technology; molecular docking analyses were used to further characterise the nature of the interactions (Chapter 5 and Chapter 6). Finally, the roles of indoxyl sulfate and indole-3-acetate were investigated in two different cell lines, by using a combination of cell-based and molecular approaches (Chapter 4 and Chapter 7). The anti-angiogenic effect of indole-3-acetate was validated *in vivo* using a subcutaneous Matrigel plug model (Chapter 8).

# 1. Cell lines and cell-based assays

In this PhD, two different cell lines (3T3-L1 and HUVEC) were used to validate the effect of indoles in *in vitro*-based assays.

## 1.1. 3T3-L1

3T3-L1 is a secondary cell line extensively used as an *in vitro* model to study lipogenesis, due to its ability to differentiate into mature adipocytes (Poulos, et al., 2010). Derived from murine fibroblasts obtained from Swiss 3T3 embryos, 3T3-L1 cells exhibit a similar cellular morphology to fibroblasts, during proliferation (Green and Kehinde, 1975; Green and Meuth, 1974). However, after reaching confluence, proliferation is inhibited; at this stage, upon the addition of pro-adipogenic compounds (insulin, dexamethasone and 3-isobutyl-1-methylxanthine (IBMX)), 3T3-L1 cells are able to differentiate into adipocyte-like cells (Green and Kehinde, 1975; Green and Meuth, 1974; Rubin, et al., 1978; Russell and Ho, 1976). The induction of differentiation initiates deep phenotypic changes in 3T3-L1, including the adoption of a spherical shape and the presence of fully developed lipid droplets (Poulos, et al., 2010). Mature adipocytes are characterised by an upregulation of the expression of pro-adipogenic genes (Gerhold, et al., 2002; Student, et al., 1980).

Being derived by clonal expansion, 3T3-L1 cells proliferate as a homogenous cellular population and, therefore, exhibit a homogenous response to treatments up to 20 cellular passages (Green and Kehinde, 1975; Green and Meuth, 1974). These characteristics support their choice as one of the most used cell lines to investigate the adipogenic potential of both endocrine and pharmacological compounds (Green and Meuth, 1974). A full description of the protocol used to maintain 3T3-L1, as well as to induce its differentiation into mature adipocytes, is shown in Chapter 5.

### 1.1.1. Cell viability

The assessment of the cell viability is a key step in cell-culture based assays. For adherent cell lines, viable cells attach to the cell culture plates; cells that undergo cell death lose their adherence and are subsequently liberated to the medium. Therefore, many cell viability methods are based on the staining of adherent cells as an indirect measure for cell viability (Xu, et al., 2015).

Crystal Violet staining is a simple and reliable method to stain adherent cells and thus quantify cell viability (Feoktistova, et al., 2016). Crystal Violet is a triarylmethane dye that binds to DNA and stains cell nuclei (Feoktistova, et al., 2016). Non-viable cells are removed by thoroughly washing the cell culture plates with water. Finally, the dye is eluted with methanol and quantified by measuring the optical density (absorbance) at 590 nm, providing a quantitative readout for cell viability (Feoktistova, et al., 2016). A full description of the method is shown in Chapter 5.

### 1.1.2. Lipid staining (Oil Red O)

The intracytoplasmic accumulation of lipids is one of the key phenotypic features of differentiated 3T3-L1 cells (Green and Meuth, 1974; Poulos, et al., 2010). This process mimics lipid accumulation (lipogenesis) observed in obesity *in vivo* and has, therefore, been extensively used as a model to validate putative pro- or anti-adipogenic compounds.

The staining of lipid droplets using Oil Red O has been used as a quantitative assay to measure lipogenesis *in vitro* (Cheung, et al., 2015; Janderova, et al., 2003; Kuri-Harcuch, et al., 1978; Rossmeislova, et al., 2013). Oil Red O is a fat-soluble diazo dye with a high affinity for lipids and cholesterol esters, but low affinity for biological membranes (Fowler and Greenspan, 1985). Also, Oil Red O has a higher affinity for lipids than for its own solvent (Fowler and Greenspan, 1985). These characteristics allow the entry of the dye to the cytoplasm, where it stains the lipid droplets. In order to allow quantitative measurements, the dye is eluted using 2-propanol and its absorbance photometrically determined at 518 nm (Kuri-Harcuch, et al., 1978).

## 1.2. Human Umbilical Vein Endothelial Cells (HUVEC)

Human Umbilical Vein Endothelial Cells (HUVEC) are the most commonly used cell line to study vascular endothelial cell function and angiogenesis (Crampton, et al., 2007). Isolated from human umbilical veins upon collagenase digestion, HUVEC have fastidious growth factor requirements, including endothelial cell growth supplements, heparin and high concentrations of Foetal Bovine Serum (FBS) (Crampton, et al., 2007; Jaffe, et al., 1973). HUVEC can be obtained either as primary (e.g. ATCC PCS-100-010) or as established cell lines (e.g. ATCC CRL-2873 HUVEC-CS or ATCC CRL-1730, HUV-EC-C), the latter of which was used in the present work.

HUVEC constitutively express VEGFR-2 (the main receptor for normal cell function and angiogenesis) and produce a range of signalling molecules associated with vascular physiology, *e.g.* nitric oxide (NO) (Boerma, et al., 2006). During proliferation, HUVEC exhibit a cobblestone phenotype (Marin, et al., 2001). However, under specific conditions, they can be differentiated into tridimensional cultures (Heiss, et al., 2015). These features are exploited in specific assays to study specific steps of angiogenesis *in vitro* (*i.e.* cell migration and tube formation), as explained in detail in *Sections 1.2.4.* and *1.2.5.*

### 1.2.1. Cell viability

See *Section 1.1.1.*

### 1.2.2. Phosphorylation assays

VEGFR-2 is the main receptor responsible for normal vascular endothelial cell function and angiogenesis (Huang, et al., 2012). In this work, VEGFR-2 phosphorylation was quantified by ELISA (PathScan Phospho-VEGFR-2 (Tyr1175) Sandwich ELISA Kit #7335; PathScan Total VEGFR-2 Sandwich ELISA Kit #7340, Cell Signalling Technology). ELISA is a highly specific, flexible and sensitive method

that exploits the formation of antigen-antibody complexes for quantification purposes (Engvall and Perlmann, 1971). Sandwich ELISA consists of adding a sample (*e.g.* cell lysate) containing the antigen (*e.g.* Total or Phospho-VEGFR-2) to a monoclonal antibody-coated well, forming an antigen-antibody complex (Engvall and Perlmann, 1971). After washing, a second enzyme-linked specific antibody for a different epitope of the antigen is added and allowed to react (Engvall and Perlmann, 1971). Finally, the substrate is added and converted by the enzyme into a coloured product (Engvall and Perlmann, 1971). The magnitude of the optical density of the developed colour is proportional to the amount of antigen.

VEGFR-2 phosphorylation activates a range of downstream signalling pathways, including the Protein Kinase B (AKT). AKT phosphorylation was quantified by the (Phospho (Ser473) / Total AKT Assay Whole Cell Lysate Kit, Meso Scale Discovery (MSD)). Although also a Sandwich immunoassay, this assay uses a Multi-Array technology that requires a minimal amount of sample (10-25  $\mu$ L) compared to traditional ELISA and allows the detection of several biomarkers simultaneously (Figure 10a). The assay consists in adding the sample (*e.g.* cell lysate) to a plate pre-coated with capture antibodies for Phospho-AKT (Ser473) and total AKT (Figure 10b). Next, a solution containing the detection antibody (conjugated with an electro chemiluminescent compound, MSD SULFO-TAG label) is added. Analytes in the sample bind to the antibodies immobilized on the working electrode surface; and recruitment of the conjugated detection antibody by bound analytes completes the sandwich (Figure 10c). Finally, a MSD read buffer is added, providing the appropriate chemical environment for electrochemiluminescence. The plate is loaded into an MSD SECTOR Imager, inside which a voltage is applied to the plate electrodes, causing the labels bound to the electrode surface to emit light. The instrument measures the intensity of emitted light to provide a quantitative measure of phosphorylated and total AKT in the sample.

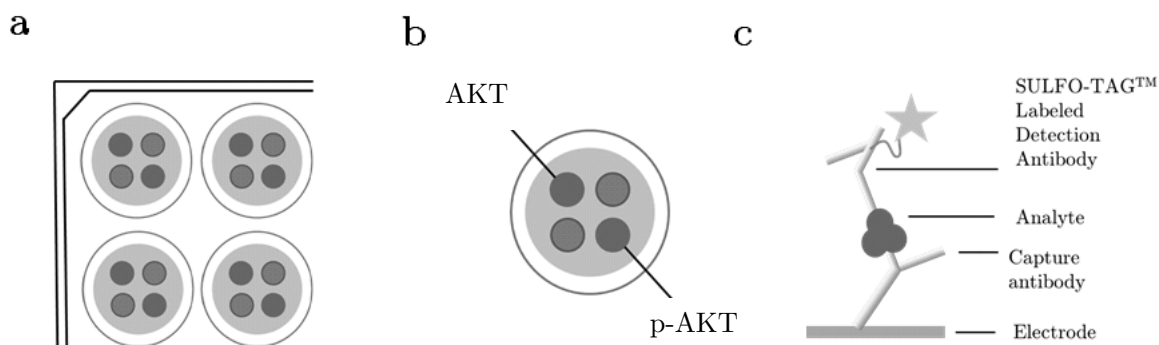


Figure 10. Principle of the MSD Phosphoprotein Assay.

Spot diagram showing the MSD® Phosphoprotein Multi-Array plate (a) and the placement of analyte capture antibody in each well (b). Analytes in the sample bind to the capture antibodies immobilized on the working electrode surface; recruitment of the conjugated detection antibody by bound analytes complete the sandwich (c). Adapted from (MSD, 2011). AKT: Protein Kinase B; p-AKT: Phospho-AKT

### 1.2.3. NO quantification

Amongst the range of signalling molecules produced by HUVEC, NO appears to be particularly relevant to vascular health (Tousoulis, et al., 2012). NO quantification was performed by the Griess Method (Tsikas, 2005), using the Total NO and Nitrate/Nitrite Parameter Assay (R&D Systems). NO is a gaseous free radical with a half-life of seconds; therefore, the levels of its more stable NO metabolites, nitrite ( $\text{NO}_2^-$ ) and nitrate ( $\text{NO}_3^-$ ), are usually used for its indirect measurement (Marletta, et al., 1988; Tsikas, 2005). This assay quantifies total NO following the enzymatic conversion of  $\text{NO}_3^-$  to  $\text{NO}_2^-$  by nitrate reductase.  $\text{NO}_2^-$  is detected by the Griess Reaction, a two-step diazotisation reaction in which acidified  $\text{NO}_2^-$  produces a nitrosating agent that reacts with sulfanilic acid to produce a diazonium ion (Figure 11a). The latter is finally coupled to *N*-(1-naphtyl)ethylenediamine to form the chromophoric azo-derivative that absorbs light at 540-570 nm (Figure 11b) (Miles, et al., 1996).

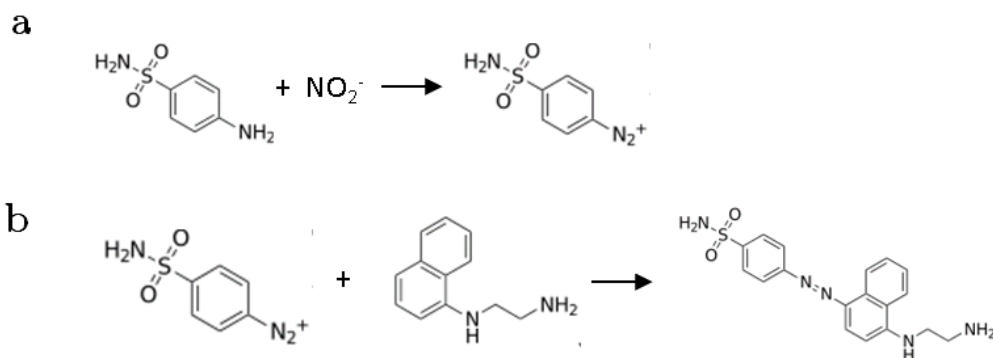


Figure 11. Griess Reaction.

The nitrosating agent generated from acidified nitrite reacts with sulfanilamide to yield a diazonium derivative (a). This reactive intermediate interacts with *N*-(1-naphthyl)ethylenediamine to yield a coloured diazo product (b) (Miles, et al., 1996).

#### 1.2.4. Cell migration assay

Cell migration is a feature of vascular endothelial cells during angiogenesis (Gerhardt, et al., 2003; Lamalice, et al., 2007; Potente, et al., 2011). In this work, an exclusion zone assay was used to assess the migratory abilities of endothelial cells, by evaluating their ability to cover a defined cell-free zone (AlMalki, et al., 2014; Staton, et al., 2004).

The Radius<sup>TM</sup> Cell Migration Assay Kit comprises a patented plate, containing a biocompatible hydrogel spot in each well, to which cells cannot attach. Cells are seeded and, once firm cell attachment is achieved, the hydrogel spot is removed, exposing a cell-free region to study cell migration (Figure 12). The percentage of cell-free area recovered by cells (Equation 1) provides a quantitative measure of their migrating ability (Yarrow, et al., 2004).

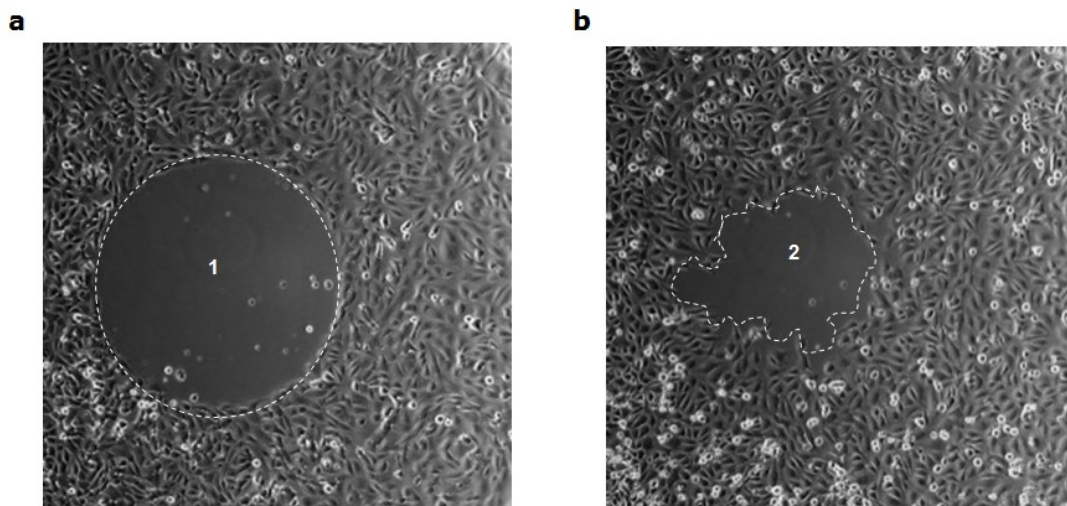


Figure 12. Cell migration assay: quantitative analysis.

a) Microscopic aspect of HUVEC attached outside of the hydrogel spot immediately after hydrogel spot removal ( $t = 0h$ ), defining a cell exclusion zone (1). b) Microscopic aspect of vehicle-treated HUVEC, 4 hours after hydrogel spot removal ( $t=4h$ ).

$$\% \text{ Area recovered} = \left( 1 - \frac{\text{cell - free area at } t = x \text{ (where } x = 2h, 4h \text{ or } 6h)}{\text{cell - free area at } t = 0h} \right) \times 100$$

Equation 1. Quantification of cell migrating activity.

### 1.2.5. Tube formation assay

The tube formation assay has been widely adopted to validate pro- and anti-angiogenic compounds *in vitro* (Arnaoutova and Kleinman, 2010; Kubota, et al., 1988). Capillary-like tube formation is a complex phenotype that requires endothelial cells to synthesise proteases, to digest the matrix (*e.g.* Matrigel), to migrate towards each other and, finally, to align in order to form tubular structures; therefore, it complements the previous cell migration assay and provides a more comprehensive assessment of angiogenesis.

Matrigel is a basal membrane extract usually obtained from Engelbreth-Holm-Swarm (EHS) mouse sarcoma cells (Passaniti, et al., 1992). Its major components include collagens, heparan sulfate proteoglycans, laminins, fibronectin, matrix metalloproteinase (MMP) proenzymes and growth factors (FGF, TGF $\beta$ , VEGF<sub>A</sub>)

(Haralabopoulos, et al., 1994). In this assay, liquid Matrigel (4°C) is used to coat cell culture labware, followed by incubation at 37 °C to allow the proteins to polymerise; HUVEC are then cultured on the newly-formed gel (Malinda, 2001). Activated by the components of the basement membrane extract, HUVEC start exhibiting complex cellular behaviours and forming capillary-like structures (Haralabopoulos, et al., 1994; Malinda, 2001). The networks are usually fully developed after 16-24h and can be quantified either by manual or automatic methods. (Auerbach, et al., 2003; Carpentier, 2012).

## **2. *In vivo* mouse study: Matrigel subcutaneous plug**

The Matrigel subcutaneous plug assay is a fast and effective method to validate stimulators or inhibitors of angiogenesis *in vivo* (Fridman, et al., 2012; Tahergorabi and Khazaei, 2012).

The principle of the assay is that, when subcutaneously injected in mice, Matrigel solidifies forming a “plug”. When angiogenic factors are present in the Matrigel, endothelial cells are chemotactically attracted into the plug, forming vessels (Baker, et al., 2006; Malinda, 2001). After 7-10 days, the Matrigel plug is removed and the newly formed blood vessels are observed and quantified for assessment of angiogenesis (Baker, et al., 2006; Hasan, et al., 2004).

## **3. <sup>1</sup>H Nuclear Magnetic Resonance Spectroscopy**

<sup>1</sup>H Nuclear Magnetic Resonance (<sup>1</sup>H NMR) spectroscopy was used to characterise the contribution of indoles to the metabolic signature of obesity and cardiometabolic disorders. This reproducible research technique detects small molecular weight molecules present in complex biological matrices, allowing the characterisation of the metabolic profile of a given individual, at a particular time point and with minimal sample preparation effort (Holmes, et al., 2008; Nicholson, et al., 2002; Nicholson, et al., 1999). Over recent decades, considerable improvements in

spectral dispersion and the sensitivity of the technique enabled this method to detect metabolites in micromolar concentrations (Holmes, et al., 1997).

### 3.1. Principles of $^1\text{H}$ NMR Spectroscopy

$^1\text{H}$  NMR spectroscopy exploits the magnetic properties of atomic nuclei to provide detailed information about the structure, chemical environment and abundance (quantification) of the molecules that contain them.

Atomic nuclei are composed of protons and neutrons. All isotopes with an odd number of protons and/or neutrons exhibit an angular momentum – the “*nuclear spin*” – different than zero. The nuclear spin quantum number ( $I$ ) depends on the atomic number (*i.e.* number of protons) and atomic mass (number of nucleons) present (Jardetzky and Roberts, 1981). Whilst all nuclei with  $I \neq 0$  are detectable by  $^1\text{H}$  NMR spectroscopy, the most frequently used in biological approaches have  $I=1/2$  (*i.e.*  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{31}\text{P}$ ) (Jardetzky and Roberts, 1981). Amongst these,  $^1\text{H}$  is the most abundant element in biological molecules and therefore the most frequently used in NMR metabonomic studies (Jardetzky and Roberts, 1981).

Upon application of a magnetic field ( $B_0$ ), the nuclear moments of the spins orient themselves in a set of possible orientations. The number of orientations is defined by the equation  $2I+1$  (Keeler, 2002). As an example, for  $I=1/2$ , there are two different orientations of the nuclear spin: a stable, “low-energy spin” (position  $\alpha$ ) and a high-energy spin (position  $\beta$ ) (Figure 13a). For each orientation, the spin is characterised by a frequency ( $\nu$ , in Hz) and an angular speed ( $\omega$ , in  $\text{Rad.s}^{-1}$ ), as defined by the Larmor precession (Keeler, 2002).

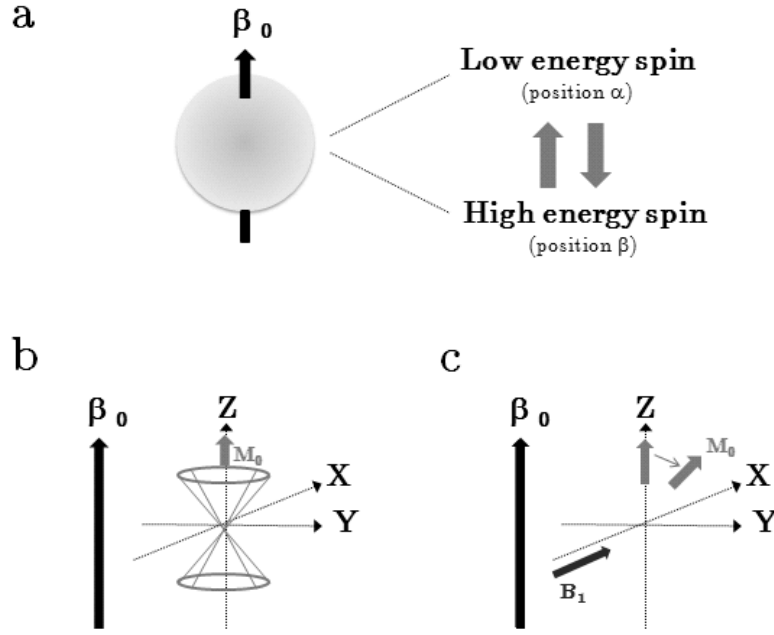


Figure 13. Principles of  $^1\text{H}$  NMR Spectroscopy.

As a consequence of the application of a magnetic field ( $B_0$ ), the nuclear moments of the spins orient themselves in several possible orientations. For  $I=1/2$ , there are two different orientations of the nuclear spin: a low-energy spin (position  $\alpha$ ) and a high-energy spin (position  $\beta$ ) (a). Motion of the magnetised nuclei, upon action of this magnetisation vector  $M_0$  (b). Upon action of a specific pulse of radiofrequency ( $B_1$ ), some nuclear spins in the lower energy level are encouraged into high energy levels, promoting a new orientation of the magnetisation vector  $M_0$  (c). When the pulse ( $B_1$ ) stops, the excited spins return to their low energy status. Adapted from (Jardetzky and Roberts, 1981; Keeler, 2002).

The population of the two energy levels is described by the Boltzmann distribution (Equation 2), where  $N_{\text{upper}}$  is the number of nuclei in the upper energy state,  $N_{\text{lower}}$  is the number of nuclei in the low energy stage,  $K$  is the Boltzmann constant and  $T$  is the absolute temperature (in Kelvin) and  $\Delta E$  is the energy difference between both levels.

$$N_{\text{upper}} / N_{\text{lower}} = e^{-\Delta E/KT}$$

Equation 2. Boltzmann distribution.

The difference on the populations of both energy levels originates a magnetisation vector ( $M_0$ ) following the same direction as  $B_0$ . The energy difference between both energy levels ( $\Delta E$ ) can be calculated by Equation 3, where  $\gamma$  is the

gyromagnetic ratio (nuclear magnetic moment),  $h$  is the Plank's constant and  $B_0$  the NMR magnetic field.

$$\Delta E = \gamma h B_0 / 2\pi$$

Equation 3. Energy distribution between high- and low-energy statuses.

If a specific pulse of radiofrequency ( $B_1$ ) is applied at the resonant frequency  $\omega$ , some nuclear spins in the lower energy level are encouraged into high energy levels, promoting a new orientation of the magnetisation vector  $M_0$  (Figure 13c). When the pulse ( $B_1$ ) stops, the excited spins return to their low energy status. This relaxation process comprises two aspects: a) the *longitudinal (spin-lattice) relaxation time*, that refers to the time constant for the return of  $M_0$  to the basal position in the z axis and b) the *transverse (spin-spin) relaxation time* that refers to time constant for the nuclear spins to return to the basal status in the X-Y plane. During this return to their basal status, the component of magnetisation releases a voltage, the *Free Induction Decay* (FID) that originates the  $^1\text{H}$  NMR signal. A Fourier transform is applied to convert the time domain signal of FID in a frequency domain and finally generate a  $^1\text{H}$  NMR spectrum (Jardetzky and Roberts, 1981; Keeler, 2002).

$^1\text{H}$  NMR spectroscopy provides not only information about the atomic nuclei, but also on the molecular structure and conformation. This is possible because each nucleus is directly affected by its surrounding chemical environment: for example, electronegativity, electron density and circulating electron ring currents generate a magnetic field around the nucleus that shields the nucleus from the external applied magnetic field. The difference between the applied field and the nuclear magnetic field (*nuclear shielding*) and the ratio between the nuclear shielding and the applied magnetic field (*chemical shift*) provide further information on the molecular structure (Jardetzky and Roberts, 1981; Keeler, 2002).

Finally, the splitting patterns of  $^1\text{H}$  NMR signals can provide additional spectral information about the connectivity of atoms within the molecule. Close magnetic nuclei can interact with each other spatially or via spin-spin coupling. In this case, spin-spin couplings from neighbouring groups belonging to the same molecule result in the

splitting of the NMR signals into two or more components. The complexity of the splitting pattern is a function of the number of neighbouring nuclei. According to the general rule of  $n+1$ , where  $n$  is the number of protons in adjacent atoms, a nucleus with two neighbouring protons will have its  $^1\text{H}$  NMR signals split in three peaks (*i.e.* a triplet) (Keeler, 2002).

### 3.2. Monodimensional and multidimensional $^1\text{H}$ NMR Spectroscopy

$^1\text{H}$  NMR spectroscopy can be mono-dimensional (1D), bi-dimensional (2D) or multi-dimensional (nD). 1D NMR experiments start with the *preparation time* (or relaxation delay), during which the spins reach their equilibrium state. Next, a radio-frequency pulse is used - the *pulse width* - to stimulate the nuclear spins into high-energy statuses. When the spins return to their basal low energy state, generate a FID that is acquired during the *acquisition time*. A 1D NMR spectrum can be obtained in less than 5 minutes, however, in order to increase the signal-to-noise ratio, the radio-frequency pulse is repeated multiple times, with a resulting increase on the time of the experiment (Van, et al., 2008). Although 1D NMR has successfully identified a range of metabolites abundant in human biofluids, less abundant metabolites are often hidden by overlap with other compounds (Van, et al., 2008). Several techniques were used to alleviate peak congestion, such as the use of isotopically enriched metabolites (Higashi, et al., 2014), sample fractionation (Borlak, et al., 2003) and spectral fitting of a set of known metabolites (Weljie, et al., 2006). Some 1D NMR pulse sequences, such as the spin-echo experiment, optimise the detection of low molecular weight compounds by reducing broad signals of macromolecules (Hahn, 1950).

Conversely, 2D NMR experiments add a second, modulated delay in the pulse sequence, and usually comprise four steps: 1) the *preparation time*, 2) the *evolution time*, where no pulses are applied and the nuclear spins freely rotate; 3) the *mixing time*, when a series of radio-frequency pulses are applied and finally the 4) *acquisition time*, during which the FID signal from the sample is detected (Jardetzky and Roberts, 1981; Keeler, 2002). In 2D NMR, each frequency axis is associated with one of the two

time variables (*evolution time* and *acquisition time*), both of which are converted from a time series into a frequency series via a two-dimensional Fourier transformation (Keeler, 2002). Due to the deconvolution of the signal across a second dimension, 2D NMR provides more information than 1D NMR and is especially useful in the determination of the molecular structure (Van, et al., 2008). Types of 2D NMR methods include Correlation Spectroscopy (COSY), Heteronuclear Single Quantum Coherence Spectroscopy (HSQC) and the J-resolved experiment (*J-Res*) (Foxall, et al., 1993; Keeler, 2002; Posma, et al., 2017; Rodriguez-Martinez, et al., 2017; Van, et al., 2008).

### 3.3. Unsupervised multivariate statistical analysis

Metabolic profiling by NMR spectroscopy generates high-density data. Multivariate methods have been proposed as a useful approach to summarise the main sources of variation in these datasets, to perform discriminant analyses and to recognise patterns (Eriksson, et al., 2004). Multivariate analyses can be either unsupervised (no hypothesis required for model construction) or supervised (model constructed based in a previously established hypothesis).

Amongst unsupervised methods, Principal Component Analysis (PCA) is one of the simplest and widest used methods (Nicholson, et al., 2007). In brief, PCA is a statistical procedure that transforms a set of observations (*e.g.* NMR spectra) into a new coordinate system of Principal Components (PC), which axes or directions describe the dominant pattern of the matrix (Hotelling, 1933; Jolliffe, 2002). The first PC (PC1), built from linear least squares analysis in order to maximise the variance of the data, accounts for the maximal variability observed (Hotelling, 1933; Pearson, 1901). The second PC is orthogonal to the first one, and is added to improve the modelling of the data variance. Each succeeding PC has the highest possible variance, provided that it is orthogonal to the previous component (Hotelling, 1933; Jolliffe, 2002; Pearson, 1901).

Geometrically, the principal components of the model ( $n=K$ ) are used to construct a  $K$ -dimensions space into which the observations are projected, according to their respective variables. The coordinates of the observations in the  $K$ -dimensions space are represented by the scores (T), whilst the directions of the PC in the initial space are represented by the loadings (P). Both scores and loadings can be used to construct, respectively, the scores plot and the loadings plot. While the scores plot shows the relationship between observations and reveals patterns in the dataset, the loadings plot provides information on the contribution of the variables to the model (Keeler, 2002; Trygg, et al., 2007). The part of the matrix that is not explained by the PCA model (the “residuals” (E)) is visually represented as the distance between each observation and its projection on the principal components space. The PCA matrix is modelled by the equation:

$$X = TP^T + E$$

Equation 4. Construction of the Principal Component Analysis matrix

### 3.4. Supervised multivariate statistical analysis

Contrarily to unsupervised methods, supervised statistical analysis models are based in previously established hypotheses. Therefore, these methods are particularly relevant when metabonomic profiling is used to find differences between groups (*e.g.* treated versus untreated) (Keeler, 2002).

Partial Least Squares (PLS) is one of the most commonly used supervised methods in metabonomic analyses (Nicholson, et al., 2007). PLS regression calculates a linear regression model between two matrices (*e.g.*  $X$  = spectral data,  $Y$  = quantitative values such as time, concentration or clinical data) and identifies the maximal multidimensional variance between them (Bollard, et al., 2005). When  $Y$  is categorical, a variant of PLS is used (Partial Least Squares Discriminant Analysis (PLS-DA) (Andersen, et al., 2012)).

Spectral data from NMR often contain a certain level of variation that is not correlated with the  $Y$  matrix, and can therefore affect the interpretation of the PLS

model. Orthogonal Partial Least Squares (OPLS) was developed with the purpose of overcoming this limitation (Cloarec, et al., 2005; Fonville, et al., 2010). In this method, an integrated Orthogonal Signal Correlation (OSC) is incorporated in the model, partitioning the X variance into predictive and orthogonal (uncorrelated) components of the Y matrix (Cloarec, et al., 2005; Fonville, et al., 2010). As consequence, the variation of interest is optimised and the uncorrelated variation is minimised, leading to improved interpretability and easier visualisation (Cloarec, et al., 2005). However, the OSC filter does not improve predictability (Trygg and Wold, 2002). OPLS-DA (Discriminant Analysis) is an extension of OPLS for discrete variables, particularly useful in classification studies (Bylesjo, et al., 2006).

The robustness of the resulting models can be inferred from the  $R^2_{Yhat}$  and the  $Q^2_{Yhat}$  values, that estimate, respectively, the percentage of variation explained and the predictive ability of the model (Steel and Torrie, 1960).  $R^2$  values range between 0 and 1 and, generally, the higher the value the better the model fits the data (Steel and Torrie, 1960).

Cross-validation is usually performed by partitioning the data into complementary subsets, performing the analysis on one subset (the *training set*) and validating the analysis in the other data subset (*validation set* or *testing set*) (Seymour, 1993). This operation is repeated until the whole dataset has been used as *validation set*, providing the value of  $Q^2Y$  (*i.e.* the percentage of the predicted variance from the Y vector in the model).  $Q^2Y$  values equal or above 0.5 are considered reliable for data interpretation (Eriksson, et al., 1999; Eriksson, et al., 2004). Another alternative for model validation consists of performing a permutation test; in this method, a number of parallel models based on the fit of randomly re-ordered Y data is generated, and the  $Q^2Y$  is considered valid if significant different ( $p < 0.05$ ) from the distribution of  $Q^2Y$  values generated in the random models (Eriksson, et al., 1999; Eriksson, et al., 2004).

### 3.5. Analytical and statistical strategies for biomarker identification

Metabolite identification is performed by comparing the spectrum with existing databases that provide information on the shape and coupling of its signals (*e.g.* Human Metabolome Database) (Wishart, et al., 2007). Whenever their assignment is not obvious, both analytical and statistical strategies can be used.

Regarding analytical strategies, 2D NMR experiments have proven extremely effective at identifying novel and known metabolites. J-resolved (*J-Res*) NMR remains one of the most popular among 2D experiments for metabolite assignment due to its simplicity and short acquisition time (Dona, et al., 2016). Through the *J-Res* experiment, a simplified projection of the proton spectrum in which all multiplet peaks appear as singlet can be extracted, with a resulting reduction in the complexity of the spectrum is achieved and clarification of most overlapping problems (Nagayama, et al., 1977). Furthermore, in 2D *J-Res* NMR the coupling constant (*J* value) of each signal is represented in a second dimension (Nagayama, et al., 1977). In order to assign a given metabolite, both the chemical shifts and the coupling constants should be close to the values reported in existing databases (Dona, et al., 2016).

Statistical techniques such as Statistical Total Correlation Spectroscopy (STOCSY), can provide further clarification (Cloarec, et al., 2005). STOCSY calculates the correlation matrix between a given point and the remainder of the spectrum, allowing the extraction of all its correlated peaks. As the signals of each molecule correlate with its other peaks across the spectrum, STOCSY provides valuable information for metabolite assignment – but also for the identification of other metabolites correlated with the driving signal (Cloarec, et al., 2005). Another statistical approach, SubseT Optimisation by Reference Matching (STORM), selects subsets of spectra with specific signatures, providing fewer false positive correlations and a simpler interpretation of the correlation matrix, in comparison with STOCSY (Posma, et al., 2012).

Even when a given metabolite is confidently identified by NMR, it can be sometimes important to further confirm its assignment using analytical approaches, for example, when identifying novel or important biomarker (Dona, et al., 2016). Finally, the gold standard for metabolite experiment is the “*spike-in experiment*” (Bouatra, et al., 2013), in which a small volume of a solution of the compound of interest is added to the sample, with a resulting increase in the intensity of its signal – and, consequently, the unequivocal identification of its NMR signature.

## 4. Kinase screening

Consisting on 518 kinases, the human kinome represents a broad repertoire of signal transduction triggers for human homeostasis and metabolism (Cohen, 2002; Manning, et al., 2002). Indoles’ targets in the human kinome targets were assessed using KdSELECT technology (DiscoverX Corporation), as described is previously published papers (Fabian, et al., 2005; Karaman, et al., 2008). This technique is based on a competition-binding assay that quantitatively measures the ability of a compound to compete with an immobilized, active site-directed ligand. The assay consists of an immobilized ligand, a DNA-tagged kinase and a test compound (Figure 14).

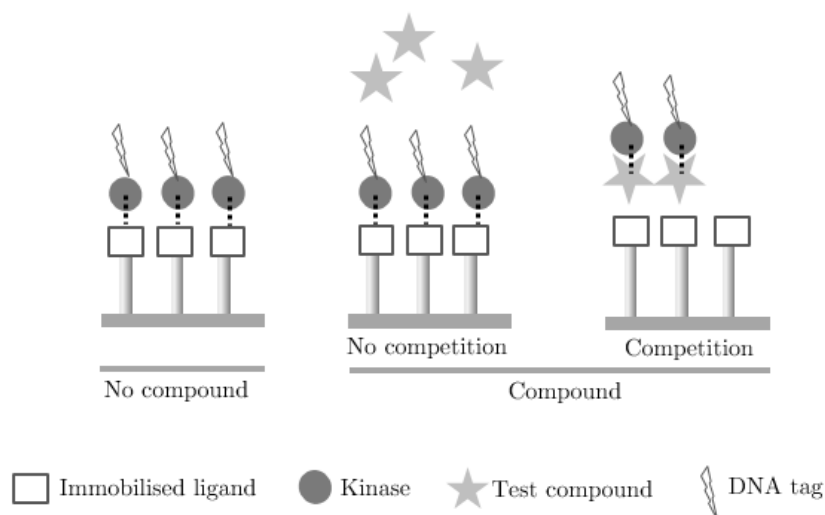


Figure 14. Kinase screening by KINOMEscan technology.

Compounds that bind the kinase active site prevent kinase binding to the immobilized ligand, reducing the amount of kinase captured on the solid support (Fabian, et al., 2005; Karaman, et al., 2008).

The ability of the test compound (*i.e.* indolic compound) to compete with the immobilized ligand was measured by quantitative PCR of the DNA tag. Results for primary screen binding interactions are reported as % of control (%Ctrl), where lower numbers indicate stronger hits (Equation 5).

$$\%Ctrl = \left( \frac{\text{test compound signal} - \text{positive control signal}}{\text{negative control signal} - \text{positive control signal}} \right) \times 100$$

Equation 5. Quantification of primary screen binding interactions in the kinome screening.

For positive hits, the binding constant ( $KD$ ) was then calculated from an 11-point dose-response curve. The TREEspot Software Tool was used to generate kinase interaction tree plots, which were reprinted with permission from KINOMEscan (DiscoverX Corporation). For compounds with inhibitory activity, the half-maximal inhibitory concentration ( $IC_{50}$ ) was determined using the Kinexus kinase-inhibitor activity profiling service (Kinexus). The technique is based on the direct quantification of radiolabelled phosphate from ATP onto a protein substrate of a target protein kinase.

## 5. Molecular docking

In this work, molecular docking analyses were used to investigate the complexes formed by a putative ligand (*i.e.* indolic compound) and a kinase target. Molecular docking is a modelling method that predicts the relative interaction of the two elements of the complex, providing information on the strength and type of signal generated by the interaction (Kitchen, et al., 2004). Molecular docking focuses on computational simulations (*e.g.* GOLD software) to predict an optimised conformation such as the free energy of the overall complex is minimised. In order to achieve this, molecular docking involves three steps: 1) input of the kinase structure, 2) a search algorithm and 3) a scoring function (Kitchen, et al., 2004).

The structure of the kinase of interest was retrieved from The Protein Data Bank (PDB), the most comprehensive database for experimentally determined protein structures (Berman, et al., 2000). Next, the search algorithm provided by GOLD® software (Cambridge Crystallographic Data Centre) provides a list of the possible orientations and conformations of the protein paired with the putative ligand. For this purpose, a genetic algorithm was used; in brief, genetic algorithms use the energy scores of each position (*pose*) as a fitness function to select input for the next iteration (Mitchell, 1996). The poses generated by the genetic algorithm are then evaluated using a scoring function (Ajay and Murcko, 1995). Most scoring functions incorporate molecular mechanics force fields estimates (*i.e.* solvent effects, conformational changes, internal rotations, association energy and free energy) to calculate the energy of each pose. Individual scores for each pose are finally ranked, lower energies indicating a more stable system and thus more likely interactions (Ajay and Murcko, 1995).

## 6. Basic Local Alignment Search Tool (BLAST)

In this project, a BLAST-based strategy was used to predict gut microbial proteins involved in the metabolism of tryptophan into indoles. BLAST is the most widely used bioinformatics algorithm to compare primary biological sequence information (*e.g.* amino acid sequences of proteins, nucleotide sequences of DNA) (Altschul, et al., 1990).

Specifically, BLAST allows the comparison between a query sequence (target sequence) with a previously compiled library or database. In this case, the tryptophan metabolism reactions pathway performed by human or prokaryotes was identified from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) using the application programming interface (API) (<http://www.kegg.jp/kegg/docs/keggapi.html>). For each reaction, the KEGG protein target sequences (FASTA file) were searched against a compiled database of prokaryotic proteomes from RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>) and cultivation-based studies targeting the gut microbiota (Browne, et al., 2016; Lagier, et al., 2016) using BLAST.

The BLAST algorithm identifies quickly and accurately high scoring sequence alignments between the target sequence and a database, using a heuristic approach (Altschul, et al., 1990). BLAST starts by removing low-complexity regions or sequence repeats in the query sequence and partitions the query protein sequence in smaller sequences, which are then listed as high-scoring or low-scoring matches. High-scoring matches are converted into an efficient search tree, allowing the program to rapidly compare database sequences (Altschul, et al., 1990).

BLAST is a bioinformatics approach that uses a family of tools, each one individually customised according to the query sequence, the database searched and the comparison intended. An overview of various BLAST approaches is shown in Table 1 (adapted from (Johnson, et al., 2008)).

Table 1. Overview of different types of Basic Local Alignment Search Tools

| BLAST type (acronym)                             | Target sequence                                    | Features                                                                                                            |
|--------------------------------------------------|----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Nucleotide-nucleotide (BLASTN)                   | Nucleotide sequence                                | Returns nucleotide sequence                                                                                         |
| Protein-protein BLAST (BLASTP)                   | Protein sequence                                   | Returns the most similar protein sequences from the protein database                                                |
| Position-Specific Iterative BLAST (PSI-BLAST)    | Protein sequence                                   | Returns protein sequences, particularly sensitive for distant evolutionary relationships                            |
| Nucleotide 6-frame translation protein (BLASTX)  | Translation product of a nucleotide query sequence | Compares the translation products of a nucleotide query sequence (both strands) against a protein sequence database |
| Protein-nucleotide 6-frame translation (TBLASTN) | Protein sequence                                   | Compares a protein query against a nucleotide sequence database.                                                    |

Information adapted from (Johnson, et al., 2008)

For protein sequence queries, the most commonly used approaches are BLASTP (for similar sequences) and PSI-BLAST (for distant evolutionary relationships) (Table 1).

## 7. Networks and graph theory

Graph theory is the study of mathematical structures used to model pairwise relationships between discrete objects (Biggs, et al., 1986). Although the first paper of the history of graph theory was first published in 1736, the actual term “graph” was coined by J. Sylvester, in 1878 (Sylvester, 1878). Graph theory has a wide range of applications in physics, computer science, engineering, finance, sociology and biology. With the advent of high-throughput technologies (‘omics’) and the exponential generation of publicly available data, the analysis of molecular networks has also gained a significant interest in biology and medicine (Bashan, et al., 2012; Habibi, et al., 2014).

### 7.1. Graphs and networks

Graphs are employed as models for systems of objects (*i.e.* nodes) and the relationships between them (*i.e.* edges) (Gentleman, et al., 2005). Edges are binary relationships that join two nodes; they can be directed or undirected. A directed edge is an edge where one extremity is designated the *head* (starting point) and the other the *tail* (endpoint). *Directed graphs* (or *digraphs*) can be used to model, for instance, transcription factor or signalling pathways, where the direction of the steps involved is well characterised. The “*underlying graph*” of a digraph is the graph resulting from converting all directed edges into undirected (Gentleman, et al., 2005). A network can be defined as a graph in which nodes and/or edges have specific attributes (*e.g.* names). Visually, nodes are represented by a dot and edges are represented by a line. In directed graphs, the direction of an edge is indicated by an arrow (Figure 15).

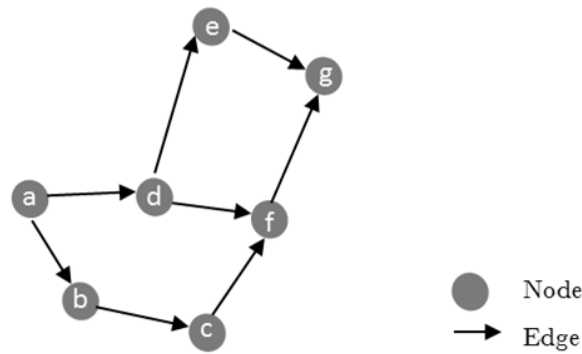


Figure 15. Example of a simple directed graph.

## 7.2. Network optimisation and centrality measures

Network analysis can be used as a tool to find optimal pathways within a given network. Furthermore, information in the relative relevance of individual nodes and edges in a graph can be identified by centrality measures (Freeman, 1977; Gentleman, et al., 2005).

Centrality measures are based on the assumption that nodes into which the highest number of paths converge are particularly relevant in the overall network (Bonacich, 1987). As an example, in biological pathways, a node with a higher centrality would have more relevance in the network, as more signalling pathways – and, therefore, more biological information – would converge or pass through it.

Betweenness centrality is a widely used measure of centrality, first described by Linton Freeman in 1977 (Freeman, 1977). For every pair of nodes in a given directed network, there is at least one shortest path between them such that either the number of edges that the path passes through is minimised. The betweenness centrality for each node is the number of these shortest paths that pass through the vertex, calculated by the equation

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Equation 6. Betweenness centrality of a given node.

where  $\sigma_{st}$  is the total number of shortest paths from node  $s$  to node  $t$ , and  $\sigma_{st}(v)$  is the number of shortest paths that pass through  $v$  (Equation 6)(Freeman, 1977).





## Chapter 3

# Overview of microbial production of indoles and prediction of species involved

## 1. Introduction

The human gut microbiota communicates with its human host through a range of signalling metabolites that bind to human targets and impact their signalling pathways, thus modulating human health (Chilloux, et al., 2016; Russell, et al., 2013). Amongst these metabolites, indoles - heterocyclic compounds produced upon bacterial degradation of tryptophan - appear to be particularly relevant for human health.

Indoles have been linked to deleterious cardiovascular phenotypes *both in vitro* and *in vivo*. In clinical studies, levels of indoxyl sulfate seem to be a powerful predictor of overall and cardiovascular mortality (Barreto, et al., 2009). Indoxyl sulfate induces oxidative stress *in vitro* (Faure, et al., 2006), a putative mechanism through which it

enhances cardiac fibrosis and cardiomyocyte hypertrophy. Indole-3-acetate is also a predictive marker of cardiovascular mortality; additionally, it induces endothelial inflammation (*via* AHR) and oxidative stress (Dou, et al., 2015). Recently, it was also suggested that indoles may be involved in obesity (Le Chatelier, et al., 2013; Zhang, et al., 2015).

The production of indoles from substrates depends on the functionalities of the gut microbiota; consistently, their production is reduced by antibiotic treatment (Sun, et al., 2013) and affected by prebiotics (Xu, et al., 2002). Several bacterial strains have been reported to be involved in the production of indoles. Tryptophan can be directly converted to indole by tryptophanase, a bacterial lyase reported to be present in *Bacteroides thetaiotaomicron*, *Proteus vulgaris* and *Escherichia coli* (Jean and DeMoss, 1968; Wesoly and Weiler, 2012); indole is further conjugated in the liver to form indoxyl sulfate. Alternatively, tryptophan can be deaminated by *Clostridium* spp. and *Lactobacillus* spp., producing indole-3-pyruvate, indole-3-lactate, indole-3-acetate and 3-methylindole (Attwood, et al., 2006; Mohammed, et al., 2003; Russell, et al., 2013; Wesoly and Weiler, 2012; Whitehead, et al., 2008; Wikoff, et al., 2009). Studies performed in gnotobiotic mice demonstrated that the production of another indolic compound, indole-3-propionate, was completely dependent on gut microbial activity and colonisation with a single bacterial species, *Clostridium sporogenes*, was able to establish its production (Wikoff, et al., 2009). Whilst the aforementioned studies evaluated the impact of single species in the formation of individual indolic compounds, there is a need for a systematic approach investigating the contribution of the gut microbiota, as a complex ecosystem, to the formation of indoles. In order to fill this gap in knowledge, a computational framework was used to predict the contribution of the gut microbiota to the production of indoles.

The specific aims of this chapter include:

- a) To identify the prokaryotic metabolic reactions involved to the production of indoles;
- b) To identify and predict gut prokaryotic species able to perform these metabolic reactions.

## 2. Materials and Methods

### 2.1. Tryptophan reaction network

A reaction network based on the tryptophan metabolism reference reaction pathway from KEGG (rn:R00380) (Kanehisa and Goto, 2000) was built in R (R Core Team, 2013). The reactions of the network were linked to their corresponding KEGG gene orthology identifiers using the KEGG API (<http://www.kegg.jp/kegg/docs/keggapi.html>, on 14 November 2016). Only reactions with gene orthologues corresponding to human, bacteria and/or archaea were retained in the network. Three additional reactions (rn:R00681, rn:R01971, rn:R01973) were included in the network despite not being linked to a orthology gene ID, because there is evidence that these reactions can occur in some species of bacteria (Du, et al., 2015; Hornemann, et al., 1971; Hornemann, et al., 1970; Jean and DeMoss, 1968; Roberts and Rosenfeld, 1977; Speedie, et al., 1975; Takai, et al., 1977). The filtered network was visualized in Cytoscape v.3.4.0 (Shannon, et al., 2003).

### 2.2. Protein Basic Local Alignment Search Tool (BLASTP)

Proteome sequences for 1313 prokaryotes were downloaded from the Human Microbiome Project (HMP) (<http://hmpdacc.org/HMRGD/all/#data>) on 14 November 2016. A non-redundant set of proteomes (in terms of species coverage) representing 11,161 bacterial and 345 archaeal species was downloaded from RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>) on 14 November 2016. Proteomes for whole-genome data described in recent cultivation-based studies targeting the gut microbiota (n=239; (Browne, et al., 2016) and n=175 (Lagier, et al., 2016)) were downloaded from GenBank and EBI. A fasta file of the concatenated proteomes was used to create a database against which the KEGG protein sequences associated with each reaction were searched using the Protein-Protein Basic Local Alignment Search Tool BLASTP (Johnson, et al., 2008). BLASTP is an local alignment search tool customised for protein searches; in brief, the algorithm compares amino-acid sequences of proteins from a query sequence with a library or database of sequences, identifying sequences

that resemble the query sequence above a certain threshold (Johnson, et al., 2008). In this case, data were filtered using an E value cut-off of  $1 \times 10^{-3}$  and 90% sequence coverage; only hits with > 95 % homology to database sequences were retained.

## 2.3. Subnetwork analysis

The results of the subnetwork analyses were visualized using different approaches. For indoxyl sulfate, relevant species of bacterial were clustered by phylum and visualised as a network in Cytoscape v.3.4.0. For indole-3-acetate, a reaction-species heatmap colour-coded according to the ability of a given bacterium to perform each reaction was generated. In order to highlight the most relevant species, the reaction-species matrix was filtered as follows. Each path was assigned a value (path-value) reflecting its contribution to the corresponding pathway (*i.e.*, total number of reactions performed by the bacterial species /total number of reactions in a given pathway) (Equation 7).

$$\text{Path - value} = \left( \frac{\text{Number of reactions performed (in a given pathway)}}{\text{Total number of reactions (in a given pathway)}} \right)$$

Equation 7. Calculation of the path-values.

Each path-value was calculated as the ratio between the number of reactions performed by a given bacterium and the total number of reactions of the metabolic pathway.

Next, each individual bacterial score was calculated as the sum of all path-values (*i.e.* bacterial contribution to each one of the paths involved in the production of a given compound) (Equation 8). According to these filtering criteria, the top 38 ranked species were selected.

$$\text{score} = \sum_{\text{Path} = 1}^n (\text{Path - value})$$

Equation 8. Calculation of the individual bacterial score values.

Individual scores were calculated as the sum of all the path-values ( $n$  = total number of paths involved in the production of a given compound).

### 3. Results

### 3.1. Contribution of prokaryotic metabolism to the production of indoles

In order to identify the contribution of the human gut microbiota to metabolic steps involved in the production of indoles from tryptophan, a network illustrating the metabolic reactions involved was computed, based on KEGG (Kanehisa and Goto, 2000) (Figure 16). Only metabolic reactions performed by human, bacterial and archaeal species were included (reactions performed by other organisms were excluded and are described in Table 2). Based on this, 16 metabolites and 18 metabolic reactions (7 prokaryotic-specific, 4 human-specific and 7 present in both human and prokaryotes) were retained (Figure 16). In total, 39% of the reactions of the network are performed exclusively by bacteria and archaea species, reflecting the relevance of prokaryotes to the metabolism of indoles.

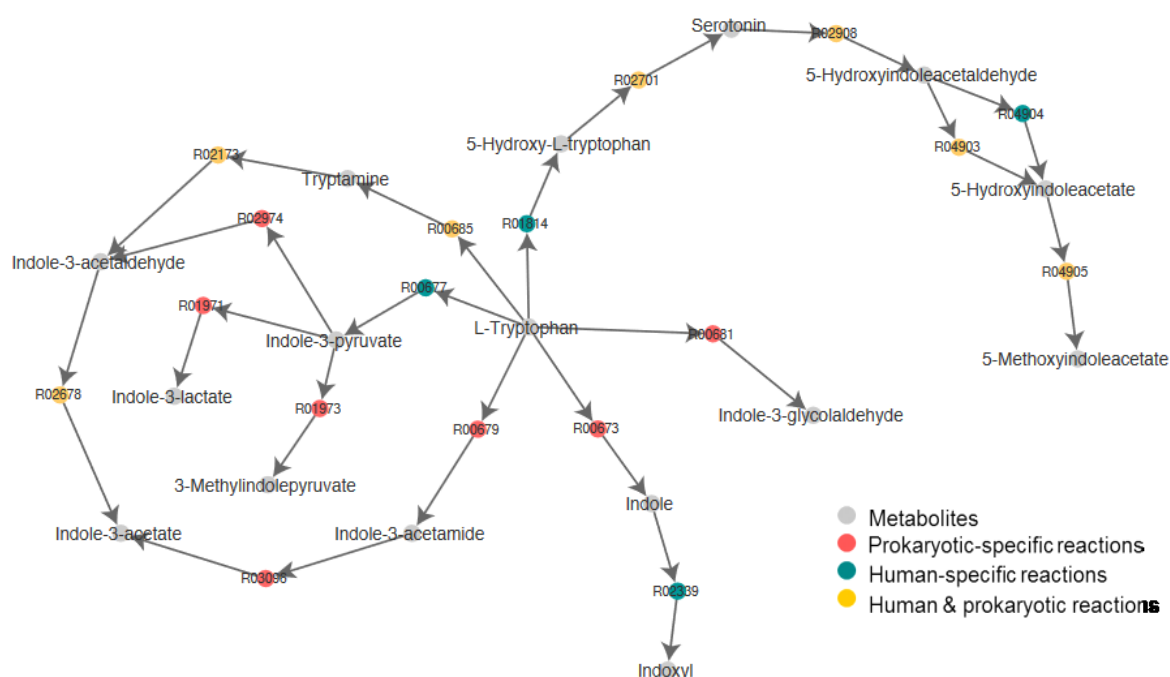


Figure 16. Overview of the production of indoles from tryptophan.  
Network illustrating the metabolic reactions involved in tryptophan catabolism; all reactions linking *L*-Tryptophan to indole-3-acetate or to indoxyl are irreversible (unidirectional).

Table 2. List of the metabolic reactions involved in tryptophan metabolism.

| KEGG<br>Reaction ID* | KEGG<br>Orthology | Gene              | Definition                                                    | Organism              |
|----------------------|-------------------|-------------------|---------------------------------------------------------------|-----------------------|
| R00673               | K01667            | <i>tnaA</i>       | Tryptophanase                                                 | Prokaryotic           |
| R00677               | K03334            | <i>IL4I1</i>      | L-amino-acid oxidase                                          | Human                 |
| R00679               | K00466            | <i>iaaM</i>       | Tryptophan 2-monooxygenase                                    | Prokaryotic           |
| R00684               | K14265            | <i>Tam1</i>       | Tryptophan aminotransferase                                   | Other                 |
| R00685               | K01593            | <i>DDC</i>        | Aromatic-L-amino-acid decarboxylase                           | Human and prokaryotic |
| R01814               | K00502            | <i>TPH1_2</i>     | Tryptophan 5-monooxygenase                                    | Human                 |
| R01971**             | n/a               | n/a               | n/a                                                           | Prokaryotic           |
| R01973**             | n/a               | n/a               | n/a                                                           | Prokaryotic           |
| R01974               | K04103            | <i>ipdC</i>       | Indole-3-pyruvate decarboxylase                               | Prokaryotic           |
| R02173               | K11182            | <i>AOC1, ABP1</i> | Diamine oxidase                                               | Human                 |
|                      | K00274            | <i>MAO</i>        | Monoamine oxidase                                             | Human and prokaryotic |
| R02339               | n/a               | n/a               | n/a                                                           | Human                 |
| R02678               | K14085            | <i>ALDH7A1</i>    | Aldehyde dehydrogenase family 7 member A1                     | Human                 |
|                      | K00149            | <i>ALDH9A1</i>    | Aldehyde dehydrogenase family 9 member A1                     | Human                 |
|                      | K00128            | <i>ALDH</i>       | Aldehyde dehydrogenase (NAD <sup>+</sup> )                    | Human and prokaryotic |
| R02679               | n/a               | n/a               | n/a                                                           | Other                 |
| R02680               | n/a               | n/a               | n/a                                                           | Other                 |
| R02681               | K11817            | <i>AAOL_2</i>     | Indole-3-acetaldehyde oxidase                                 | Other                 |
| R02700               | K14265            | <i>Tam1</i>       | Tryptophan aminotransferase                                   | Other                 |
| R02701               | K01593            | <i>DDC</i>        | Aromatic-L-amino-acid decarboxylase                           | Human and prokaryotic |
| R02908               | K00274            | <i>MAO</i>        | Monoamine oxidase                                             | Human and prokaryotic |
| R03092               | n/a               | n/a               | n/a                                                           | n/a                   |
| R03096               | K01426            | <i>amiE</i>       | Amidase                                                       | Prokaryotic           |
| R04093               | K11868            | <i>CYP71A13</i>   | Indoleacetaldoxime dehydratase                                | Other                 |
| R04094               | K01237            | <i>E3.2.1.147</i> | Myrosinase                                                    | Other                 |
| R04903               | K00128            | <i>ALDH</i>       | Aldehyde dehydrogenase (NAD <sup>+</sup> )                    | Human and prokaryotic |
|                      | K00149            | <i>ALDH9A1</i>    | Aldehyde dehydrogenase family 9 member A1                     | Human                 |
|                      | K14085            | <i>ALDH7A1</i>    | Aldehyde dehydrogenase family 7 member A1                     | Human                 |
| R04904               | K00157            | <i>AOX</i>        | Aldehyde oxidase                                              | Human                 |
| R04905               | K00543            | <i>ASMT</i>       | Acetylserotonin <i>N</i> -methyltransferase                   | Human and prokaryotic |
| R04906               | n/a               | n/a               | n/a                                                           | n/a                   |
| R06804               | n/a               | n/a               | n/a                                                           | Other                 |
| R08160               | K11813            | <i>CYP79B2</i>    | Tryptophan <i>N</i> -monooxygenase                            | Other                 |
| R08164               | K11820            | <i>UGT74B1</i>    | <i>N</i> -hydroxythioamide <i>S</i> -beta-glucosyltransferase | Other                 |
| R08167               | K11821            | <i>5T5A_B_C</i>   | Desulfoglucosinolate sulfotransferase A/B/C                   | Other                 |
| R08168               | K11818            | <i>CYP83B1</i>    | Cytochrome P450 family 83 subfamily B                         | Other                 |
| R08170               | K11819            | <i>SUR1, RTY</i>  | <i>S</i> -alkyl-thiohydroximate lyase SUR1                    | Other                 |
| R08605               | n/a               | n/a               | n/a                                                           | n/a                   |
| R08606               | n/a               | n/a               | n/a                                                           | n/a                   |
| R08609               | n/a               | n/a               | n/a                                                           | n/a                   |
| R10180               | K16903            | <i>TAA1</i>       | L-tryptophan-pyruvate aminotransferase                        | Other                 |
| R10181               | K11816            | <i>YUC2</i>       | Indole-3-pyruvate monooxygenase                               | Prokaryotic           |

\*Metabolic reactions performed by human, bacterial and archaeal species are shown in bold and were included in the metabolic network. Reactions performed by other organisms, or without assigned KEGG ID were excluded. \*\*Reactions known to be carried out by bacteria, but without KEGG ID assigned. n/a, none assigned.

According to KEGG, indole-3-pyruvate is the sole indolic compound produced only by human metabolism. The production of all other indoles requires prokaryotic metabolism, either for all reactions (indole-3-acetate *via* indole-3-acetamide and indole) or at least for one reaction (indoxyl and indole-3-acetaldehyde). Bacterial metabolism is also required for the production of indole-3-lactate, indole-3-acetaldehyde and 3-methylindolepyruvate, although these reactions have not been characterised yet.

### 3.2. BLASTP as a tool to predict gut microbial species involved in tryptophan metabolism

Next, it was explored which prokaryotic species may contribute to the metabolic network leading to the production of indoles. Although KEGG represents a popular reference database, its description of the prokaryotes involved in indole production is not comprehensive. In order to augment the coverage of prokaryotes involved in indole production reported in KEGG, BLASTP of KEGG proteins against databases of curated proteome sequences was used. In brief, the protein sequences associated with each reaction of the KEGG-based network were searched against compiled databases of prokaryotic protein sequences, including RefSeq (n = 11,506), HMP (n=1,313) and cultivation studies (Browne et al., 2016 (n=239); Lagier et al., 2016 (n=175)); only hits with > 90% sequence coverage and 95% homology were retained. Globally, BLASTP allowed the identification of 404 new prokaryotic gene homologues potentially implicated in indole metabolism (Figure 17a). At the single reaction level, increases were found in the number of prokaryotic species able to perform a given reaction, ranging from +7.8% (aldehyde dehydrogenase (NAD<sup>+</sup>)) to +57.1% (Acetylserotonin-*N*-methyltransferase) (Figure 17b). Furthermore, human-specific reactions reported in KEGG (see Table 2) were also searched using BLASTP, confirming they were not performed by prokaryotes.

A filtering strategy to select prokaryotic species present in the human gut was applied. For this purpose, a database including microbial species present in the gastrointestinal tract from the Human Microbiome Project (<http://hmpdacc.org/HMRGD/all/#data>) and in the culture-based studies by (Lagier

et al., 2016) and (Browne et al., 2016) was compiled. In the gut-filtered dataset, BLASTP searches predicted a total of 49 new prokaryotic genes (Figure 17c). While some reactions are potentially performed by a reduced number of species (acetylserotonin-*N*-methyltransferase (n=2) and aromatic-*L*-amino-acid decarboxylase (n=6); corresponding respectively to 0.007 and 0.02% of the gut prokaryotic dataset), others are predicted to be performed by 48% gut prokaryotes (*e.g.* aldehyde dehydrogenase (NAD<sup>+</sup>), n=132) (Figure 17d). The conversion of *L*-tryptophan into indole by tryptophanase shows the highest percentage increase on the number of predicted species (+60%) (Figure 17d-e).

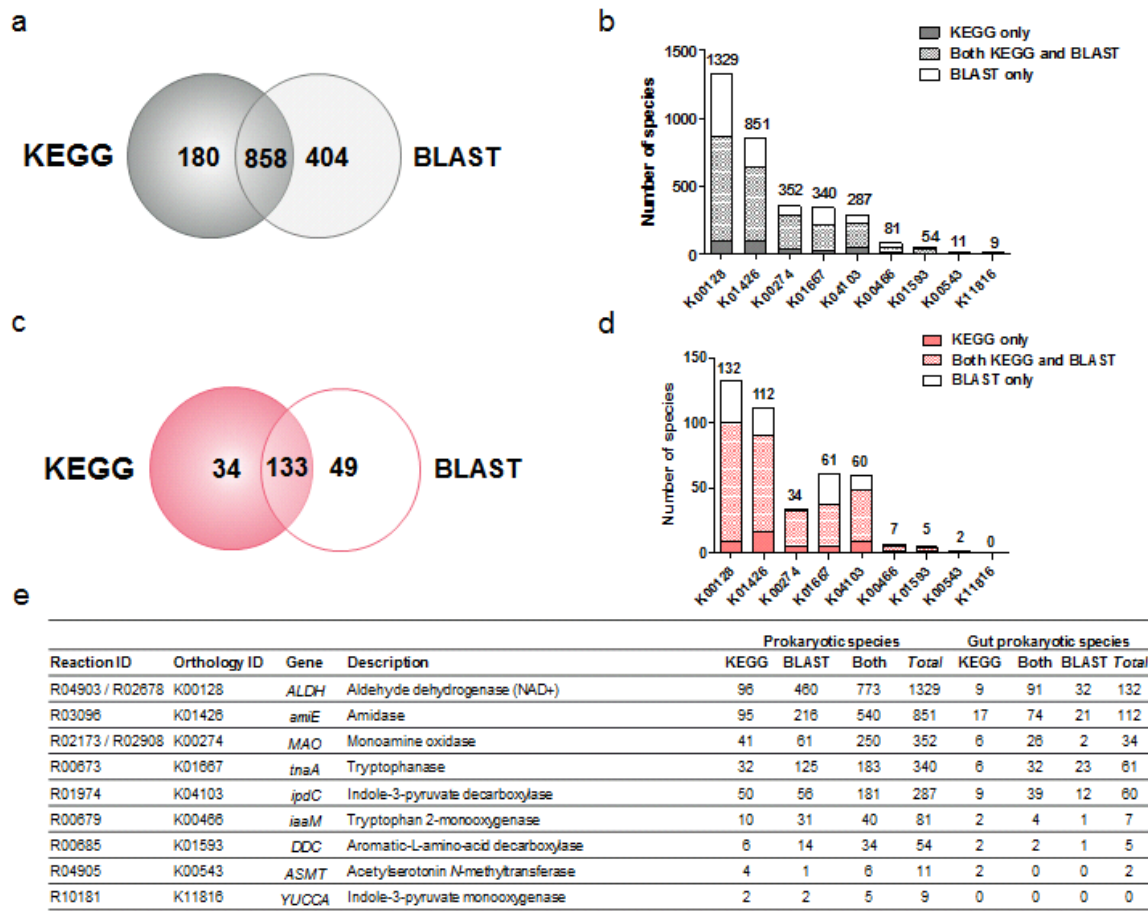


Figure 17. Prediction of prokaryotic species involved in the production of indoles. a) Number of prokaryotic genes involved in the production of indoles, detected by KEGG, BLASTP or both. b) Number of prokaryotic species involved in each metabolic reaction. c) Venn diagram representing the number of gut prokaryotic genes involved in the production of indoles, detected by KEGG, BLASTP or both, after filtering. d) Number of gut prokaryotic species involved in each metabolic reaction, after filtering. e) List of reactions of the metabolic network, with the respective contribution of KEGG and BLASTP to the identification of associated prokaryotic species, before and after filtering.

### 3.3. Prokaryotic communities involved in the production of indole

Having provided an overview of the prokaryotic metabolism of indoles, the next results focus on specific indolic compounds reported to be relevant for human health: indole and indole-3-acetate.

Indole is synthesised in the gut by the action of a prokaryotic lyase, tryptophanase, on tryptophan and is then converted by hepatic enzymes into indoxyl and indoxyl sulfate (Figure 18a). By merging the KEGG + BLASTP gut dataset, 61 prokaryotic species involved in this metabolic conversion were identified (Figure 18b). Data visualisation of these species, in the form of a phylum-clustered network, reveals a marked relative contribution of the phyla Bacteroidetes ( $n = 17$ ) and Proteobacteria ( $n = 30$ ) to this metabolic process (Figure 18b).

BLASTP predicted 23 more species than recognised previously able to potentially perform this reaction, including, amongst others: *Klebsiella oxytoca*, *Citrobacter farmeri*, *Shigella flexneri*, *Enterobacteriaceae bacterium*, *Aeromonas* spp. e *Escherichia* spp. (Proteobacteria); *Prevotella pallens*, *Prevotella nigriscens*, *Odoribacter laneus*, *Porphyromonas uenonis* and *Bacteroides* spp. (Bacteroidetes); *Propionibacterium* spp. and *Streptomyces* spp. (Actinobacteria); *Fusobacterium sp.* (Fusobacteria) and *Clostridium* spp. (Firmicutes) (Figure 18b). No new species were identified in the phyla Spirochaetes and Verrucomicrobia.

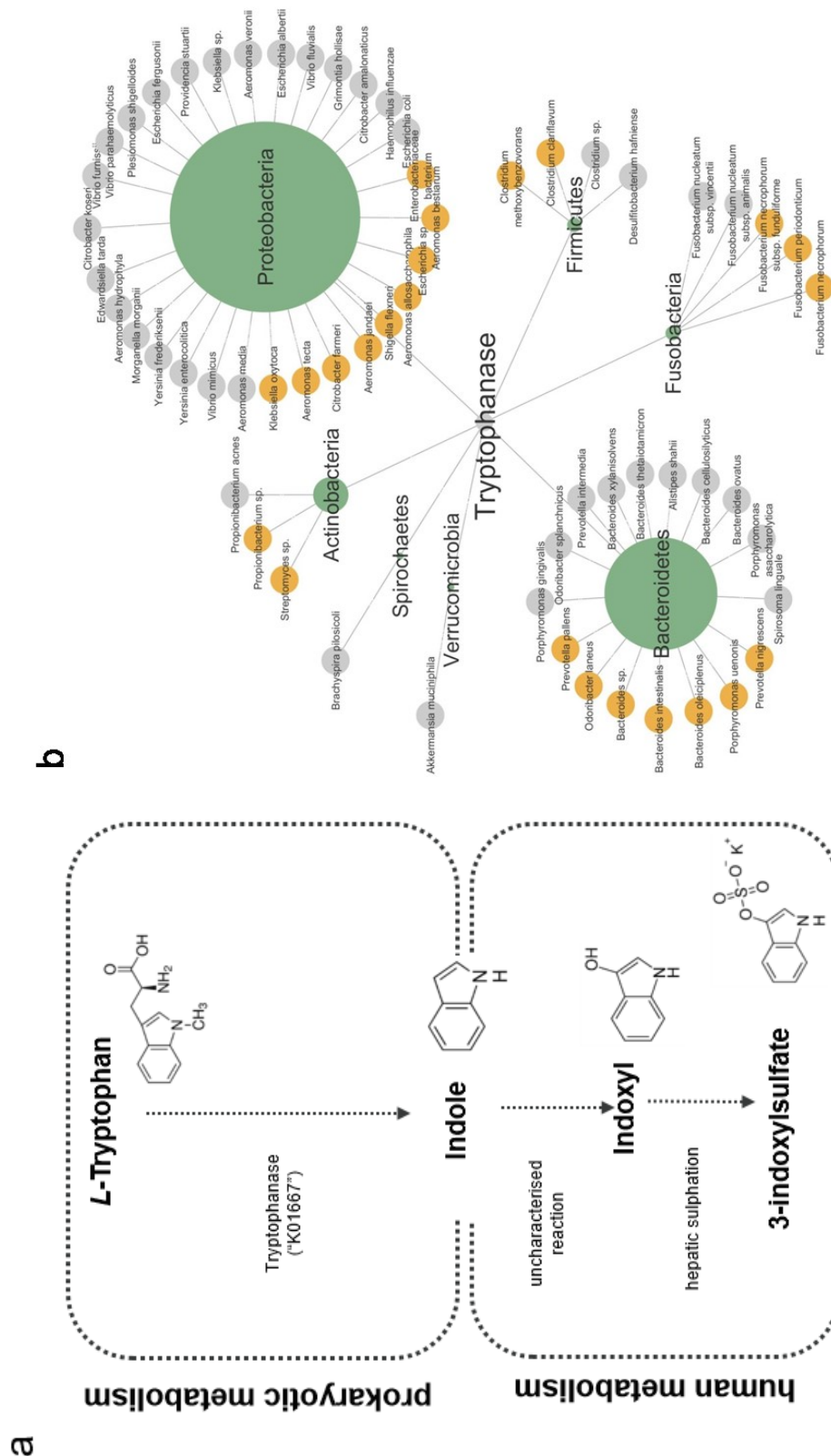


Figure 18. Contribution of the gut microbiota to the production of indole.

(a) Schematic representation of the production of indole *via* tryptophanase ("K01667"); (b) Network showing the prokaryotic species involved in the reaction, where species newly predicted by BLASTP are coloured in yellow. The size of the phylum nodes (green) reflects the number of species belonging to each phylum.

### 3.4. Prokaryotic communities involved in the production of indole-3-acetate

The production of indole-3-acetate exhibits increased metabolic complexity, occurring through three metabolic paths: *via* indole-3-acetamide (Path 1), indole-3-pyruvate (Path 2) or tryptamine (Path 3). Interestingly, while Path 1 is exclusively bacterial, both Path 2 and Path 3 exhibit microbial and mammalian co-metabolism (Figure 19a). An overview of the potential contribution of prokaryotic species to the individual reactions of the three paths is shown in Figure 19b. In each path, the first step is consistently the one performed by a lower number of prokaryotic species, suggesting a critical role for a subset of species at the initiation of these metabolic pathways.

A scoring system was created to highlight and rank the most relevant species, reflecting their overall contribution to the three metabolic paths. Score values assumed a bimodal distribution (Figure 19c), with higher scores corresponding to the ability of a microbe to perform a higher number of potential reactions (*for further details on how the scores were calculated, see Section 2. Methods*). After filtering the original heatmap using a score cut-off value of 1.66 (corresponding to the ability of a given species to perform 4 out of 7 reactions), 38 species were identified. As shown in Figure 19d, the species with highest scores belong almost exclusively to the phyla Proteobacteria (68.4%) and Firmicutes (28.9%). At the individual species level, *Pseudomonas putida* and *Pseudomonas fluorescens* exhibited the highest scores (score = 2.49), corresponding to their ability to potentially perform 6 out of the 7 metabolic reactions involved in the production of indole-3-acetate (Figure 19d). While most of these species were reported in KEGG, BLASTP searches allowed the identification of several additional species with high scores, such as *Pseudomonas nitroreducens* and *Raoultella terrigena*. *Pseudomonas nitroreducens* is potentially able to perform all the steps of Path 1 (exclusively bacterial) and complete the last steps of the other two paths. Conversely, *Raoultella terrigena* performs all the steps of Path 2 (initiated by human metabolism only, L-amino-acid oxidase (“K03334”)) and completes the last metabolic steps of the remaining paths.

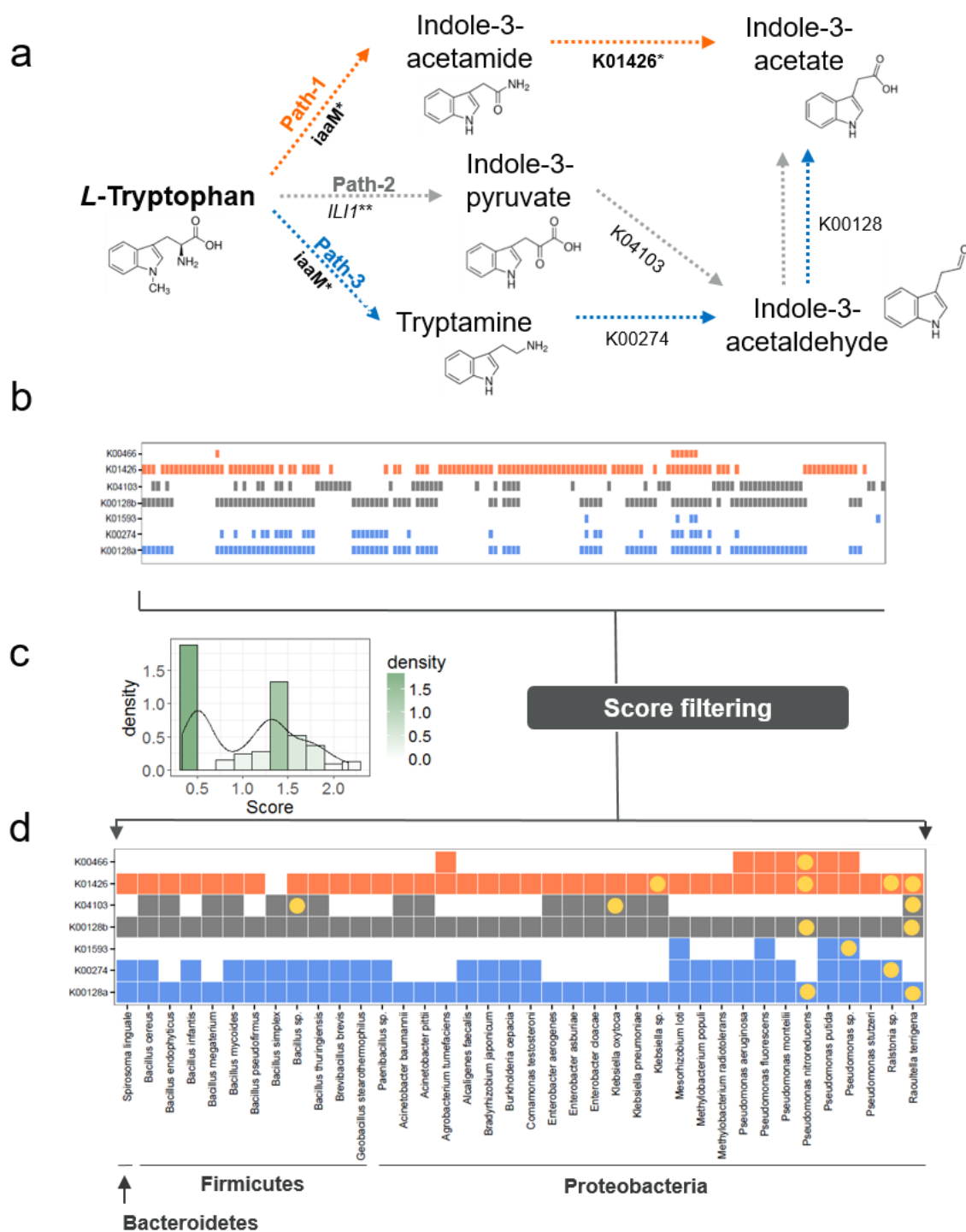


Figure 19. Contribution of the gut microbiota to the production of indole-3-acetate.

(a) Schematic representation of the production of indole-3-acetate by three possible metabolic pathways, represented with orange (path 1), grey (path 2) or blue (path 3) arrows. (b) Heatmap showing the contribution of individual species to each the reactions involved in the production of indole-3-acetate. (c) Histogram showing the distribution of the score values. (d) Filtered heatmap selecting the 38 species with the highest contribution (score  $\geq 1.66$ ). \*Exclusively bacterial reactions, \*\*exclusively human reactions.

It should also be noted, however, that the scoring system proposed merely allows to identify the species worthy of further study *in vitro* and *in vivo*; however, it does not allow to make predictions as to which species make the greatest contribution to metabolism of indoles in the human gut.

## 4. Discussion

In this chapter, the contribution of the gut microbiota to the production of indoles was investigated. In particular, a systematic computational approach was used to identify the metabolic reactions and potential prokaryotic species involved in the production of indoles, as well as to predict members of the gut microbiota involved in these metabolic processes.

### 4.1. Gut prokaryotic metabolism controls the majority of the reactions leading to the production of indoles

First, the contribution of prokaryotes to the metabolic network leading to the production of indoles was investigated. Gut prokaryotic metabolism potentially contributes to 78% of the reactions of the metabolic network; however, the relevance of its contribution varies depending on the indolic compound considered. According to KEGG (and further validation by BLASTP-searches), indole-3-pyruvate is produced by human metabolism only, upon the action of the *L*-amino-acid oxidase; therefore, it should not be considered a microbiota-associated metabolite. *L*-amino-acid oxidase is a secreted enzyme encoded by the gene *IL4II* and is highly expressed in inflammatory cells, but also in the gastrointestinal tract, particularly in the colon and rectum (Ponten, et al., 2008). The enzyme is released into the intestinal lumen (Ponten, et al., 2008), where it can convert tryptophan into indole-3-pyruvate, which can then be used as a substrate for gut prokaryotic metabolism. Therefore, although indole-3-pyruvate is not a gut microbial metabolite *per se*, it can be used as a substrate for the production of microbial metabolites, exemplifying an interesting case of mammalian-microbial

metabolic cooperation. In this study, it was not possible to conduct BLASTP searches on the reactions involved in the production of indole-3-lactate and 3-methylindolepyruvate from indole-3-acetate, as these reactions are not linked with a KEGG ID orthology. However, according to the literature, these reactions have been identified, respectively, in *Clostridium sporogenes* (Jean and DeMoss, 1968) and *Streptomyces griseus* (Hornemann, et al., 1970). To our best knowledge, no genes have been assigned for these reactions.

Apart from indole-3-pyruvate, our results suggest that all indolic compounds included in this network involve prokaryotic metabolism, at least partially. Further results focus on the production of a) indole (exclusively prokaryotic metabolism) and b) indole-3-acetate (both shared by human and prokaryotic metabolism).

## **4.2. Indole is exclusively produced by prokaryotic metabolism: Proteobacteria and Bacteroidetes are the main contributors**

Using a combined KEGG + BLASTP strategy on the gut-specific database, a total of 61 prokaryotes species were identified/predicted to be involved in the production of indole. Most species were from the phyla Proteobacteria (n = 30) and Bacteroidetes (n = 17). Amongst the species identified, many have been previously reported as tryptophanase-containing, thus validating the identification strategy adopted in this study (*e.g.* *Bacteroides thetaiotaomicron*, (Tannock, 1977) *Escherichia coli* (Smith, 1897), *Escherichia fergusonii* (Farmer, et al., 1985), *Edwardsiella tarda* (Amandi, et al., 1982), *Haemophilus influenzae* (Stull, et al., 1995)). The BLASTP strategy predicted, based on protein homology, 23 species previously unknown to be involved in indole production. Interestingly, some of them (*Klebsiella oxytoca*, *Prevotella pallens* and *Prevotella nigriscens*) were recently confirmed as indole-producing species, encoding tryptophanase (*TnaA*) in their genomic DNA and/or producing detectable levels of indole (Alves, et al., 2006; Sasaki-Imamura, et al., 2011).

As for the other predicted species, it is important to note that the detection of a protein sequence in a microbe's genome is only an indicator that it may produce a

given metabolite from a substrate; it does not guarantee functionality and further *in vitro* studies are needed to confirm metabolic activity. Validation studies should include qualitative (*e.g.* Kovac's reagent or Ehrlich's reagent (Lombard and Dowell, 1983)) or quantitative assessment (*e.g.* Liquid Chromatography - Mass Spectrometry (LC-MS) (Badenoch-Jones, et al., 1982)) of indole in pure cultures of bacteria, after growth in sterile tryptophan-containing medium.

#### **4.3. Proteobacteria and Firmicutes are the main contributors to the production of indole-3-acetate**

The contribution of gut prokaryotes to the production of indole-3-acetate was also evaluated. Three distinct prokaryotic biosynthetic pathways leading to the production of indole-3-acetate were identified, based on KEGG. Amongst these, two have already been described (*via* indole-3-acetamide and indole-3-pyruvate) (Patten, et al., 2013).

The relative contributions of the microbiota and human metabolism varied between pathways. The production of indole-3-acetate *via* indole-3-acetamide is prokaryotic-specific (Path 1); *Streptomyces* spp. are amongst the species previously demonstrated to be involved in this conversion (Manulis, et al., 1994), but they are not commensals of the human gut. Conversely,, the production of indole-3-acetate *via* indole-3-pyruvate (Path 2) requires a host-specific enzyme, thus representing an example of host-microbiota co-metabolism, as suggested here and previously demonstrated by others (Sridharan, et al., 2014). Finally, all the reactions involved in the production of indole-3-acetate *via* tryptamine (Path 3) can be produced by both prokaryotic and human metabolism.

Using a combined KEGG + BLASTP strategy on the gut microbiota database, a total of 205 prokaryotes involved in the production of indole-3-acetate were identified / predicted; most species derived from the phyla Proteobacteria and Firmicutes. After implementing a scoring system to identify the species contributing to the highest number of steps of the overall metabolic network, the species with highest scores were

*Pseudomonas putida* and *Pseudomonas fluorescens*. Interestingly, both species were already proposed to be involved in the production of indole-3-acetate, in plant studies (Karnwal, 2009; Patten and Glick, 2002). Other high score species, such as *Klebsiella oxytoca* and *Klebsiella pneumoniae*, were also demonstrated to produce indole-3-acetate in plant-based studies (Celloto, et al., 2012; Sachdev, et al., 2009). Furthermore, amongst the high-scoring species, two other gut microbial species were predicted (*Pseudomonas nitroreducens* and *Raoultella terrigena*). According to our best knowledge, these species have not been previously reported or predicted to produce indole-3-acetate. As mentioned above, these predictions require further *in vitro* validations (quantification of indole-3-acetate in pure cultures of bacteria by LC-MS, after growth of the organisms in sterile tryptophan-containing medium (Lin, et al., 2015)).

It is not surprising that many studies targeting indole-3-acetate-producing bacteria are performed in plants; for more than 70 years, indole-3-acetate has been recognised as the main auxin plant hormone, playing a critical role in cell division, vascular development and tropic response to light and gravity (Zhao, 2012). However, a recent study pointed out several commonalities between plant and the human gut microbial ecosystems (Ramirez-Puebla, et al., 2013). In particular, as indole-3-acetate-producing species present in plants are also found in the human gut, they can produce indole-3-acetate to be putatively incorporated into human metabolic and signalling pathways.

## 5. Conclusion

Altogether, the results shown in this chapter highlight the contribution of the human gut-associated prokaryotes to the production of indoles, potentially performing 78% of the reactions included in this metabolic network. The metabolic pathways identified converged in the production of one of two indolic compounds: indoxyl sulfate (*via* indole) or indole-3-acetate (*via* indole-3-acetamine, indole-3-pyruvate or tryptamine). The results suggest that the contribution of the gut prokaryotes varies depending on the indolic compound considered: while indole-3-pyruvate is produced only *via* human metabolism, indole requires prokaryotic metabolism. For others, such as indole-3-acetate, both mammalian and microbial metabolism are involved.

In this chapter, a phylum-level overview of the species involved in the production of indoxyl sulfate (*via* indole that undergoes further sulfation in human liver) and indole-3-acetate was provided. While indole is potentially mostly produced by Proteobacteria and Bacteroidetes, the species showing higher scores for the production of indole-3-acetate belong to the phyla Proteobacteria and Firmicutes. Although the computational-based strategy used herein allowed the prediction of a wider range of species able to perform these metabolic reactions, the functional activity of these species needs to be confirmed by *in vitro* studies.

As a natural expansion of this work, this computational strategy could represent a useful approach for those interrogating the metabolism of other gut-derived microbial metabolites relevant to human health.







## Chapter 4

# Contribution of indoles to the metabolic signature of human obesity

## 1. Introduction

A growing body of evidence suggests that gut microbiota have a critical role in human obesity and metabolism. This effect results from a complex interaction that incorporates not only differences in prokaryotic species, but also in their metabolic functions and, consequently, in the levels of resulting gut microbial metabolites (Cani, 2014; Chilloux, et al., 2016). Various microbial metabolites have been demonstrated to play a role in obesity (Chilloux, et al., 2016), and SCFA have been a particularly studied group. While acetate exhibits anorexigenic properties by altering the hypothalamic expression of neuropeptides involved in appetite suppression (Frost, et al., 2014), propionate upregulates the release of appetite-suppressing gut hormones, such as glucagon-like peptide-1 and peptide YY (Psichas, et al., 2015). TMAO, a

choline-derived gut microbial metabolite, also limits impaired glucose tolerance in mice and in isolated pancreatic islets (Dumas, et al., 2017).

Indoles have also emerged as metabolites with a putative role in obesity. As explored in the last chapter, bacterial degradation of tryptophan converges into the production of two indolic compounds: indoxyl sulfate (*via* indole) or indole-3-acetate (*via* indole-3-acetamine, indole-3-pyruvate or tryptamine). Indoles are then absorbed in the colon, thus entering systemic circulation (Sender, et al., 2016). Metabonomic studies have consistently identified indoles in human biofluids (Bouatra, et al., 2013; Danaceau, et al., 2003; Duranton, et al., 2012; Elliott, et al., 2015).

In 2013, Chatelier et al. found that obese subjects have a lower bacterial gene count and exhibit significantly higher modules for the degradation of aromatic amino acids (*e.g.* tryptophan) (Le Chatelier, et al., 2013), therefore having a higher genomic potential for the production of tryptophan-derived metabolites. Recently, Zhang et al. reported for the first time an association between indoxyl sulfate and obesity, in a paediatric cohort (Zhang, et al., 2015). Obesity in children with Prader-Willi syndrome was associated not only with a higher production of indoxyl sulfate, but also with a higher abundance of the genes involved in the production of this metabolite (Zhang, et al., 2015). It is important to note, however, that these findings were reported in a context of genetic obesity, and cannot be directly extrapolated to non-syndromic obesity.

In this chapter, a  $^1\text{H}$  NMR-based metabolic profiling approach was used to explore the contribution of indoles to the metabolic signature of obesity. The specific aims of this chapter include:

- a) To investigate the association between obesity and indoles, in a non-syndromic adult cohort;
- b) To investigate the impact of putative confounders and/or mediators.

## **2. Materials and Methods**

### **2.1. Study participants**

The MetaHit study participants were randomly selected from the Inter99 study, a randomized intervention study conducted at the Research Centre for Prevention and Health in Glostrup, Denmark, between 1999 and 2006. Inclusion criteria included: BMI < 25 kg.m<sup>-2</sup> or BMI > 30 kg.m<sup>-2</sup>, no known gastrointestinal disease, no previously bariatric surgery, no medications known to affect the immune system, and no antibiotics 2 months before fecal sample collection. A detailed description of the study is provided in (clinicalTrials.gov: NCT00289237). The study was approved by the local Ethical Committees of the Capital Region of Denmark (HC-2008-017) and was in accordance with the principles of the Declaration of Helsinki. All individuals gave written informed consent before participating in the study.

### **2.2. Phenotyping, biochemical and dietary assessment**

Sample collection, phenotypical, biochemical and dietary assessments performed in Chapter 4 (MetaHit Study) were performed by collaborators from the MetaHIT consortium, as described in (Le Chatelier, et al., 2013). Participants were examined in the morning after an overnight fast. Height and weight were measured in participants without shoes and wearing light clothes. Waist circumference was recorded using a non-expandable measuring tape, and measured between the lower rib margin and the iliac crest. Percentage of whole body fat (%Fat) was assessed by DEXA and data analyses were conducted with the integrated software (Hologic Discovery A, Santax).

Blood samples were collected after an overnight fast. Plasma glucose was analysed by glucose oxidase method (Granutest, Merck) and HbA<sub>1c</sub> was measured using a TOSOH G7 by ion-exchange high performance liquid chromatography. Plasma total cholesterol, plasma HDL-cholesterol and triglycerides were measured using Vitros 5600 using reflect-spectrophotometrics. Plasma total free fatty acids were analysed using standard biochemical methods (Modular Evo). Homeostatic Model Assessment

for Insulin Resistance (HOMA-IR) was calculated as: (fasting plasma glucose ( $\text{mmol.L}^{-1}$ ) x fasting serum insulin ( $\text{mU.L}^{-1}$ ))/22.5. Serum insulin was measured using the AutoDELFIA insulin kit (Perkin-Elmer, Wallac). Serum adiponectin was analysed using a two-site-sandwich ELISA kit for measuring total human adiponectin (TECO). Serum leptin was measured using the Bio-Plex Pro diabetes assay. Plasma high-sensitivity CRP was analysed by a particle-enhanced immunoturbidmetric assay on Modular Evo using the CRPL3 kit (Roche). Serum IL-6 and serum TNF- $\alpha$  were analysed by Luminex using the Bio-Plex Pro cytokine assay (Bio-Rad). A detailed description of the biochemical measurements was published elsewhere (Le Chatelier, et al., 2013). Participants completed a self-administered 198-item Food Frequency Questionnaire, where they were asked to report their dietary intake during the previous month. A detailed description of the questionnaire and estimation of the dietary intake in the population was previously published (Lau C. et al, 2004).

### 2.3. Sample preparation for metabolic profiling

Urine samples were stored at  $-80^{\circ}\text{C}$  and prepared by using 540  $\mu\text{L}$  of urine mixed with 60  $\mu\text{L}$  of urine buffer solution (0.1 M phosphate buffer solution (10%  $^2\text{H}_2\text{O}/\text{H}_2\text{O}$  vol/vol for 100 mL), 18 g  $\text{NaH}_2\text{PO}_4$  in 80 mL  $\text{D}_2\text{O}$ , 100 mg trimethylsilylpropanoic acid (TSP), 13 mg  $\text{NaN}_3$ ; pH 7.4 with KOH).

### 2.4. Spectra acquisition and pre-processing

Standard  $^1\text{H}$  NMR spectra were acquired using a Bruker Advance III spectrometer (Bruker, Rheinstetten, Germany), operating at 600 MHz  $^1\text{H}$  frequency. The  $^1\text{H}$  NMR spectra were phased and baseline corrected by using in-house software and imported into Matlab (Mathwork, Inc) at high resolution. Each NMR spectrum was calibrated to the TSP signal at  $\delta$ -0.00, using the RSPA algorithm and normalised (probabilistic quotient normalisation) (Dieterle, et al., 2006). The water signal was removed at  $\delta$ -4.68-5.00.

## 2.5. Metabolite assignment

Metabolite assignment was performed using a combination of resources, including NMR Chenomx NMR suite 7.0 software, Human Metabolome Database (HMDB), Matlab and in-house databases (Nicholson, et al., 1995), STOCSY (Cloarec, et al., 2005) and spike-in experiment. The spike-in experiment for indoxyl sulfate was performed on the urine sample with the highest quantity of the metabolite, by successively adding increasing volumes of indoxyl sulfate (5  $\mu$ L of a stock solution, 3 M). Spectra were acquired using a Bruker Advance III spectrometer (Bruker, Rheinstetten, Germany), operating at 600 MHz  $^1$ H and processed as described in *Section 2.4. Spectra acquisition and pre-processing*.

## 2.6. Statistical analysis

For statistical comparisons between variables, total and partial correlations were performed, followed by correction for multiple testing using the Benjamini-Hochberg (BH) method. Spearman or Pearson correlations were used, respectively, for non-parametric and parametric analyses. Adjustment for co-variables was performed using partial correlations (further details are provided in individual figure captions). Linear regression models were performed using the least squares approach. Significance was established on a  $p$ -FDR<0.05 threshold of 0.05.

For metabonomic studies, common multivariate statistical approaches were used: PCA (Cloarec, et al., 2005; Gartland, et al., 1990; Gartland, et al., 1991; Holmes, et al., 1994) and OPLS-DA (Blaise, et al., 2009; Bylesjo, et al., 2006; Davidovic, et al., 2011). The intensity of the signal of indoxyl sulfate, expressed as the peak integral (7.5040-7.5090 ppm), was  $\log_{10}$ -transformed and used for quantification and investigation of putative associations with clinical variables. A statistical threshold of  $p$ -FDR<0.10 was adopted.

Statistical analysis was performed in R v.3.3.0 using the MWASTools package (Rodriguez-Martinez, et al., 2017), Matlab v.r2016a and GraphPad Prism v.5.

### 3. Results

#### 3.1. MetaHit cohort characterisation

The MetaHit cohort (Le Chatelier, et al., 2013) comprises 279 subjects (53.8% female), with an average age of  $38.80 \pm 7.38$  years. The cardiometabolic risk of the participants was assessed throughout an extensive characterisation of a) obesity-related parameters, b) metabolic homeostasis and c) inflammatory markers. A summary of these characteristics is shown in Table 3.

Table 3. Metabolic summary description of the MetaHit study population.

|                                           | Reference values | MetaHit |       |                  |
|-------------------------------------------|------------------|---------|-------|------------------|
|                                           |                  | Mean    | SD    | 95% CI           |
| Obesity parameters                        |                  |         |       |                  |
| Whole body fat (%Fat)                     | [12.00 – 28.00]  | 32.40   | 9.10  | [31.33 - 33.47]  |
| BMI (kg m <sup>-2</sup> )                 | [18.50 – 25.00]  | 29.60   | 6.03  | [28.89 - 30.31]  |
| Weight (kg)                               | n/a              | 87.77   | 19.77 | [85.44 - 90.10]  |
| Waist (cm)                                | n/a              | 99.97   | 16.39 | [98.04 - 101.90] |
| Metabolic Profile                         |                  |         |       |                  |
| Plasmatic glucose (mmol L <sup>-1</sup> ) | [4.0 – 6.0]      | 5.74    | 0.53  | [5.68-5.80]      |
| HbA1c (%)                                 | < 6.5            | 5.56    | 0.33  | [5.52-5.60]      |
| Insulin (pmol L <sup>-1</sup> )           | < 174            | 53.01   | 33.79 | [49.02-57.01]    |
| Total cholesterol (mmol L <sup>-1</sup> ) | [3.00 – 5.50]    | 5.52    | 0.95  | [5.41-5.63]      |
| HDL-cholesterol (mmol L <sup>-1</sup> )   | [1.04 – 2.07]    | 1.55    | 0.53  | [1.49-1.61]      |
| Triglycerides (mmol L <sup>-1</sup> )     | [0.56 – 1.69]    | 1.34    | 0.70  | [1.26-1.42]      |
| Leptin (µg L <sup>-1</sup> )              | n/a              | 21.64   | 29.06 | [18.20-25.07]    |
| Adiponectin (mg L <sup>-1</sup> )         | n/a              | 11.10   | 7.17  | [10.26-11.95]    |
| Inflammatory Profile                      |                  |         |       |                  |
| hsCRP <sup>I</sup> (mg L <sup>-1</sup> )  | < 5.0            | 2.65    | 2.82  | [2.317-2.986]    |
| Interleukin-6 (pg mL <sup>-1</sup> )      | n/a              | 19.97   | 21.63 | [17.41-22.53]    |
| TNF-α                                     | n/a              | 26.23   | 49.43 | [20.36-32.10]    |

BMI: Body Mass Index: defined as body mass (kg) divided by the square of the height (m<sup>2</sup>); HbA1c: Haemoglobin A1c, also referred as glycated haemoglobin; HDL-cholesterol: High-Density Lipoprotein cholesterol; TNF-α: Tumour Necrosis Factor-α. Reference values: %Fat (Hewitt, et al., 1993); BMI (WHO, 2016); fasting plasmatic glucose and HbA1c (IDF, 2005); total cholesterol, HDL-cholesterol, triglycerides and hsCRP (Medscape, 2014); for weight, waist, leptin, adiponectin, interleukin-6 and TNF-α, no consensual reference ranges are available up to the date (n/a).

A more detailed description on the relationship between obesity parameters and other clinical and analytical measures is shown as a correlation matrix (Figure 20). The correlation matrix revealed a strong hierarchical clustering between several measures of obesity (Figure 20). Amongst these measures, the percentage of whole body fat (%Fat) was the parameter showing the strongest correlation with plasma leptin ( $r = 0.852$ ,  $p$ -FDR<0.001), a direct marker of adipocyte size (hypertrophy) and number (hyperplasia) (Figure 20), thus emerging as the ideal parameter to measure obesity.

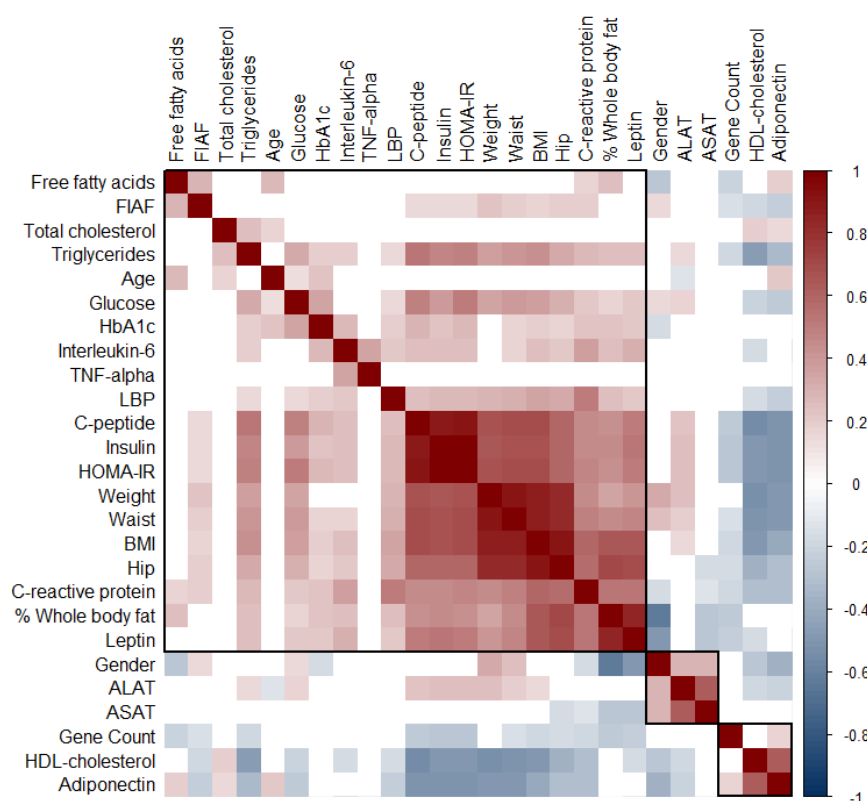


Figure 20. Correlation matrix between obesity parameter and other metabolic variables in the MetaHit study. Associations were computed using Spearman correlations. The  $p$ -values of the correlations were adjusted for multiple-testing using the BH criterion. Only significant correlations ( $p$ -FDR < 0.05) are shown with colour matching the direction and the strength of the correlation. ALAT: Alanine transaminase; ASAT: Alanine aminotransferase; BMI Body Mass Index; HbA1c: Haemoglobin A1c, also referred as glycated haemoglobin; HDL-cholesterol: High-Density Lipoprotein cholesterol; HOMA-IR: Homeostatic model assessment for insulin resistance; TNF- $\alpha$ : Tumour Necrosis Factor- $\alpha$ ; LPS: Lipopolysaccharide; FIAF: Fasting-Induced Adipose Factor.

Furthermore, %Fat was also the obesity measure exhibiting the highest correlation with HbA1c ( $r = 0.223$ ,  $p\text{-FDR} < 0.001$ ), the current gold-standard for management of type 2 diabetes mellitus (Figure 20, Table 4). In a non-adjusted model, %Fat also positively correlated with other markers of metabolic dysfunction, including HOMA-IR, triglycerides, free fatty acids, interleukin-6 and C-reactive protein. All correlations remained significant after adjustment for age and gender, except for circulating free fatty acids (Table 4).

Table 4. Spearman correlations between the percentage of body fat (%Fat) and other metabolic parameters

|                                        | Unadjusted |                         |                         | Age- and gender-adjusted |                         |                         |
|----------------------------------------|------------|-------------------------|-------------------------|--------------------------|-------------------------|-------------------------|
|                                        | $r$        | $p\text{-value}$        | $p\text{-FDR}$          | $r$                      | $p\text{-value}$        | $p\text{-FDR}$          |
| <b>Demographic</b>                     |            |                         |                         |                          |                         |                         |
| Age                                    | 0.009      | $8.90 \times 10^{-01*}$ | $9.06 \times 10^{-01}$  | -                        | -                       | -                       |
| Gender                                 | -0.637     | $2.27 \times 10^{-30*}$ | $2.02 \times 10^{-29*}$ | -                        | -                       | -                       |
| <b>Metabolic Profile</b>               |            |                         |                         |                          |                         |                         |
| Plasma glucose                         | 0.173      | $5.65 \times 10^{-03*}$ | $1.07 \times 10^{-02}$  | 0.348                    | $1.29 \times 10^{-08*}$ | $2.40 \times 10^{-08*}$ |
| HbA1c (%)                              | 0.223      | $3.33 \times 10^{-04*}$ | $7.67 \times 10^{-04*}$ | 0.140                    | $2.57 \times 10^{-02*}$ | $3.34 \times 10^{-02*}$ |
| Insulin (pmol l <sup>-1</sup> )        | 0.443      | $1.11 \times 10^{-13*}$ | $4.93 \times 10^{-13*}$ | 0.623                    | $1.51 \times 10^{-28*}$ | $4.38 \times 10^{-28*}$ |
| C-peptide                              | 0.425      | $1.32 \times 10^{-12*}$ | $5.70 \times 10^{-12*}$ | 0.607                    | $7.41 \times 10^{-27*}$ | $1.93 \times 10^{-26*}$ |
| HOMA-IR                                | 0.438      | $2.20 \times 10^{-13*}$ | $9.66 \times 10^{-13*}$ | 0.635                    | $5.25 \times 10^{-30*}$ | $1.71 \times 10^{-01*}$ |
| Total cholesterol                      | 0.055      | $3.83 \times 10^{-01}$  | $4.65 \times 10^{-01}$  | 0.034                    | $5.93 \times 10^{-01}$  | $5.93 \times 10^{-01}$  |
| HDL-cholesterol                        | -0.123     | $4.93 \times 10^{-02*}$ | $7.72 \times 10^{-02}$  | -0.412                   | $8.45 \times 10^{-12*}$ | $1.83 \times 10^{-11*}$ |
| Triglycerides (mmol l <sup>-1</sup> )  | 0.256      | $2.67 \times 10^{-05*}$ | $7.33 \times 10^{-05*}$ | 0.369                    | $1.43 \times 10^{-09*}$ | $2.85 \times 10^{-09*}$ |
| FFA (mmol l <sup>-1</sup> )            | 0.246      | $7.40 \times 10^{-05*}$ | $1.87 \times 10^{-04*}$ | 0.083                    | $1.86 \times 10^{-01}$  | $2.10 \times 10^{-01}$  |
| Leptin (μg l <sup>-1</sup> )           | 0.852      | $3.89 \times 10^{-73*}$ | $6.26 \times 10^{-72*}$ | 0.812                    | $1.23 \times 10^{-60*}$ | $7.97 \times 10^{-60*}$ |
| Adiponectin (mg l <sup>-1</sup> )      | 0.018      | $7.75 \times 10^{-01}$  | $8.14 \times 10^{-01}$  | -0.327                   | $9.91 \times 10^{-08*}$ | $1.72 \times 10^{-07*}$ |
| <b>Inflammatory Profile</b>            |            |                         |                         |                          |                         |                         |
| CRP <sup>f</sup> (mg l <sup>-1</sup> ) | 0.524      | $2.10 \times 10^{-19*}$ | $1.48 \times 10^{-18*}$ | 0.554                    | $8.76 \times 10^{-22*}$ | $2.07 \times 10^{-21*}$ |
| Interleukin-6                          | 0.257      | $3.33 \times 10^{-05*}$ | $9.09 \times 10^{-05*}$ | 0.227                    | $2.71 \times 10^{-04*}$ | $4.14 \times 10^{-04*}$ |
| TNF-α                                  | 0.045      | $4.78 \times 10^{-01}$  | $5.63 \times 10^{-01}$  | 0.045                    | $4.78 \times 10^{-01}$  | $4.97 \times 10^{-01}$  |
| FIAF (μg l <sup>-1</sup> )             | 0.077      | $2.19 \times 10^{-01}$  | $2.92 \times 10^{-01}$  | 0.220                    | $4.24 \times 10^{-04*}$ | $6.12 \times 10^{-04*}$ |

Associations were computed using Spearman correlations. The p-values of the correlations were adjusted for multiple-testing using the BH criterion. Only significant correlations ( $p\text{FDR} < 0.05$ ) are shown with colour matching the direction and the strength of the correlation. ALAT: Alanine transaminase; ASAT: Alanine aminotransferase; BMI: Body Mass Index; FFA: Free Fatty Acids; HbA1c: Haemoglobin A1c, also referred as glycated haemoglobin; HDL-cholesterol: High-Density Lipoprotein cholesterol; HOMA-IR: Homeostatic model assessment for insulin resistance; TNF-α: Tumour Necrosis Factor- ; LPS: Lipopolysaccharide; FIAF: Fasting-Induced Adipose Factor.

### 3.2. Identifying the clinical phenotype to investigate the association between obesity and gut microbial changes

Le Chatelier et al. have previously found that obesity, insulin resistance and dyslipidaemia are associated with lower bacterial richness (*i.e.* lower number of

bacterial genes) (Le Chatelier, et al., 2013). In this chapter, the relationship between bacterial gene count and various measures of obesity was explored. Bacterial gene count was significantly anti-correlated with all anthropometric measures of obesity, except with body weight (Figure 21). Importantly, %Fat exhibited the most significant correlation with gene count in both unadjusted ( $r = -0.242$ ,  $p\text{-FDR} < 2.36 \times 10^{-4}$ ) and age- and gender- adjusted models ( $r = -0.213$ ,  $p\text{-FDR} < 1.28 \times 10^{-3}$ ) (Figure 21 and Table 5).

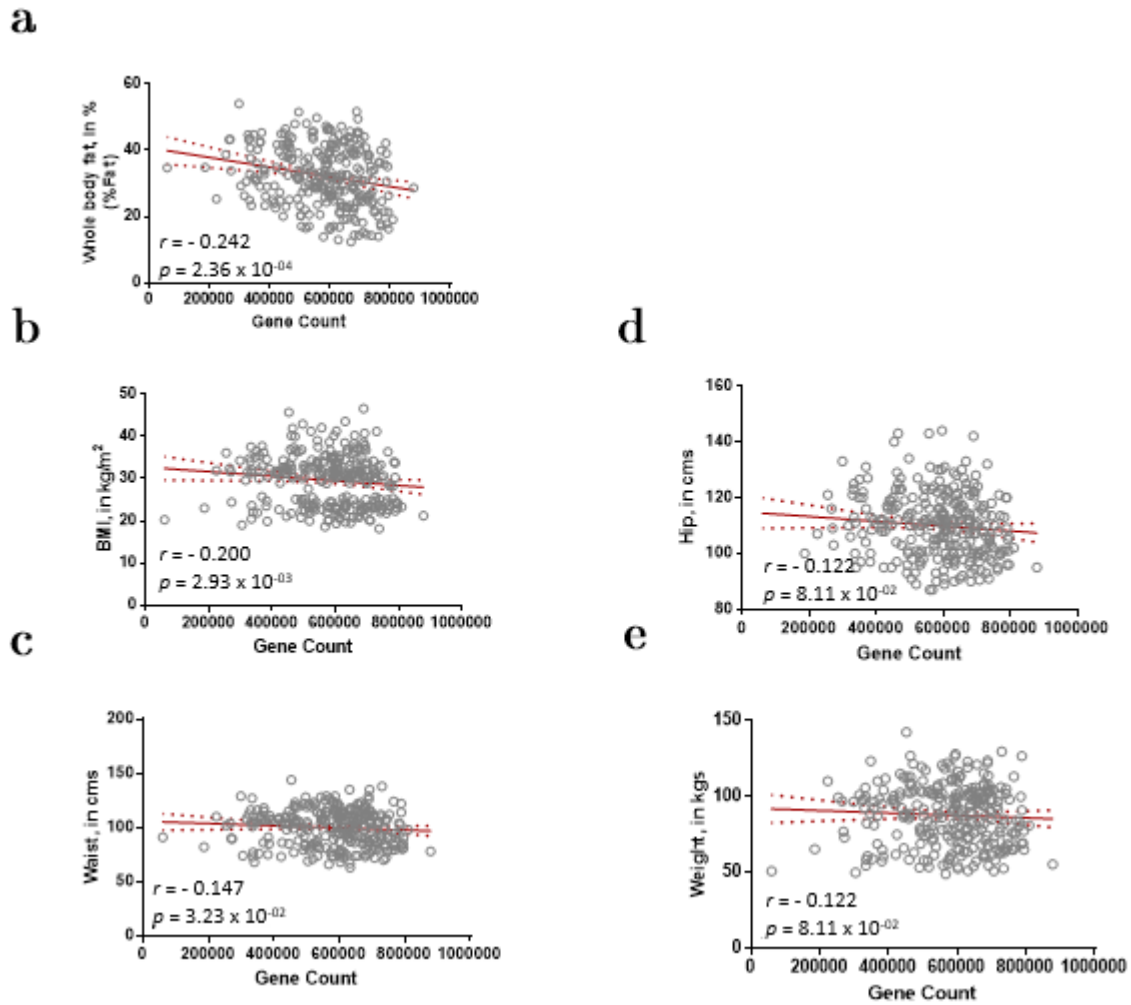


Figure 21. %Fat exhibits the strongest anti-correlation with bacterial gene count. a-e) Associations were computed using Spearman correlations.  $p$  values were corrected for multiple testing using the Benjamini-Hochberg (BH) criterion. Trend lines were calculated by linear regression using the least squares approach. A statistical threshold of 0.05 was adopted. BMI: Body Mass Index.

Table 5. Spearman correlations between obesity measures and gene count

|                           | Gene Count |                          |                          |                          |
|---------------------------|------------|--------------------------|--------------------------|--------------------------|
|                           | Unadjusted |                          | Age- and gender-adjusted |                          |
|                           | <i>r</i>   | <i>p</i> -FDR            | <i>r</i>                 | <i>p</i> -FDR            |
| %Fat (%)                  | -0.242     | 2.36x10 <sup>-04</sup> * | -0.213                   | 1.28x10 <sup>-03</sup> * |
| Weight (kg)               | -0.122     | 8.11x10 <sup>-02</sup>   | -0.140                   | 1.13x10 <sup>-02</sup> * |
| Waist (cm)                | -0.147     | 3.23x10 <sup>-02</sup> * | -0.191                   | 4.12x10 <sup>-03</sup> * |
| BMI (kg m <sup>-2</sup> ) | -0.200     | 2.93x10 <sup>-03</sup> * | -0.220                   | 2.81x10 <sup>-03</sup> * |
| Hip (cm)                  | -0.168     | 1.30x10 <sup>-02</sup> * | -0.155                   | 2.20x10 <sup>-02</sup> * |

Total (unadjusted) and partial (age- and gender-adjusted) Spearman correlations were calculated, followed by correction for multiple testing using the BH method. A statistical threshold of 0.05 was adopted. %Fat: whole body fat (%); BMI: Body Mass Index.

As expected, %Fat exhibits the strongest correlation with leptin (a direct marker of increased adipocyte number and size) and the strongest anti-correlation with gut microbial dysfunction, and therefore was chosen to further investigate the contribution of microbial metabolites in subjects with increased adiposity.

### 3.3. Investigating the metabolic profiles of patients with increased adiposity (%Fat) by <sup>1</sup>H NMR

A <sup>1</sup>H NMR metabonomic approach was used to investigate the metabolic signature of patients with increased adiposity and, specifically, to assess the contribution of indoles. Urinary metabolic profiles were acquired using a standard 1D experiment with water pre-saturation; the chemical shifts of the most abundant metabolites assigned are presented in Table 6. <sup>1</sup>H NMR profiling allowed the identification of urinary metabolites related to a wide range of both human and microbial-derived products. Concerning human metabolism, it was possible to identify metabolites involved in branched-chain amino acid and energy metabolism (*e.g.* valine (M1); citrate (M9) and lactate (M4)) and creatinine biosynthesis and skeletal muscle metabolism (*e.g.* creatinine (M11), glycine (M14), taurine (M13), creatine (M15), histidine (M21), alanine (M5)) (Table 6).

Regarding gut bacterial metabolites, <sup>1</sup>H NMR profiling identified products of three different host-microbial co-pathways. The first includes the bacterial conversion

of dietary choline into methylamines (*e.g.* trimethylamine (TMA, M10), trimethylamine-*N*-oxide (TMAO, M12) and dimethylamine (DMA, M9)). The second co-metabolic pathway involves the production of hippurate (M20) - produced from dietary phenolics, aromatic amino acids and other related compounds upon bacterial action and further conjugated with glycine in the liver (Lees, et al., 2013). Metabolites from the third pathway, involved in the distal colonic protein putrefaction and metabolism of aromatic amino acids, include *N*-PAG (M7, produced from phenylalanine) (Holmes, et al., 2011; Russell, et al., 2013)) and indoxyl sulfate (produced from tryptophan, as previously described in Chapters 1 and 3). Apart from indoxyl sulfate (M19); no other potential indolic compounds were unambiguously assigned.

Table 6. List of most abundant urinary metabolites assigned and respective chemical shift

| Metabolite key | Metabolite name                 | Chemical shift ( $\delta$ $^1\text{H}$ ppm) |
|----------------|---------------------------------|---------------------------------------------|
| M1             | Valine                          | 0.98 (d), 1.03 (d)                          |
| M2             | 3-Hydroxybutyrate               | 1.19 (d), 2.30 (m)                          |
| M3             | Lactate                         | 1.32 (d), 4.11 (q)                          |
| M4             | 3-Hydroxy-3-methylglutarate     | 1.30 (s), 2.40 (m)                          |
| M5             | Alanine                         | 1.47 (d)                                    |
| M6             | Acetate                         | 1.92 (s)                                    |
| M7             | <i>N</i> -phenylacetylglutamine | 7.36 (m), 7.40 (m)                          |
| M8             | Citrate                         | 2.53 (d), 2.67 (d)                          |
| M9             | Dimethylamine                   | 2.71 (s)                                    |
| M10            | Trimethylamine                  | 2.87 (s)                                    |
| M11            | Creatinine                      | 3.05 (s), 4.07 (s)                          |
| M12            | Trimethylamine- <i>N</i> -Oxide | 3.26 (s)                                    |
| M13            | Taurine                         | 3.27 (t), 3.43 (t)                          |
| M14            | Glycine                         | 3.57 (s)                                    |
| M15            | Creatine                        | 3.93 (s)                                    |
| M16            | Trigonelline                    | 4.43 (s), 8.07 (t), 8.83 (t)                |
| M17            | Urea                            | 5.8 (s)                                     |
| M18            | Tyrosine                        | 6.89 (d), 7.18 (d)                          |
| M19            | Indoxyl sulfate                 | 7.51 (d), 7.71 (d)                          |
| M20            | Hippurate                       | 7.56 (t), 7.65 (t), 7.84 (t)                |
| M21            | Histidine                       | 7.08 (s), 7.90 (s)                          |
| M22            | Formate                         | 8.44 (s)                                    |

(s) singlet; d (doublet); (t) triplet; (m) multiplicities

A representative spectrum of a urine sample with the most abundant metabolites assigned is shown in Figure 22.

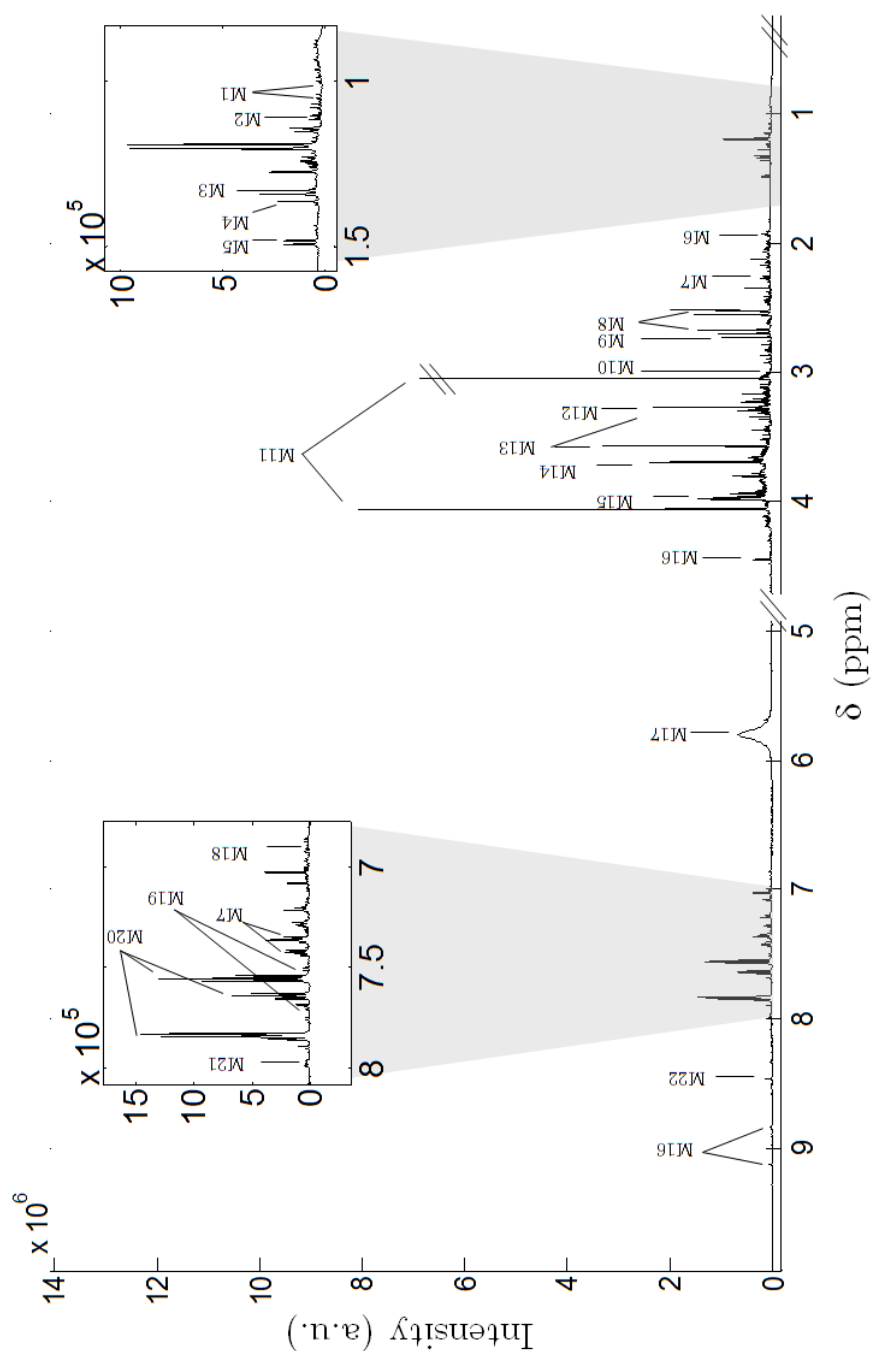


Figure 22. A typical urine  $^1\text{H}$  NMR spectrum (standard 1D spectrum with water pre-saturation). Most abundant metabolites are numbered as follows: M1: valine; M2: 3-hydroxybutyrate; M3: lactate; M4: 3-hydroxy-3-methylglutarate; M5: alanine; M6: acetate; M7: *N*-phenylacetylglutamine; M8: citrate; M9: dimethylamine; M10: trimethylamine; M11: creatinine; M12: trimethylamine-*N*-oxide; M13: taurine; M14: glycine; M15: creatine; M16: trigonelline; M17: urea; M18: tyrosine; M19: indoxyl sulfate; M20: hippurate; M21: histidine; M22: formate. a.u.: arbitrary units; ppm: parts per million.

In order to explore the overall differences on the metabolic signature of patients with increased adiposity (%Fat), both unsupervised and supervised methods were used. Although no clear discrimination was observed in a PCA model including the metabolic profiles of individual on extreme quartiles of %Fat (5 Principal Components, mean-centred model, (Figure 23a-b), a relatively good separation between groups was observed in an OPLS-DA model (5 Orthogonal Components,  $R^2=0.9335$ ,  $Q^2=0.067$ , Figure 23c-d). Even though the  $Q^2/R^2$  ratio is low (and, therefore, overfitting cannot be excluded), the separation between groups observed in this model suggests that individuals with increased adiposity exhibit a specific metabolic signature.

In order to better explore the metabolic signature of individuals with increased adiposity, the total correlations between %Fat and  $^1\text{H}$  NMR metabolites were computed and visualised as a skyline plot (Figure 23e). In the spectral region between  $\delta 7.00$ – $8.00$ , there were observed several signals positively associated with %Fat, which were putatively identified as indoxyl sulfate (Figure 23e).

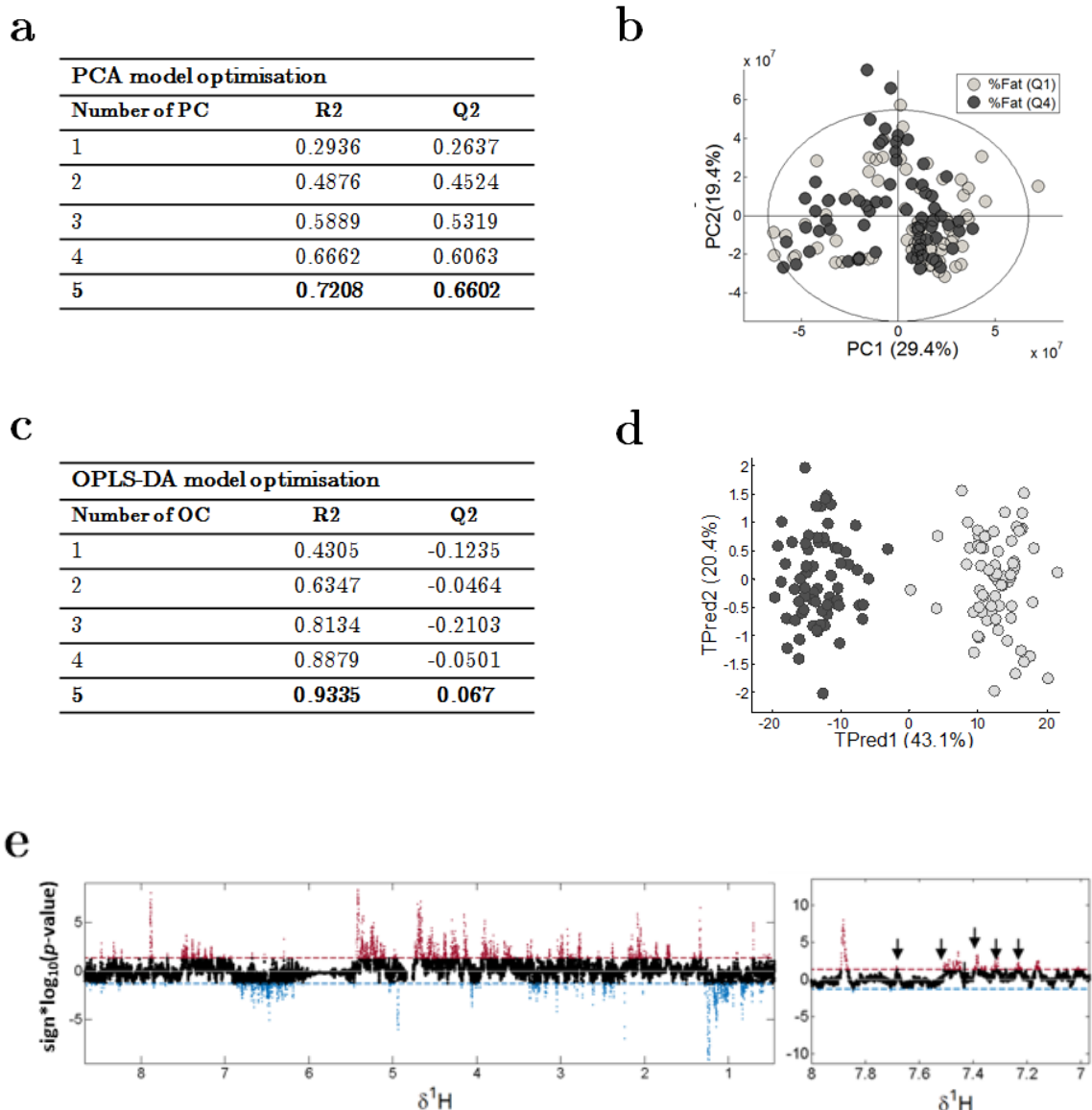


Figure 23. Metabonomic signature of increased adiposity: Principal Component Analysis (PCA), Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) and skyline plot. a) Optimisation of the Principal Component Analysis model. b) PCA scores plot (5 PC, mean-centred model). Optimisation (c) and scores plot (d) of the OPLS-DA model. e) Visualization of the associations of %Fat with urine  $^1\text{H}$  NMR metabolites. The associations were computed using total Spearman correlations. Total and partial ( $\delta$  7.00–8.00) skyline plots show the  $-\log_{10}(p\text{-val}) \times \text{sign}$  of beta coefficient of each NMR signal. Statistically significant signals positively associated with %Fat were coloured in red, and significant signals negatively associated with %Fat were coloured in blue. The arrows indicate signals putatively assigned as indoxyl sulfate.

### 3.4. Exploring the relationship between obesity and indoxyl sulfate

The assignment of indoxyl sulfate was confirmed by acquiring a standard 1D spectrum and by performing a spike-in experiment (Figure 24).

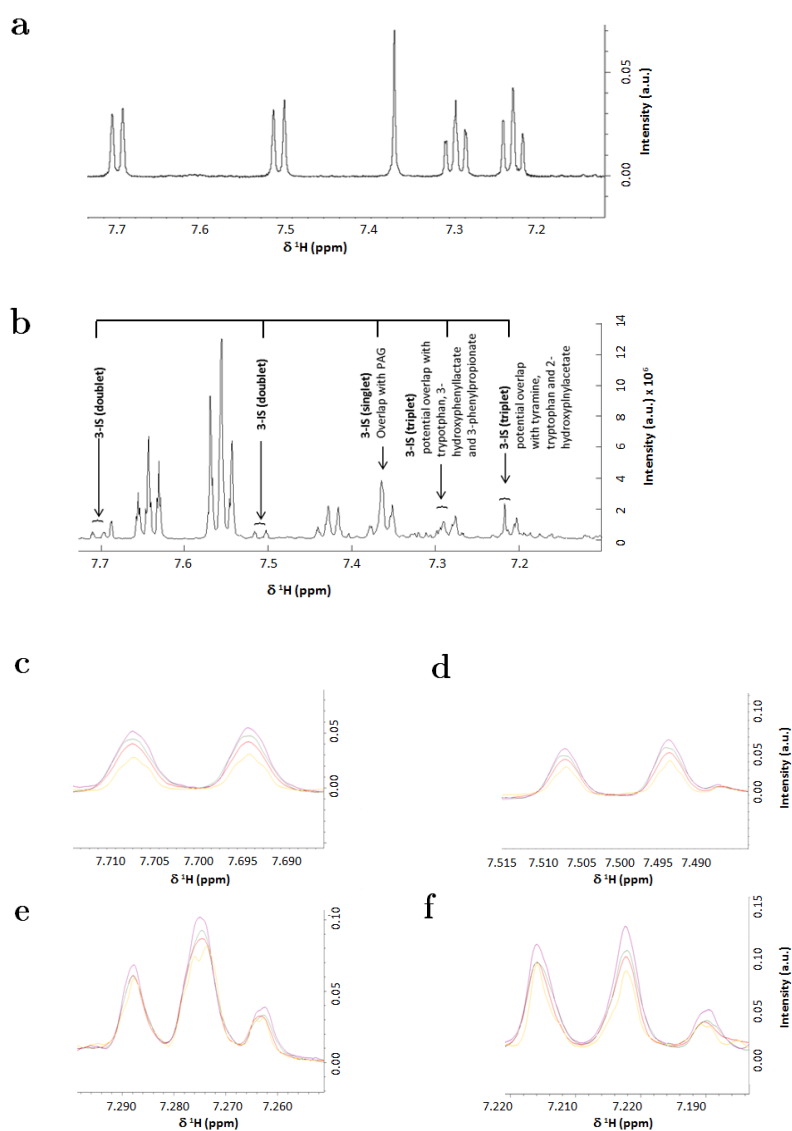


Figure 24. Confirmation of the assignment of indoxyl sulfate by a spike-in experiment. a) Standard spectrum of indoxyl sulfate. b) Detailed view (aromatic region) of a urine sample, highlighting the assignment of indoxyl sulfate and potential overlap with other metabolites. c-f) Three cumulative additions of 5  $\mu$ L of a stock solution of indoxyl sulfate (baseline: yellow; 5  $\mu$ L: red; 10  $\mu$ L: green; 5  $\mu$ L: purple) resulted in a progressive increase of the intensity of the metabolite's signal. a.u.: arbitrary units; 3-IS: indoxyl sulfate; a.u.: arbitrary units; ppm: parts per million.

Comparing the standard spectrum with a urine sample, it was possible to identify the optimal signals for quantification, for which no overlap with other compounds was observed (doublets at 7.70 and 7.50 ppm) (Figure 24a-b). The spike-in experiment resulted in a progressive increase of the intensity of the metabolite's signal (Figure 24c-f), thus unambiguously confirming its assignment.

In order to further investigate the association between indoxyl sulfate and adiposity (%Fat), correlations with both anthropometric and analytical markers of obesity (*e.g.* leptin and adiponectin) were performed. For this targeted analysis, the signal of indoxyl sulfate at 7.50 (doublet) was integrated as the area under the curve (AUC) and  $\log_{10}$ -transformed (Figure 25a). Indoxyl sulfate positively correlated with %Fat ( $r = 0.1402$ ,  $p$ -adaptiveSTS = 0.06). No other significant correlations were observed between indoxyl sulfate and other obesity-related anthropometric parameters (Figure 25b). Correlations with other obesity-related analytical markers were also addressed; interestingly, indoxyl sulfate also correlated with increased plasmatic levels of leptin ( $r = 0.1576$ ,  $p$ -adaptiveSTS = 0.06), a marker for adipocyte number and size (Figure 25b). An overview of distribution of individual data points (and respective logistic regression models) for the associations between indoxyl sulfate and adiposity (%Fat), and indoxyl sulfate and leptin concentration, are shown in Figure 25c-d. The stability of the significant associations was confirmed by bootstrapping (resampling with replacement) with 10,000 iterations. The 95% confidence intervals (95%CI) for the correlations between indoxyl sulfate and %Fat and leptin were, respectively, [0.0322, 0.2419] and [0.0341, 0.3155].

### 3.5. Assessing the effect of potential confounders, mediators and co-variables

In order to identify potential confounders, mediators and co-variables implicated in the associations of indoxyl sulfate with obesity-related parameters, Pearson correlations were performed between indoxyl sulfate intensities and several demographic, metabolic and inflammatory variables; however, no significant associations were found (Figure 25e).

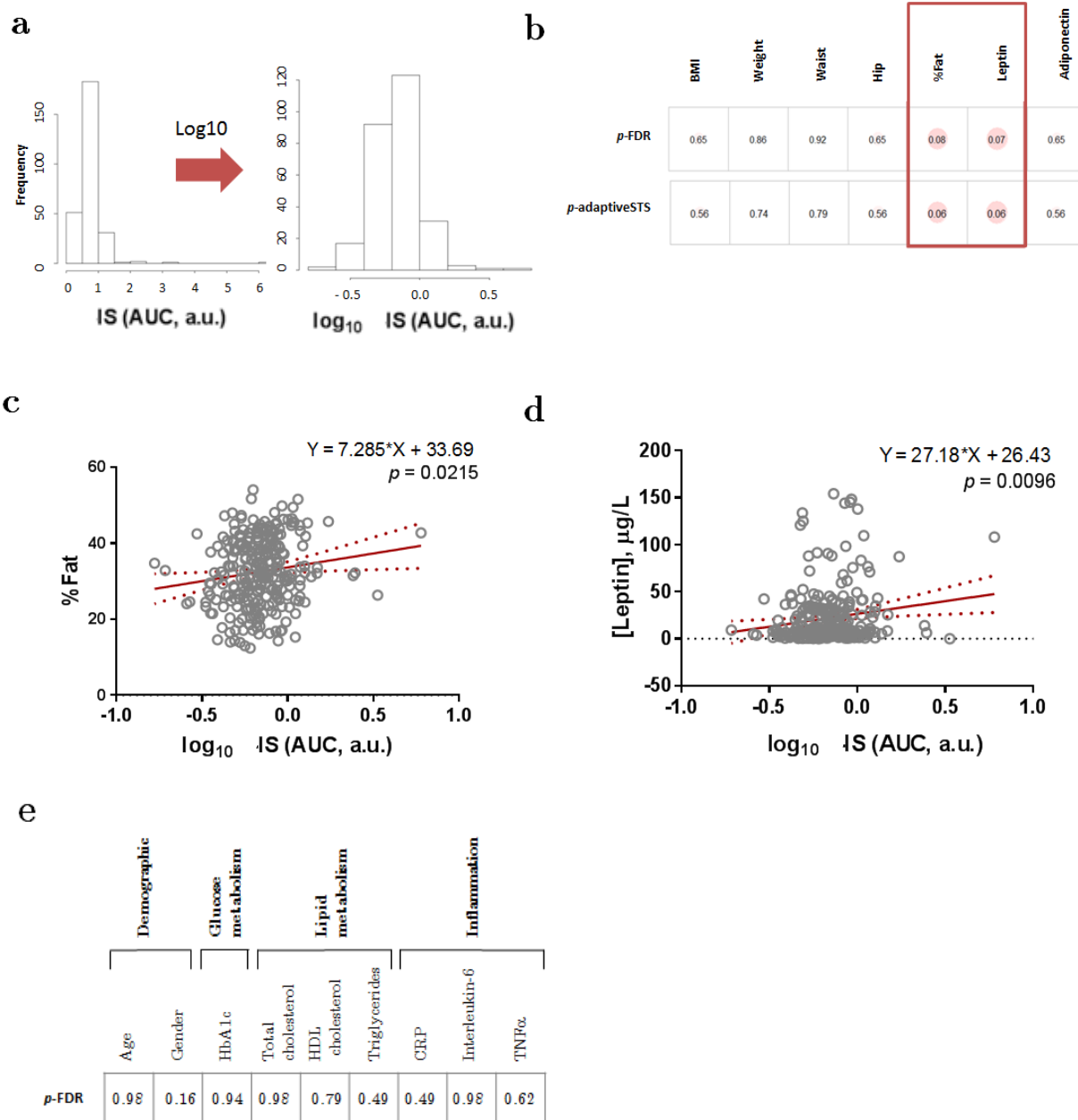


Figure 25. Correlations between indoxyl sulfate and other obesity-related parameters. a) The intensity of indoxyl sulfate intensity, expressed as the area under the curve (7.5040-7.5090 ppm), was log<sub>10</sub>-transformed; histograms represent the frequency distribution before and after normalisation; b) Pearson correlations were performed; Benjamini-Hochberg (FDR) and adaptive Storey-Taylor-Sigmund (adaptiveSTS) corrections for multiple testing were applied as mentioned. Estimates (*r*) are colour-coded and plotted in red (if positive) or blue (if negative); *p*-values are shown. c-d) Distribution of individual data points and linear regression models for indoxyl sulfate and %Fat (c) and leptin (d). e) Potential confounders were assessed by performing Pearson correlations, BH-adjusted for multiple testing. Statistical analysis was performed in R v.3.3.0 using the MWAStools package. a.u.: arbitrary units; AUC: Area Under the Curve; BMI: Body Mass Index; CRP: C-Reactive Protein; HbA1c: Haemoglobin A1c; IS: Indoxyl Sulphate; TNF $\alpha$ : Tumour Necrosis Factor- $\alpha$ .

Although indoxyl sulfate is positively associated with increased adiposity (%Fat), these findings do not exclude, *per se*, the possibility of a direct dietary effect. Being a product of the bacterial degradation of tryptophan, increased levels of indoxyl sulfate in obese subjects could be explained by an increased consumption of this amino acid in these subjects. However, no significant associations were observed between indoxyl sulfate and tryptophan, nor with any other dietary compounds (Figure 26a-b).

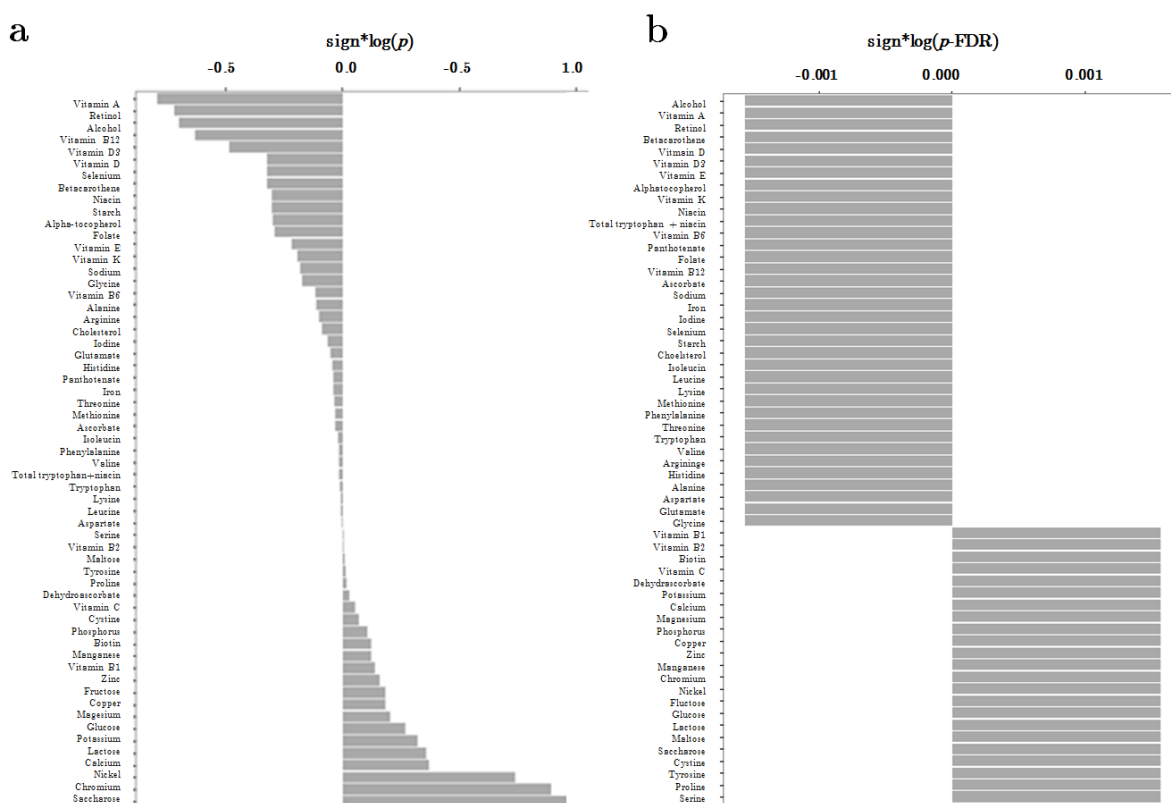


Figure 26. Total correlations between indoxyl sulfate and dietary compounds.

The intensity of indoxyl sulfate, expressed as the area under the curve (7.5040-7.5090 ppm), was log<sub>10</sub>-transformed; Pearson correlations, both uncorrected (a) and BH-corrected (b) were performed. *p* values are plotted as sign\*log(*p*), where positive bars represent positive estimates (*r*) and negative bars represent negative estimates. Bars (r) are colour-coded and plotted in grey (not-significant), red (significant, positive estimates) or blue (significant, negative estimates).

## 4. Discussion

In this chapter, the contribution of indoles to the metabolic signature of obesity was investigated. First, the cohort was characterised and relationships between obesity and other variables were assessed, in order to identify the best obesity measurement to use for further correlative analyses. Next, a  $^1\text{H}$  NMR-based strategy was used to profile indoles in urine samples, and the association between 3-indoxysulfate and adiposity was investigated. Finally, the impact of putative co-variables, confounders or mediators was explored.

### 4.1. Whole body fat (%Fat) exhibits the most significant correlation with leptin and bacterial gene count

Amongst all obesity measures these measures, %Fat was the parameter showing the strongest correlation with plasmatic leptin ( $r = 0.852$ ,  $p\text{-FDR} < 0.001$ ). Leptin is a product of the obese (*ob*) gene and is produced primarily in the adipocytes of white adipose (Green, et al., 1995). Although the primary function of leptin is to regulate energy balance by inducing satiety, its levels are increased in obesity, probably due to decreased sensitivity (leptin resistance) (Pan, et al., 2014). Leptin levels are known to be increased in obesity and to vary exponentially with the amount the fat mass; therefore, leptin levels have been proposed as direct marker of adipocyte size and number (Caro, et al., 1996).

The strongest correlation between %Fat and leptin, can be explained by the higher accuracy of this measure to quantify obesity. Whilst the quantification of obesity by DEXA provides a direct quantification, other anthropometric models, such as BMI, are surrogate markers based on mathematical predictive models (Shah and Braverman, 2012; Wells and Fewtrell, 2006). Furthermore, the accuracy of BMI and weight measurements can be compromised by the contribution of other body components as muscle mass or bone mass (Donini, et al., 2013).

As described by Le Chatelier et al, gene count is anti-correlated with obesity (Le Chatelier, et al., 2013). Importantly, amongst all obesity measures, %Fat also exhibits the most significant statistical association with gene count. Again, this stronger association might be explained by the higher accuracy of this measure and, therefore, the higher statistical power regarding the detection of associations with markers of gut microbial dysfunction. Altogether, these results support the use of %Fat, whenever available, as the optimal measure to perform correlative analyses aiming to investigate the impact of the gut microbiota in obesity.

## 4.2. %Fat is positively correlated with indoxyl sulfate

According to Le Chatelier et al., obese subjects exhibit not only a significantly lower gene count, but also a higher genomic potential for the production of tryptophan-derived metabolites (Le Chatelier, et al., 2013). In this chapter, metabonomic profiling was used to investigate the putative association obesity (expressed as %Fat) and indoles.

$^1\text{H}$  NMR has been extensively used for metabolic profiling due to its high reproducibility and easy sample preparation; in particular, it is a widely validated approach for the study of the microbial-mammalian interaction in the development of obesity (Dumas and Davidovic, 2013). Consistent with previous studies (Calvani, et al., 2010; Elliott, et al., 2015), we were able to identify a range of mammalian products in urine metabolic profiles, including metabolites involved in branched-chain amino acid and energy metabolism (*e.g.* valine, citrate, lactate) (Giskeodegard, et al., 2015) and in biosynthesis and skeletal muscle metabolism (*e.g.* creatinine, glycine, taurine, alanine) (Elliott, et al., 2015). As in previous studies, a range of gut microbial metabolites were also identified, including methylamines, hippurate and products of aromatic amino acids degradation (*e.g.* *N*-PAG) (Dumas, et al., 2006; Elliott, et al., 2015; Giskeodegard, et al., 2015).

In this study, the only indolic compound identified in urine metabolic profiles was indoxyl sulfate. Indoxyl sulfate has been repeatedly identified in  $^1\text{H}$  NMR

metabonomic studies, mostly in human urine (Bouatra, et al., 2013; Danaceau, et al., 2003; Duranton, et al., 2012; Elliott, et al., 2015; Giskeodegard, et al., 2015). However it is the first time that a positive association is reported between this metabolite and obesity (expressed as %Fat), in the context of non-genomic obesity. These results are consistent with a recently published study by Zhang et al., reporting a positive association between indoxyl sulfate and obesity, in the context of genetic obesity (*i.e.* Prader-Willi syndrome). Children diagnosed with Prader-Willi exhibited higher levels of indoxyl sulfate and increased abundance of genes involved on its production (Zhang, et al., 2015). Nonetheless, Prader-Willi syndrome is a very specific context characterised by an epigenetic phenomenon of imprinting (*i.e.*, the deletion of paternal copies of the *SNRPN* and *necdin* genes along with clusters of small nucleolar RNAs) and therefore could not be generalised. By replicating the same findings in the MetaHit cohort, we demonstrate that the association between indoxyl sulfate and obesity is also present in non-genetic obesity.

Indoxyl sulfate was previously identified by  $^1\text{H}$  NMR in the Intercontinental Study in Macronutrients in Arterial Pressure (Elliott, et al., 2015); however, no significant association was observed between its levels and BMI. As previously discussed, BMI is a surrogate marker, and its low accuracy might compromised the detection of associations with gut microbial metabolites.

### **4.3. Levels of indoxyl sulfate are not correlated with tryptophan dietary intake**

The identification of confounding factors is vital to identify robust disease markers in clinical association studies (Vandeputte, et al., 2016). Amongst those, history of antibiotic treatment (Cho, et al., 2012), inflammation status (Le Chatelier, et al., 2013) (Hildebrand, et al., 2013) and diet (David, et al., 2014; Rezzi, et al., 2007; Stella, et al., 2006) are thought to be particularly relevant. For this study, only participants without previous antibiotic treatment 2 months before faecal sample collection were included, thus excluding the need to explore previous antibiotic intake as a potential confounder.

No associations were found between indoxyl sulfate and inflammatory markers. Previous studies found that indoxyl sulfate increases the inflammatory response to LPS *in vitro*, by increasing TNF- $\alpha$  and IL-6 (Adesso, et al., 2013). However, it is important to note that these studies used very high concentrations, similar to the ones detected in uremic patients. Another study from (Brito, et al., 2016) also found a positive correlation between IL-6 levels and indoxyl sulfate, but again, participants included only patients with severe chronic renal disease. No studies validated similar findings in non-uremic patients. In this cohort, the absence of significant associations between the level of indoxyl sulfate and inflammatory cytokines suggests that the compound does not impact the pro-inflammatory status in the context of non-uremic patients. Therefore, it is not likely that the association between this gut microbial metabolite is obesity is mediated by a pro-inflammatory effect.

Similarly, no significant correlations were found between tryptophan intake and indoxyl sulfate urinary levels. One of the main concerns regarding novel associations between obesity and gut microbial markers is whether the variations on these compounds reflect changes in the gut microbial metabolism, or simply reflect an increased consumption of the respective substrate (in this case, tryptophan). In line with the findings of Brito et al (Brito, et al., 2016), no significant correlation were found between indoxyl sulfate and tryptophan consumption (or any other nutrients), thus excluding a potential dietary confounding effect.

## 5. Conclusion

In this chapter, indoxyl sulfate was identified as part of the metabolic signature of individuals with increased adiposity. Indoxyl sulfate correlated not only with whole body fat, but also with the plasmatic levels of leptin, which are a proxy for adipocyte size and number. Being physiologically plausible, the co-identification of both correlations strengthens a putative role of this metabolite in the context of obesity. These results are consistent with previous findings in the context of paediatric genetic obesity, thus suggesting that the association between indoxyl sulfate and obesity is not specific of a specific syndromic effect.

Of note, as in other studies, no significant association was observed between indoxyl sulfate metabolite and BMI, a common but less accurate proxy for human obesity. The strength of the association between body fat (%Fat) and gut microbial changes (*i.e.* gene count or the production of this specific metabolite), supports its use as an optimal variable for studies aiming to investigate the impact of the microbial-mammalian axis in human obesity.

It is important to note that this significant association, is not, however, sufficient to generate causal inferences. In the next chapter, the effect of indoxyl sulfate will be tested *in vitro* and its targets and signalling pathways explored by a combined strategy including a kinase screening and network analysis.







## Chapter 5

# Validation of the pro-adipogenic effect of indoxyl sulfate *in vitro*

### 1. Introduction

In the previous chapter, a  $^1\text{H}$  NMR-based metabolic profiling strategy showed that indoxyl sulfate is positively correlated with adiposity (%Fat) and plasma levels of leptin (Margetic, et al., 2002). Although indoxyl sulfate has been previously associated with several cardiovascular effects (Adesso, et al., 2013; Barreto, et al., 2009; Dou, et al., 2007; Faure, et al., 2006; Muteliefu, et al., 2009; Yisireyli, et al., 2013), its association with increased adiposity only recently started being explored. In particular, the cellular responses, mechanism and targets involved in the potential pro-adipogenic effect of indoxyl sulfate remain unclear.

In order to understand the effect of a given pro-adipogenic compound, it is important to revisit the biomolecular basis of obesity. Two major mechanisms contribute to the expansion of the adipose tissue: a) increased adipocyte cell number (hypertrophy) and b) increased adipocyte cell size (hyperplasia) (Jo, et al., 2009) (Figure 27). Adipocyte hyperplasia results of increased lipid intracellular accumulation of lipids, a consequence if the unbalance between lipid accumulation (lipogenesis) and lipid breakdown (lipolysis) (Jo, et al., 2009). Lipogenesis is induced when nutritional abundance or lipogenic stimuli promote the entry of FFA into the adipocyte, both through passive and active mechanisms (Hajri and Abumrad, 2002). Intracellular FFA are converted into acyl-CoA by the acyl-CoA synthase, which is then then used as a substrate in the endoplasmic reticulum (ER) to produce triacylglycerol (TAG) (Shi and Burn, 2004). Nascent lipid droplets are released from the ER and are ulteriorly coated with perilipin, thus becoming mature lipid droplets (Tansey, et al., 2003).

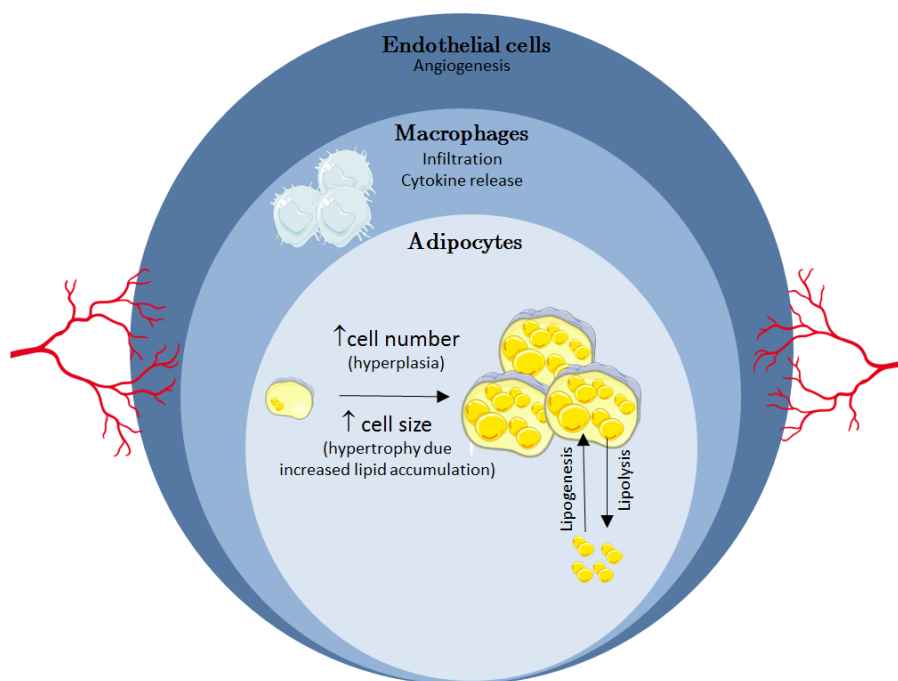


Figure 27. Schematic representation of the several cell types involved in obesity development *in vivo*. During adipose tissue expansion, both adipocyte number (hyperplasia) and size (hypertrophy) increase. The increase on adipocyte size results of increased intracellular lipid accumulation, a consequence of the unbalance between lipogenesis and lipolysis. During the development of obesity *in vivo*, the infiltration of macrophages also occurs, with subsequent release of pro-inflammatory cytokines. Activated adipocytes also produce multiple angiogenic factors that stimulate the development of a vascular network, in order to supply the energetic needs of the growing fat mass. Designed using Servier Medical Art designed medical elements.

Conversely, lipolysis occurs when fasting or lipolytic stimuli (*e.g.* catecholamines, glucagon) induce the phosphorylation of perilipin or of the hormone-sensitive lipase (HSL) (Tansey, et al., 2003). The phosphorylation of perilipin exposes the lipid droplets to the action of HSL, resulting in the hydrolysis of TAG and subsequent release of FFA (Holm, 2003). FFA finally enter the bloodstream and are used as energy sources in the muscle and the liver, where they undergo mitochondrial  $\beta$ -oxidation (Duncan, et al., 2007).

The regulation of the relative rates of lipogenesis and lipolysis are determined by the nutritional status (*e.g.* glucose, FFA) and influenced by a wide range of endocrine factors (Rosen and MacDougald, 2006) (further details in Figure 28). Insulin is the best characterised pro-lipogenic extracellular factor; produced in the pancreatic Langerhans islets, it is secreted in response to the elevation of glucose levels in plasma (Rorsman and Braun, 2013). Other pro-lipogenic factors include the Bone Morphogenic Protein (BMP) and Fibroblast Growth Factor (Rosen and MacDougald, 2006) (Figure 28). Lipogenesis suppressors have also been identified (*e.g.* Wnt-family proteins, Transforming Growth Factor  $\beta$ ). Recently, the Platelet Derived Growth Factor  $\beta$  (PDGFR $\beta$ ) was also proposed as a novel lipogenesis inhibitor (Fitter, et al., 2012). Protein kinase-phosphoproteins interactions also play an important role in the regulation of lipogenesis and lipolysis. Kinases involved in the regulation of glucose metabolism (but also lipid metabolism, proliferation and inflammation) include:  $\alpha$ -Jun N-terminal Kinase (JNK), Mitogen-Activated Protein Kinases (MAPK), Inhibitor of Kappa B-Kinase (IKK), Adenosine Monophosphate-activated Protein Kinase (AMPK), Protein Kinase B (AKT) and Extracellular Signal-Regulated Protein Kinase (ERK), amongst others (Engin, 2017).

Signals from activators and repressors of lipogenesis are integrated in the nucleus by transcription factors that directly or indirectly regulate the expression of peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ ) and CCAAT-enhancer-binding protein  $\alpha$  (C/EBP $\alpha$ ) (Rosen and MacDougald, 2006).

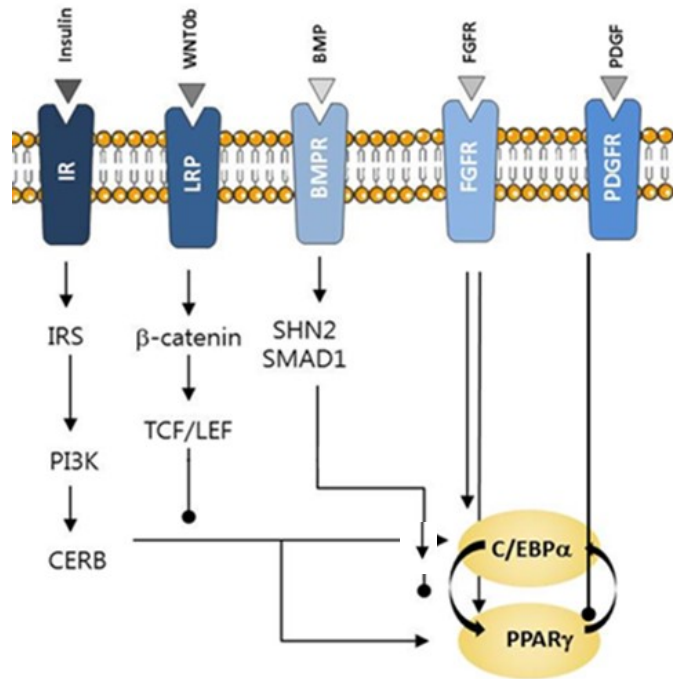


Figure 28. Regulation of lipogenesis by extracellular factors.

Insulin binds to the insulin receptor (IR), inducing the phosphorylation of the insulin receptor substrate (IRS). Bone morphogenetic protein (BMP) promotes lipogenesis by stimulating nuclear localization of Schnurri-2 (SHN2) and SMAD1; the Fibroblast Growth Factor (FGF) also enhances the expression of pro-lipogenic genes, subsequently to the activation of C/EBP $\alpha$  and PPAR $\gamma$ . Conversely, activation of  $\beta$ -catenin signalling by Wnt-family proteins (*e.g.* Wnt10b) represses lipogenesis. Platelet Derived Growth Factor  $\beta$  (PDGFR $\beta$ ) also inhibits lipogenesis (Artemenko, et al., 2005). Activator and repressor signals converge into C/EBP $\alpha$  (CCAAT-enhancer-binding protein  $\alpha$ ) and PPAR $\gamma$  (peroxisome proliferator-activated receptor  $\gamma$ ). BMP: Bone morphogenetic protein; CERB: cAMP response element-binding protein; FGFR: Fibroblast Growth Factor Receptor; LEF/TCF: T-Cell Factor/Lymphoid Enhancer-Binding Factor; PDGFR: Platelet Derived Growth Factor Receptor; PI3K: phosphoinositide-3-kinase; LRP: lipoprotein receptor-related protein receptor. Adapted from (Rosen and MacDougald, 2006) and designed using Servier Medical Art designed medical elements.

The phenotypic changes observed in adipocytes during intracellular lipid accumulation are mirrored by metabolic changes. The increase in adipocyte size is positively correlated with an increase in the expression of leptin, an hormone involved in the regulation of energy regulation and body fat storage (Caro, et al., 1996; Maffei, et al., 1995). Adipocytes also produce adiponectin, a hormone that regulates glucose homeostasis and fatty acid oxidation (Diez and Iglesias, 2003). Adiponectin levels are reduced in obese patients, while weight loss significantly enhances its circulating concentrations (Coppola, et al., 2009). Activated adipocytes also produce multiple angiogenic factors to stimulate the recruitment of macrophages and the development

of a vascular network to supply the growing fat mass (*e.g.* angiopoietin, insulin-like growth factor) (Samad, et al., 1998) (Figure 27).

A growing body of evidence suggests that indoles may play a role in obesity and energy homeostasis. Le Chatelier et al. showed that obese patients show higher numbers of genes associated with the degradation of tryptophan, suggesting that circulating indoles may also be increased in obesity (Le Chatelier, et al., 2013). A recent study in a paediatric cohort reported that obese children, in the context of the Prader-Willi syndrome, exhibit higher abundance of genes involved in the production of indoxyl sulfate and a higher production of this metabolite (Zhang, et al., 2015). In the previous chapter, it was demonstrated that indoxyl sulfate positively correlates with adiposity and leptin levels in adults. However, although high-throughput methods identified associations between increased adiposity and specific aspects of the human gut microbiome, it remains unclear whether these associations have a causal nature, and which are the cellular mechanisms and targets involved (Yassour, et al., 2016).

Although indoxyl sulfate binds AHR, this agonistic activity is not likely to mediate its putative pro-adipogenic effect. First, AHR is depleted during adipose differentiation (Shimba, et al., 1998). Also, in *knock-in* experiments, enhanced expression of AHR promoted inhibition, not stimulation of lipogenesis (Shimba, et al., 2003). Therefore, the identification of its pharmacological targets is critical to understand the molecular crosstalk underlying the effect of indoxyl sulfate in human metabolism (Dumas, 2011). Despite being the second most targeted gene class in all therapeutic areas – in this context, the human kinome remains mostly unexplored (Cohen, 2002; Manning, et al., 2002).

In this chapter, a combined strategy was implemented to clarify the cellular effects of indoxyl sulfate in adipocyte metabolism. Specific aims included:

- a) To validate the effect of indoxyl sulfate on cell viability, lipogenesis and adipocytokine production (*in vitro* model);
- b) To explore novel targets of indoxyl sulfate in the human kinome, and to predict the signalling pathways involved in the microbiome-kinome crosstalk (high-throughput kinome screening combined with network analysis).

## **2. Materials and Methods**

### **2.1. Cell culture**

3T3-L1 preadipocytes were obtained from Sigma Aldrich (Sigma Aldrich Co, LCC) and cultured in Dulbecco's Minimal Eagle Medium with glucose (4500 mg/L) (DMEM, Gibco), supplemented with foetal bovine serum (FBS, 10%), penicillin-streptomycin (1%) and of glutamine (1%) at 37°C, with an atmosphere of 5% CO<sub>2</sub>. When the cell monolayer reached 80% confluency, cells were split and seeded in new T75 flasks at a density of 2-5x10<sup>4</sup> cells/cm<sup>2</sup>. Cells between passage 3 and 10 were used for these experiments.

### **2.2. 3T3-L1 differentiation**

Cells were plated in 24-well plates at a seeding density of 2000 cells/cm<sup>2</sup>. Two days after reaching confluence, 3T3-L1 preadipocytes were induced with DMEM containing 0.4 µM insulin, 500 µM isobutylmethylxanthine (IBMX), 10% FBS, 1% penicillin-streptomycin and 1% glutamine (day 0). Cells were treated in differentiating medium (0.4 µM insulin in DMEM containing 10% FBS, 1% penicillin-streptomycin and 1% glutamine) at day 2, day 4 and day 7, supplemented with either vehicle or indoxyl sulfate (Sigma Aldrich Co, LCC) at several concentrations (0, 0.01, 1 and 100 µM). Ten days after the induction (day 10), 3T3-L1 cells reached full differentiation into adipocytes.

### **2.3. Cell viability assay**

At day 10, 3T3-L1 cells were carefully rinsed with phosphate-buffered saline (PBS, Sigma Aldrich Co, LCC), fixed with paraformaldehyde (3.7%, Sigma Aldrich Co, LCC) for 20 min and finally stained with Crystal Violet (0.05% (w/v), Sigma Aldrich Co, LCC) for 30 minutes. Cells were washed twice with tap water and allowed to dry at

room temperature. Methanol (1 mL) was then added to eluate the dye. Cell viability was measured by reading the absorbance of the eluates (200  $\mu$ L) at 540 nm, using a spectrophotometer plate reader (Synergy H1 Hybrid Reader, BioTek).

## 2.4. Lipogenesis assay

At day 10, lipogenesis was assessed by lipid staining with freshly prepared Oil Red O (Sigma Aldrich Co, LCC). After fixation with paraformaldehyde (further details in *Section 2.3*), 3T3-L1 cells were stained with previously filtered Oil Red-O solution (0.5% (w/v) in isopropanol) and incubated for 1 hour at room temperature. Cells were carefully washed for four times with distilled water, and the stained lipid droplets were observed and photographed (Olympus CK2 Inverted Microscope). The Oil Red-O dye retained in the cell was quantified by elution into straight 2-isopropanol and the absorbance of the eluates (200  $\mu$ L) was measured at 500 nm, using a spectrophotometer plate reader (Synergy H1 Hybrid Reader, BioTek).

## 2.5. Adipocytokine quantification

Cells were treated during differentiation with either vehicle or indoxyl sulfate, as previously described. At day 10, the supernatant medium of each replicate was collected, transferred into a clean Eppendorf tube and centrifuged at 8000 rpm for 5 minutes. Leptin and adiponectin were quantified in the supernatants using, respectively, the Leptin Mouse Enzyme-Linked Immunosorbent Assay (ELISA) kit (ThermoFisher Scientific) and the Quantikine Mouse Adiponectin / Acrp30 ELISA Kit (R&D Systems), accordingly to the instructions of the manufacturer. Each sample was quantified in duplicate; the average of the readings was then calculated and the average zero standard O.D. was subtracted. For both assays, a four-parameter logistic (4-PL) standard curve was generated in GraphPad Prism; sample concentrations were interpolated from their respective standard curve.

## 2.6. Kinome screening

Indoxyl sulfate kinome targets were assessed using the KdSELECT screening service (DiscoverX®), as previously described (Fabian, et al., 2005; Karaman, et al., 2008). This technique is based on a competition-binding assay that quantitatively measures the ability of a compound to compete with an immobilized, active site-directed ligand. The ability of indoxyl sulfate to compete with the immobilized ligand was measured by quantitative PCR of the DNA tag. Results for primary screen binding interactions are reported as % of control (%Ctrl), where lower numbers indicate stronger hits (for further details, see *Chapter 2. General Methodology*).

The binding constant ( $K_D$ ) was calculated from an 11-point dose-response curve for the strongest hits identified in the kinome screening. The TREEspot Software Tool was used to generate kinase interaction tree plots, which were reprinted with permission from KINOMEScan (DiscoverX Corporation). The  $IC_{50}$  of the ten strongest hits was determined using the Kinexus kinase-inhibitor activity profiling service (Kinexus, Vancouver, Canada). The technique is based on the direct quantification of radiolabelled phosphate from ATP onto a protein substrate of a target protein kinases. For further methodological details on the kinome screening, see *Chapter 2. General Methodology*.

## 2.7. Network analysis

The genes of the strongest kinases identified (%Ctrl < 1%) were linked to their corresponding signaling pathways using KEGG API (<http://www.kegg.jp/kegg/docs/keggapi.html>) (Kanehisa and Goto, 2000) and the MetaboSignal package (Rodriguez-Martinez A, 2016). A directed network table (*i.e.* two column matrix) was generated where each row represents an edge connecting two nodes (from node in column 1 to node in column 2). The betweenness-ranked shortest paths ( $n < 8$ ) were retained. In order to identify critical signalling nodes, the standardised betweenness ranks were calculated for each one of the subnetwork

comprising the shortest paths for each one of the kinases. The network was created and visualized in Cytoscape (Shannon, et al., 2003).

## 2.8. Statistical analysis

Normality was tested using the D'Agostino-Pearson Omnibus normality test, and parametric or non-parametric testing chosen accordingly. Comparisons between groups were performed by one-way analysis of variance (ANOVA) combined with Dunnett's *post-hoc* test. The individual number of experiments performed is described in the individual figures. Data are displayed as % of control (vehicle), mean  $\pm$  s.e.m in all figures. A *p*-value of 0.05 was used as threshold for statistical significance.

### 3. Results

#### 3.1. Effect of indoxyl sulfate on 3T3-L1 viability

Briefly, 3T3-L1 cells were treated during induction and differentiation with either vehicle or various concentrations of indoxyl sulfate (0.01, 1 and 100 $\mu$ M), according to the differentiation protocol previously described (see *Section 2.2*). At day 10 post-induction, 3T3-L1 cells were fixed and stained with Crystal Violet in order to quantitatively assess their viability (Figure 29a-b). 3T3-L1 cell viability was not significantly affected by treatment with indoxyl sulfate at any of the doses tested (one-way ANOVA,  $p=0.1770$ ) (Figure 29c).

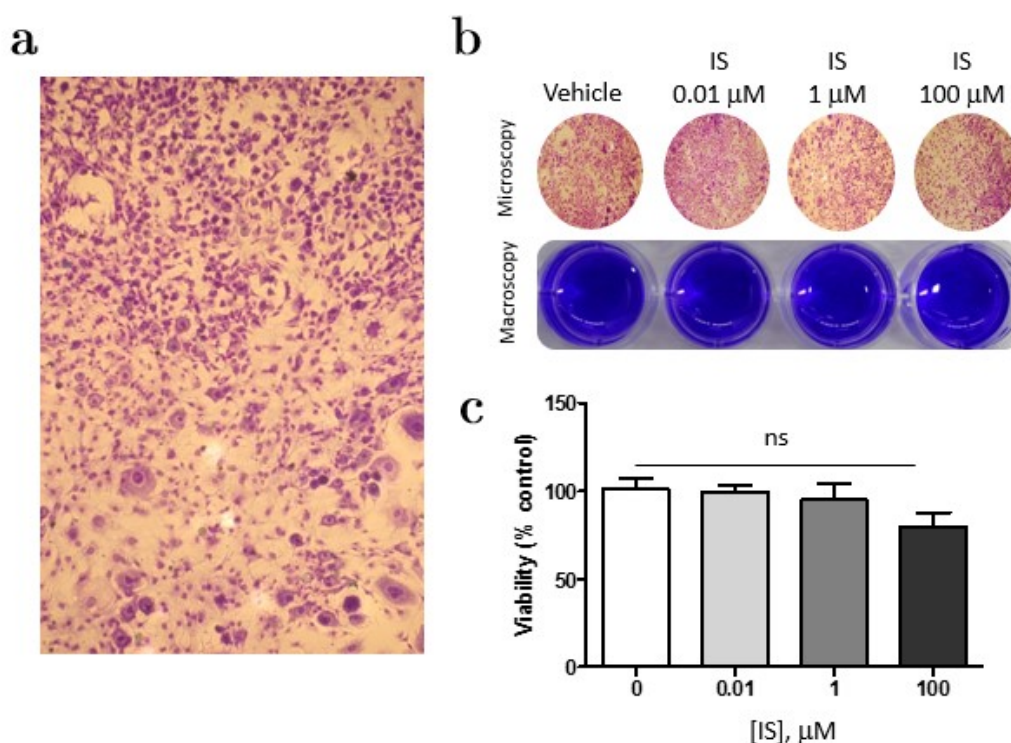


Figure 29. Effect of indoxyl sulfate on cell viability.

3T3-L1 adipocytes were treated with indoxyl sulfate (as indicated) or vehicle (PBS) during differentiation. At day 10, cell number was assessed by Crystal Violet. a) Microscopic cell morphology of 3T3-L1 preadipocytes at day 10 post differentiation (treated with vehicle, 40x magnification). b) Macroscopic and microscopic (10x) representative photographs of the various conditions tested. c) Viability was quantified as the optical density (O.D.) at 540nm. Results are expressed as % of control (vehicle), mean  $\pm$  s.e.m. of two independent experiments with six replicates each. One-way analysis of variance was performed ( $p=0.1770$ ). No significant differences between groups were found using Dunnett's multiple comparisons test. ns: non-significant, IS: indoxyl sulfate.

### 3.2. Effect of indoxyl sulfate on lipid accumulation

The effect of indoxyl sulfate on intracellular lipid accumulation was subsequently tested. 3T3-L1 cells were treated with either vehicle or different doses of the compound (0.01, 1 and 100  $\mu\text{M}$ ), during induction and differentiation. At day 10, intracellular lipid content was quantified by Oil Red O staining (Figure 30).

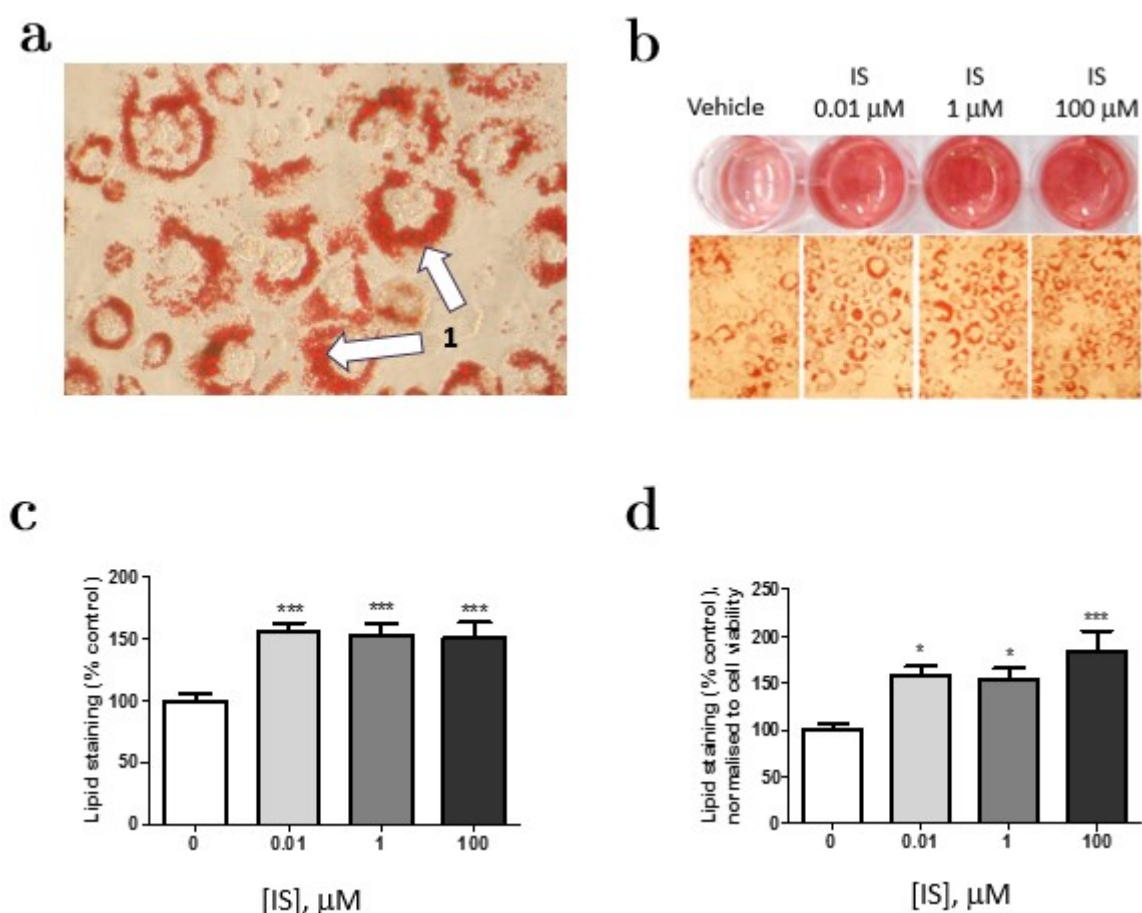


Figure 30. Indoxyl sulfate promotes lipogenesis in vitro in 3T3-L1.

3T3-L1 adipocytes were treated with indoxyl sulfate (as indicated) or vehicle during differentiation. a) Microscopic cell morphology of 3T3-L1 preadipocytes at day 10 post differentiation, after staining with Oil Red O (40x magnification), when multiple lipid droplets were observed (1). b) Macroscopic and microscopic (40x) representative photographs of the conditions tested. c-d) Indoxyl sulfate increases lipid accumulation in both non-normalised and normalised data. Results are expressed as % of control, mean  $\pm$  s.e.m. of two independent experiments with six replicates each. One-way analysis of variance was performed for non-normalised and normalised lipid staining ( $p=0.0002$  and  $p=0.0007$ , respectively) and comparisons between groups were performed by Dunnett's *post-hoc* testing. IS: Indoxyl sulfate. \* $p < 0.05$ , \*\*\* $p < 0.001$ .

Treatment with 0.01, 1 and 100  $\mu\text{M}$  resulted, respectively, in increases of 55.6% ( $p < 0.05$ ), 52.3% ( $p < 0.05$ ) and 51.2% ( $p < 0.001$ ) in lipid content (normalised to cell number), when compared to vehicle-treated cells (Figure 30d). Interestingly, the pro-adipogenic effect of indoxyl sulfate does not seem to be dose-dependent; instead, it reaches a *plateau* somewhere below the concentration of 0.01  $\mu\text{M}$  (Figure 30d).

### 3.3. Effect of indoxyl sulfate on adipocytokines secretion

Apart from accumulating triglycerides in the form of lipid droplets, differentiated adipocytes also synthesize and secrete several adipocytokines, involved in energy homeostasis (Gregoire, et al., 1998). In this experiment, the effect of indoxyl sulfate on the production of leptin and adiponectin was assessed (Figure 31).

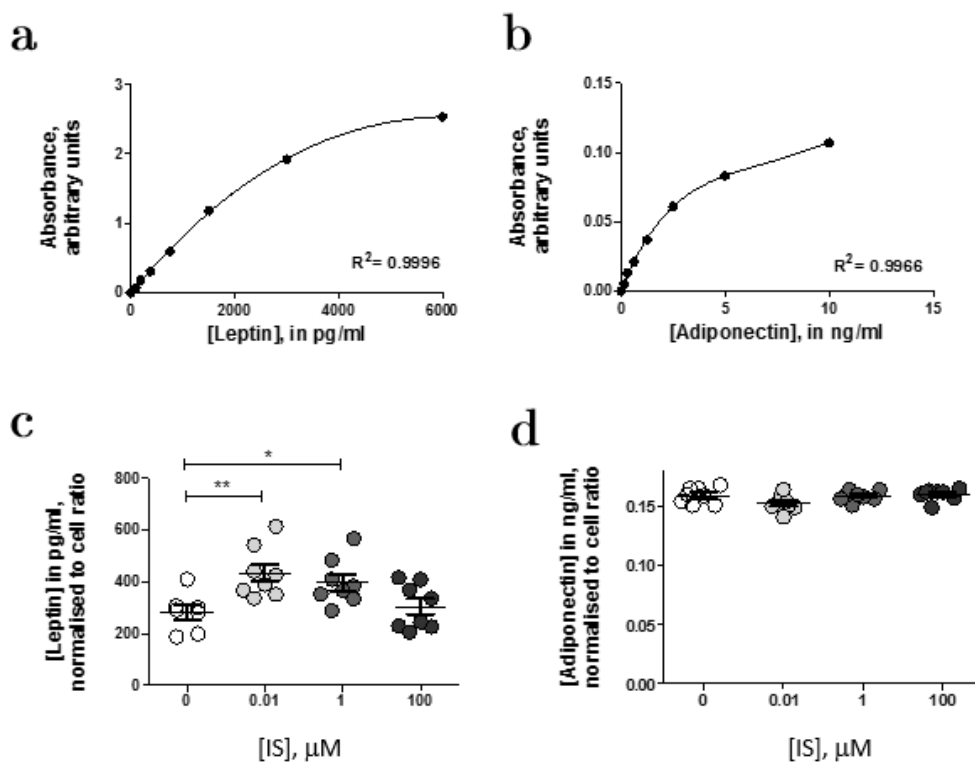


Figure 31. Effect of indoxyl sulfate on adipocytokine secretion in 3T3-L1.

a-b) For both assays, a four-logistic parameter standard curve was generated, as recommended by the manufacturer. c-d) Results were normalised to cell ratio, per replicate, and are expressed as the mean  $\pm$  s.e.m. of two independent experiments with four replicates each. One-way analysis of variance was performed for leptin and adiponectin quantification ( $p=0.006$  and  $p=0.099$ , respectively). Comparisons between groups were performed by Dunnett's *post-hoc* testing. IS: Indoxyl sulfate. \* $p < 0.05$ , \*\* $p < 0.01$ .

Consistently with its pro-adipogenic effect, indoxyl sulfate significantly increased leptin production at both 0.01  $\mu\text{M}$  (+50.6%,  $p < 0.01$ ) and 1  $\mu\text{M}$  (+30.5%,  $p < 0.05$ ) (Figure 31c). Of note, at the highest concentration tested (100  $\mu\text{M}$ ), indoxyl sulfate still increased leptin production by 17.5% compared to vehicle-treated cells, but in this case the effect was not statistically significant ( $p < 0.05$ ) (Figure 31c). However, treatment with indoxyl sulfate did not show any significant effect on adiponectin production at any of the concentrations tested (one-way ANOVA,  $p = 0.099$ ) (Figure 31d).

### 3.4. Investigating the putative targets of indoxyl sulfate in the human kinome

In order to explore novel putative targets mediating the pro-adipogenic effect of indoxyl sulfate in the human kinome, a high-throughput assay including a panel of 456 clinically-relevant kinases was used. Kinases screened were grouped in seven families: Tyrosine Kinases (TK), Tyrosine Kinase-like (TKL), STE kinase group (STE); Cell Kinase 1 (CK1), Protein kinase A, G and C families (AGC), Calcium and Calmodulin-regulated kinases (CAMK) and CMGC kinase group (CMGC). As observed in Figure 32, indoxyl sulfate was able to inhibit 46.3% ( $n=183$ ) of the kinases tested (%Ctrl < 35%).

Eight kinases exhibited extremely high binding affinity in the single-dose kinome screening, expressed as %Ctrl values < 1%. These kinases include Platelet Derived Growth Factor Receptor  $\beta$  (PDGFRB), Serine/Threonine Kinase 32C (STK32C), KIT Proto-Oncogene Receptor Tyrosine Kinase (*c*-KIT), CDC42 Binding Protein Kinase  $\alpha$  (MRCKA), Myosin Light Chain Kinase 3 (MYLK3), CDC42 Binding Protein Kinase  $\beta$  (MRCKB), GLK (Germinal Center Kinase-like kinase), CLK3 (CDC Like Kinase 3) and ZAK (Sterile alpha motif and leucine zipper-containing kinase). The kinases exhibiting highest binding were selected for further confirmation of the physical interaction, by defining a multiple-dose binding curve, and thus calculating the respective dissociation constant ( $K_D$ ) values.  $K_D$  values ranged from 9.9 to 730  $\mu\text{M}$  PDGFRB exhibited the lowest  $K_D$  and, therefore, the highest affinity (Table 7).

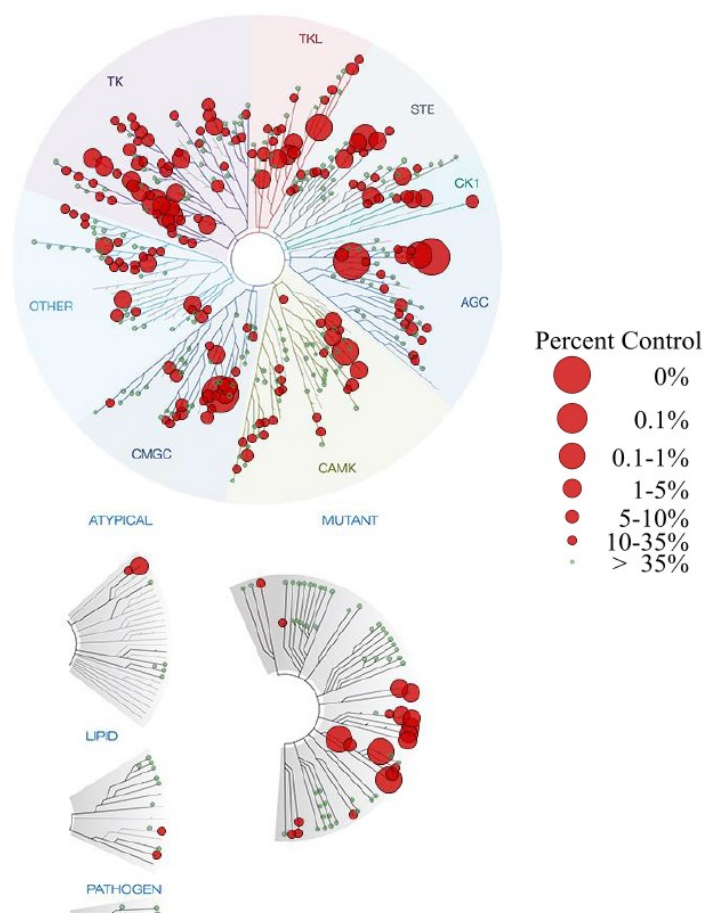


Figure 32. Indoxyl sulfate exhibits kinome-wide effects in a single-dose screening.

Indoxyl sulfate at a single dose (10  $\mu$ M) was screened against a panel of 456 clinically-relevant kinases. Preliminary screening identified 186 positive hits with potential binding <35 % *vs* control (DMSO), represented by a red dot whilst non-binding kinases are represented by a green dot. Kinase group names: TK, Tyrosine Kinases; TKL, Tyrosine Kinase-like; STE, STE kinase group; CK1, Cell Kinase 1; AGC, Protein kinase A, G and C families; CAMK, Calcium and Calmodulin-regulated kinases; CMGC, CMGC kinase group.

Table 7. High binding affinity kinases for indoxyl sulfate, ranked by  $IC_{50}$

| Acronym       | Name                                                     | $K_D$       | IS Activity | $IC_{50}/EC_{50}$    |
|---------------|----------------------------------------------------------|-------------|-------------|----------------------|
| PDGFRB        | Platelet Derived Growth Factor Receptor $\beta$          | 9.9 $\mu$ M | Inhibitor   | $IC_{50}$ = 11.7 nM  |
| STK32C        | Serine/Threonine Kinase 32C                              | 17 $\mu$ M  | Activator   | $EC_{50}$ = 2.89 mM  |
| $\alpha$ -KIT | KIT Proto-Oncogene RTK                                   | 20 $\mu$ M  | Inhibitor   | $IC_{50}$ = 10.15 mM |
| MRCKA         | CDC42 Binding Protein Kinase $\alpha$                    | 26 $\mu$ M  | Inhibitor   | $IC_{50}$ = 6.12 mM  |
| MYLK3         | Myosin light chain kinase 3                              | 45 $\mu$ M  | Inhibitor   | $IC_{50}$ = 7.00 mM  |
| MRCKB         | CDC42 Binding Protein Kinase $\beta$                     | 49 $\mu$ M  | Inhibitor   | $IC_{50}$ = 33.25 mM |
| GLK           | Germinal center kinase-like kinase                       | 210 $\mu$ M | Inhibitor   | $IC_{50}$ = 4.61 mM  |
| ZAK           | Sterile $\alpha$ motif /leucine zipper-containing kinase | 730 $\mu$ M | Inhibitor   | $IC_{50}$ = 6.84 mM  |

Half maximal inhibitory concentration ( $IC_{50}$ ) or half maximal effective concentration ( $EC_{50}$ ) values for each kinase are shown in Table 7. STK32C is the only kinase activated by indoxyl sulfate, but only at a very high, non-physiological concentration (10 mM). All other kinases assessed are inhibited by indoxyl sulfate, with the strongest inhibitory effect observed for PDGFRB ( $IC_{50} = 11.7$  nM). Other kinases inhibited by indoxyl sulfate ( $c$ -KIT, MRCKA, MYLK3, MRCKB, GLK, CLK3 and ZAK) exhibit  $IC_{50}$  values in the micromolar range (Table 7).

Amongst the 186 kinases screened, PDGFRB emerged as a strong candidate as a human target for indoxyl sulfate, with both the lowest  $KD$  and  $IC_{50}$  values. Interestingly, PDGFRB has previously been identified a negative regulator of lipogenesis (Artemenko, et al., 2005). Imatinib (a PDGFR-B inhibitor) promotes lipogenesis and increases the expression of key adipogenic transcription factors (*e.g.* C/EBP and PPAR $_{\gamma}$ ) (Fitter, et al., 2012). Therefore, it was hypothesised that the pro-adipogenic effect of indoxyl sulfate could be mediated *via* PDGFRB inhibition.

As expected, PDGF-DD (a PDGFRB agonist) reduced lipogenesis in 52.7% compared to the vehicle ( $p < 0.0001$ ); this effect was reverted by imatinib ( $p < 0.0001$ ). Consistently with previous results, indoxyl sulfate (100  $\mu$ M) increased lipid accumulation in 15.5% compared to vehicle ( $p < 0.05$ ). However, indoxyl sulfate is not able to counteract the anti-adipogenic effect of PDGFRB activation, thus suggesting that its effect is not mediated by PDGFRB inhibition (Figure 33).

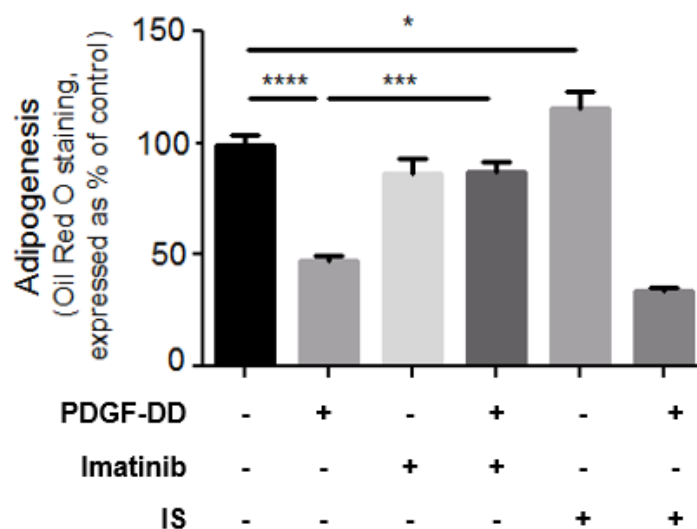


Figure 33. Indoxyl sulfate does not counteract the anti-adipogenic effect of PDGF-DD.

3T3-L1 adipocytes were treated vehicle, PDGF-DD (1  $\mu$ g/ml), imatinib (5  $\mu$ g/ml) or indoxyl sulfate (100  $\mu$ M). At day 10, lipogenesis was quantified by Oil Red O staining. PDGF-DD decreases lipid accumulation ( $p<0.0001$ ) and this effect is lost in the presence of imatinib ( $p<0.0001$ ). Results are expressed as mean  $\pm$  s.e.m. of at two independent experiments with 4 replicates each. One-way analysis of variance was performed. Tukey's post-hoc test was used for comparisons between groups \* $p<0.05$ ; \*\*\* $p<0.001$ . IS: indoxyl sulfate; PDGF-DD: Platelet Derived Growth Factor DD.

### 3.5. Network prediction of targets of indoxyl sulfate involved in lipogenesis-related signalling pathways

In order to identify critical signalling nodes between the identified kinome targets identified and lipogenesis, a network analysis approach was adopted. The KEGG API (Kanehisa and Goto, 2000) and the MetaboSignal package (Rodriguez-Martinez A, 2016) were used to calculate the betweenness-ranked shortest paths between the eight validated high-affinity kinases and triacylglycerol were calculated. Six out of the eight kinases validated were retained in the network analysis (PDGFRB, ZAK,  $c$ -KIT, MRCKA, MYLK3 and MRCKB) (Figure 34). The shortest paths (shortest geodesic distance) were observed for PDGFRB and  $c$ -KIT (both *via* 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase  $\beta$ -2 (PLCB2)-1,2 diacyl-*sn*-glycerol).

Next, the betweenness rank was calculated for each node of the subnetwork comprising the shortest paths. The node with a highest standardized betweenness rank was PLCB2 ( $3.57 \times 10^{-3}$ ) (Table 8), into which converge the signaling of 66.7% ( $n=4$ ) of the shortest paths (PDGFRB, ZAK, c-KIT and MRCKA), including the two with the shortest geodesic distance. According to the network analysis, PLCB2 emerges as a critical signalling node into which converge the multi-kinase inhibition induced by indoxyl sulfate, and that can subsequently impact the adipogenic process itself.

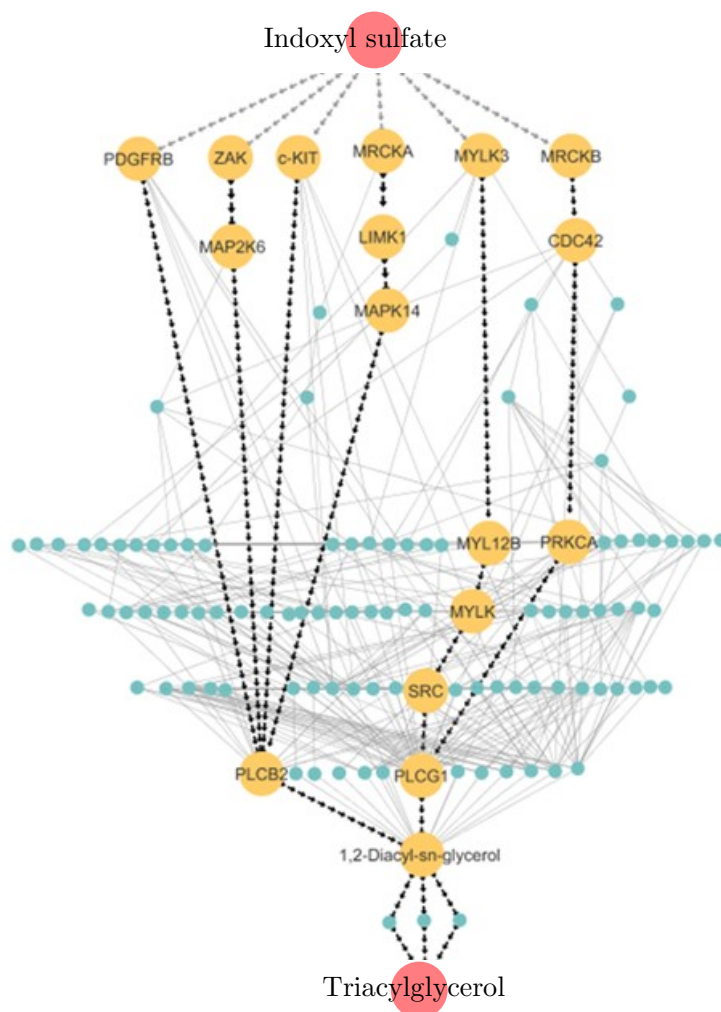


Figure 34. Prediction of relevant signalling pathways and metabolic nodes involved in the pro-adipogenic effect of indoxyl sulfate.

The signalling network was computed using the previously identified kinases as starting point and triacylglycerol as a final point. Betweenness-ranked shortest paths were computed using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) Application Programming Interface (Kanehisa and Goto, 2000) and the MetaboSignal package (Rodriguez-Martinez A, 2016). The nodes comprised in the shortest paths for each kinase are highlighted in yellow. A full description of the acronyms of the nodes is provided in Table 8.

Table 8. Betweenness ranks for the nodes of the shortest-paths subnetwork.

| Node acronym | Node name                                                            | <i>Betweenness rank</i> |
|--------------|----------------------------------------------------------------------|-------------------------|
| PLCB2        | 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase $\beta$ -2 | $3.57 \times 10^{-3}$   |
| MYLK         | Myosin light chain kinase                                            | $2.01 \times 10^{-3}$   |
| CDC42        | Cell division control protein 42 homolog                             | $1.82 \times 10^{-3}$   |
| SRC          | Proto-oncogene tyrosine-protein kinase Src                           | $1.57 \times 10^{-3}$   |
| PRKCA        | Protein kinase C $\alpha$                                            | $1.22 \times 10^{-3}$   |
| MAPK14       | Mitogen-activated protein kinase kinase 14                           | $1.07 \times 10^{-3}$   |
| MYLK12B      | Myosin light chain kinase 12B                                        | $5.07 \times 10^{-4}$   |
| MAPK2K6      | Dual specificity mitogen-activated protein kinase kinase 6           | $4.96 \times 10^{-4}$   |
| LIMK1        | LIM domain kinase 1                                                  | $3.15 \times 10^{-4}$   |

## 4. Discussion

In the previous chapter, a  $^1\text{H}$  NMR-based metabolic profiling strategy was used to demonstrate that indoxyl sulfate is positively correlated with adiposity in a human cohort. Increased adiposity in human is a consequence of the increase on the number of adipocytes present in the adipose tissue (hyperplasia) and altered cellular metabolism (Jo, et al., 2009). In this chapter, a 3T3-L1 cell line model was used to validate *in vitro* the effect of indoxyl sulfate in both cell viability and metabolic functions (*i.e.* lipogenesis and leptin production). Furthermore, by combining a kinome screening high-throughput strategy with network analysis, relevant targets and signalling pathways putatively involved in the pro-adipogenic effect were identified.

### 4.1. Indoxyl sulfate does not affect 3T3-L1 cell viability

In healthy adults, the concentration of indoxyl sulfate in plasma is in the micromolar range (maximal plasmatic levels varying between 2.50 and 14.0  $\mu\text{M}$  (Duranton, et al., 2012)); higher concentrations can be observed in uremic patients (21.7  $\mu\text{M}$ , (Wu, et al., 2011)). Therefore, in this study, the doses used were in the micromolar range (0.01, 1 and 100  $\mu\text{M}$ ). Indoxyl sulfate did not induce significant

changes on 3T3-L1 viability at any of the concentrations tested. These results are consistent with a recently published study showing that indoxyl sulfate does not exert significant deleterious effects on 3T3-L1 cell viability up to 1000  $\mu$ M (Stockler-Pinto, et al., 2016). Therefore, the pro-adipogenic effect of indoxyl sulfate is not exercised through a stimulating effect on adipocytes (hyperplasia). These concentrations, at which no toxic effects were observed, were used to perform the following experiments.

## 4.2. Indoxyl sulfate promotes lipid accumulation in 3T3-L1

3T3-L1 cells treated with indoxyl sulfate during induction and differentiation exhibited increased accumulation of intracellular lipids, at all doses tested (51.2 - 55.6%,  $p < 0.05$ ), when compared to vehicle. This is the first time that a pro-adipogenic effect is reported for indoxyl sulfate *in vitro*.

In order to confirm the pro-adipogenic effect of indoxyl sulfate, the levels of leptin and adiponectin were also evaluated. Previous kinetic studies demonstrated that leptin production increases during adipocyte differentiation, subsequently to the transcriptional activation of its gene *obese* (MacDougald, et al., 1995). In agreement with these results, it was found that treatment with indoxyl sulfate increased leptin production at both 0.01  $\mu$ M and 1  $\mu$ M, thus reinforcing the putative pro-adipogenic role of indoxyl sulfate *in vitro*. Of note, at the highest concentration tested (100  $\mu$ M), indoxyl sulfate still increased leptin production by 17.5% compared to vehicle-treated cells, however, this increase was not significant. In striking contrast, in the present study, indoxyl sulfate did not affect adiponectin secretion. Although adiponectin and leptin both increase during adipocyte differentiation, they are secreted through distinct intracellular trafficking pathways; in particular, functional endosomes are required for adiponectin, but not for leptin secretion (Xie, et al., 2008). The lack of effect observed in adiponectin production might suggest that, concurrently with its pro-adipogenic effect, indoxyl sulfate may interfere with other specific aspects of the adiponectin-secreting pathway.

As future work, these findings can be strengthened by increasing the number of experiments and by validating the pro-adipogenic effect using other molecular biology techniques. Adipocyte differentiation can be further assessed by quantifying the expression of adipocyte marker genes, including PPAR $\gamma$ , a nuclear receptor from the ligand-activated transcription factors family and vital for adipocyte differentiation (Gregoire, et al., 1998). Other gene markers of adipocyte differentiation include *C/EBP $\alpha$* , *FASN* and *ADIPOQ* (Gregoire, et al., 1998; Sul, et al., 2000).

#### **4.3. Advantages and limitations of cell-based assays in the study of adipocyte metabolism and obesity**

In this study, 3T3-L1 were used as a model to study adipocyte cell metabolism *in vitro*. Originally isolated from Swiss 3T3 mouse embryos (Green and Kehinde, 1975), 3T3-L1 offer an abundant source of homogenous cells responsive to metabolic stimuli and are therefore often used to investigate the effect of a given compound on lipogenesis (Ruiz-Ojeda, et al., 2016). However, one of its main disadvantages is that both the initial subculture and the adipogenic differentiation are time-consuming and that, being a mouse cell line, they may not fully reflect the effects on human metabolism (Student, et al., 1980). Also, as previously described, the increase of adipose tissue *in vitro* is the final result of an orchestrated effect that involves not only adipocytes, but also macrophages and vascular endothelial cells. Therefore - and since indoxyl sulfate has known effects in vascular endothelial cells (Dou, et al., 2007; Faure, et al., 2006) - *in vitro* studies using co-cultures of both cell types, as well as *in vivo* studies, are required to provide a more comprehensive overview on the effect of indoxyl sulfate on adipose tissue.

#### **4.4. Network analysis predicts that indoxyl sulfate has a multi-kinase pro-adipogenic effect**

After validating the pro-adipogenic effect of indoxyl sulfate *in vitro*, putative targets mediating this effect were investigated. One of the best studied human targets of indoxyl sulfate is AHR; however, since this receptor is depleted during adipose

differentiation (Shimba, et al., 1998) and its overexpression inhibits lipogenesis (Shimba, et al., 2003), it does not seem to be involved in the pro-adipogenic effect of indoxyl sulfate.

The high-throuput kinome screening, followed by confirmation of the binding and functional activity, identified eight high affinity targets for indoxyl sulfate in the human kinome (PDGFRB,  $\alpha$ -KIT, MRCKA, MYLK3, MRCKB and ZAK). Amongst them, PDGFRB exhibited the lowest  $KD$  and  $IC_{50}$  values. PDGFRB was previously proposed as a negative regulator of lipogenesis (Artemenko, et al., 2005); its inhibition by imatinib promotes lipogenesis and increases the expression of key adipogenic transcription factors (*e.g.* C/EBP and PPAR  $\gamma$ ) (Fitter, et al., 2012). Based on these results, it was hypothesised that the pro-adipogenic effect of indoxyl sulfate could be mediated *via* PDGFRB inhibition. However, in the concentrations tested, this hypothesis was not confirmed: indoxyl sulfate was not able to counteract the anti-adipogenic effect of PDGFRB activation, thus suggesting that its pro-adipogenic effect is not mediated, at least solely, by PDGFRB inhibition. Additionally, given the kinome-wide impact of indoxyl sulfate, it is more likely that its impact on adipocyte metabolism is mediated by multiple kinases.

The network analysis predicted six kinases (PDGFRB,  $\alpha$ -KIT, MRCKA, MYLK3, MRCKB and ZAK) with shortest betweenness-ranked paths to triacylglycerol; shortest betweenness ranks, in this case, are a proxy measure to estimate their relevance in the signalling network leading to lipogenesis. Amongst these, PDGFRB, C-KIT, MRCKA, MRCKB are expressed in the adipose tissue (Ponten, et al., 2008). Similarly to PDGFRB,  $\alpha$ -KIT has also been implicated in the development of obesity, as a loss-of-function mutation of these gene results in increased obesity and decreased energy expenditure in mice (Huang, et al., 2014). Both PDGFRB and c-KIT signalling pathways converge into PLCB2, a critical node showing the highest standardized betweenness rank in the shortest-paths subnetwork. PLCB2 is expressed in adipose tissue (Ponten, et al., 2008) and codes for the enzyme 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase  $\beta$ 2, that converts phosphatidylinositol 4,5-biphosphate into inositol 1,4,5-triphosphate and diacylglycerol, which is further esterified into triacylglycerol by the diglyceride acyltransferase (Saponaro, et al., 2015).

Altogether, the network analysis presents a comprehensive map of the strongest candidate targets in the human kinome, revealing key targets and signalling key nodes for future functional investigations. As future work, phosphoproteomics approaches combined with network analysis may identify selected kinases and signalling pathways for individual validations.

## 5. Conclusion

In this chapter, it was demonstrated that indoxyl sulfate exhibits a pro-adipogenic effect *in vitro*, characterised by increased intracellular lipid accumulation and enhanced leptin production. These results are consistent with the findings from the previous chapter, in which indoxyl sulfate significantly correlated with adiposity and plasma levels of leptin. By proving a direct pro-adipogenic effect in adipocytes *in vitro*, the results shown in the chapter demonstrate that these associations do not rely on pure correlation: conversely, they might represent a causal effect of gut microbial metabolism on human adipose tissue function.

Furthermore, a combined strategy comprising a high-throughput kinome screening and a network analysis approach was used to investigate the potential targets in the human kinome mediating this effect. This strategy revealed that indoxyl sulfate has kinome-wide effects, and network analysis shortlisted six kinases associated to shortest betweenness-ranked paths to triacylglycerol (PDGFRB,  $\alpha$ -KIT, MRCKA, MYLK3, MRCKB and ZAK). Importantly, PLCB2 was identified as a critical node into which several of the shortest-path signalling pathways converge – however, further studies (*i.e.* phosphoproteomics combined with network analysis) are needed to further clarify the individual contribution of each pathway, followed by subsequent validations.

By identifying and validating a role for indoxyl sulfate as a gut microbial metabolite with pro-adipogenic properties, these results propose a novel mechanism through which the gut microbiota can impact human obesity.





## Chapter 6

# Exploring the effect of indole-3-acetate in the human kinome

### 1. Introduction

Indole-3-acetate is one of the most abundant indolic compounds in human plasma, with plasmatic concentrations raising up to 3  $\mu\text{M}$  and 19  $\mu\text{M}$ , respectively, in healthy and uremic patients (Calaf, et al., 2011; Duranton, et al., 2012). Once in bloodstream, indole-3-acetate can bind human pharmacological targets, therefore impacting signalling networks and metabolic pathways relevant for cardiometabolic diseases. The identification of these pharmacological targets can, therefore, improve our understanding of the microbial-mammalian molecular crosstalk, and *in fine* the impact of gut microbial metabolites in human health and disease (Dumas, 2011).

Previous studies attempted to identify putative targets for indoles, mostly GPCR and nuclear receptors (Chilloux, et al., 2016; Gondouin, et al., 2013; Jin, et al., 2014; Venkatesh, et al., 2014). Indoles were shown to act as ligands for AHR; however, this effect is not class-specific, with different indolic compounds exhibiting either agonist or antagonist activities (Jin, et al., 2014). For indole-3-acetate, only a very weak AHR agonism was observed *in vitro* when dosing young adult mouse colonocytes (YMAC cells) at high, non-physiological, doses (250  $\mu$ M); a partial was also observed, but only at even higher concentrations (500  $\mu$ M) (Cheng, et al., 2015).

In this chapter, it was hypothesised that indole-3-acetate can impact human metabolism by binding kinome targets. Whilst most of the studies to the date focused on defining agonisms/antagonisms of gut microbial metabolites for GPCR and nuclear receptors, screening the kinome opens a novel research front to understand the impact of gut microbial metabolites in human health.

The primary aim of this chapter is a) to identify indole-3-acetate's targets in the human kinome using a high-throughput kinase screening, followed by confirmation of the binding (dissociation constant,  $K_D$ ) and of the modulation of the phosphorylation activity ( $IC_{50}$ ). After confirming the inhibitory effect of indole-3-acetate in a single kinase (VEGFR-2), the secondary aims included:

- b) To model the physical interaction between indole-3-acetate and VEGFR-2, using molecular docking analysis;
- c) To validate the effect of indole-3-acetate on VEGFR-2 phosphorylation *in vitro*;
- d) To assess the impact of indole-3-acetate on signalling pathways downstream of VEGFR-2.

## 2. Materials and Methods

### 2.1. Kinase screening

Indole-3-acetate targets in the human kinome were explored using the KdSELECT screening service (DiscoverX®), as previously described (Fabian, et al., 2005; Karaman, et al., 2008). The ability of indole-3-acetate to compete with the immobilized ligand was measured by quantitative Polymerase Chain Reaction (PCR) of the DNA tag. Results for primary screen binding interactions are reported as % of control (%Ctrl), where lower numbers indicate stronger hits (for further details, see *Chapter 2. General Methodology*).

The binding constant ( $K_D$ ) was then calculated from an 11-point dose-response curve. The TREEspot Software Tool was used to generate kinase interaction tree plots, which were reprinted with permission from KINOMEScan® (DiscoverX Corporation®). The  $IC_{50}$  of indole-3-acetate on VEGFR-2 was determined using Kinexus kinase-inhibitor activity profiling service (Kinexus, Vancouver, Canada). The technique is based on the direct quantification of radiolabelled phosphate from ATP onto a protein substrate of a target protein kinase - in this case, VEGFR-2.

### 2.2. Molecular docking

The macromolecular structure data of VEGFR-2 (id: 5EW3) was obtained from the Protein Database of the Research Collaboratory for Structural Bioinformatics (RCSB PDB) and pre-processed using the SYBYL® platform (Tripos, St. Louis, MO). Water molecules were removed and the protein was protonated at  $\text{pH} = 7$  and minimised using 100 interactions. The structure of indole-3-acetate was obtained from PDBeChem (EMBL-EBI, Cambridge, UK). Structure data were loaded into GOLD software, and the docking parameters were optimised and validated by binding a known inhibitor (AAL993) to VEGFR-2 (Bold, et al., 2016). The validation results were used to define the binding site as a sphere with a radius of 8 Å around the centre

carbon atom in Val-848. Automatic detection of the receptor cavity was applied in order to limit atom binding events to solvent-accessible structures. Docking solutions were calculated by the ChemPLP function, a piecewise linear potential empirical scoring function, that models the steric interactions between two molecules as a function of heavy-atom clash potential, torsional potential and hydrogen donors and acceptors (Korb, et al., 2009). Model visualisations were performed in PyMOL®.

### 2.3. Cell culture

Human umbilical vein endothelial cells (HUVEC) were obtained from BD Biosciences. In brief, cells were grown in collagen I-coated flasks and maintained in endothelial growth medium (EGM-2, Lonza), in an atmosphere of 5% CO<sub>2</sub> at 37°C. After reaching 80% confluence, cells were trypsinised and split at the ratio of 1:2 to 1:6. Cells between passage number 3 and 10 were used for the experiments described.

### 2.4. Total and phosphorylated VEGFR-2 quantification

HUVEC cells were plated in 6-well collagen-coated plates (BD BioCoat. After reaching 80% confluence, cells were starved for 2 hours in EBM (Lonza). Cells were treated with vehicle (0.1% ethanol) or indole-3-acetate in maintenance medium without added VEGF<sub>A</sub> for 15 minutes. Stimulation with VEGF<sub>A</sub> (2 ng/mL, Sigma Aldrich) was then performed, and cell lysates were collected at 10 seconds, 10 minutes, 60 minutes and 240 minutes post-stimulation; for further experiments, the 10 minutes post-stimulation time point was adopted.

For preparation of the cell lysates, cells were first washed with ice-cold phosphate buffer saline, followed by lysis in RIPA buffer (Sigma Aldrich), supplemented with Protease Inhibitor Cocktail (Phosphatase Inhibitor Cocktail 2 and Phosphatase Inhibitor Cocktail 3, both at 1:100, Sigma Aldrich) and phenylmethylsulfonyl fluoride (PMSF) 1mM (Sigma Aldrich). After a 5-minute incubation at 4°C, cells were scraped and transferred to an Eppendorf tube and centrifuged for 10 minutes (14.000g, at 4°C). Cell lysates were finally transferred to a new tube and stored at -80 °C. Total and phosphorylated VEGFR-2 were quantified

on cell lysates using the PathScan Phospho-VEGFR-2 (Tyr<sup>1175</sup>) Sandwich ELISA kit and PathScan Total VEGFR-2 Sandwich ELISA kit, both from Cell Signalling Technology, according to the manufacturer's instructions.

## **2.5. AKT phosphorylation assay**

HUVEC cells were cultured as described in 2.4. Cell lysates were prepared with Phosphatase Inhibitor I (1:100), Phosphatase Inhibitor II (1:100), and Protease Inhibitor Solution (1:100) diluted in Assay Diluent 39 (Meso Scale Discovery). Total and phosphorylated AKT were measured in undiluted samples using the MSD Multi-Spot Assay System for Phospho-AKT (Ser473) Kit, a sandwich immunoassay performed, according to the manufacturer's instructions. The plates were read using a MESO Sector Imager 2400 (Meso Scale Discovery).

## **2.6. NO quantification assay**

HUVEC cells were plated in 6-well collagen-coated plates (BD BioCoat), in maintenance medium (EGM-2, Lonza ®). After reaching 80% confluence, cells were starved for 2 hours in basal medium (EBM, Lonza). Cells were treated with vehicle (0.1% ethanol) or indole-3-acetate in normal maintenance medium (without added VEGF<sub>A</sub>) for 15 minutes. Stimulation with VEGF<sub>A</sub> (2 ng/mL) was then performed, and cell supernatants were collected after 24h. Production of NO was measured using the Total NO and Nitrate/Nitrite assay (R&D systems).

## **2.7. Statistical analysis**

For statistical comparisons between groups, normality was tested using D'Agostino & Pearson Omnibus normality test. Parametric or non-parametric one-way analysis of variance (ANOVA) was used, followed by *post hoc* testing; for AKT and NO experiments, one-tail t-tests were performed. A threshold of  $p < 0.05$  was adopted for statistical significance. Data are displayed as mean  $\pm$  s.e.m in all figures. The number of cell culture experiments is indicated in individual figure legends.

## 3. Results

### 3.1. Exploring the targets of indole-3-acetate in the human kinome

In order to discover physical interactions between indole-3-acetate and human kinases, a high-throughput assay was used to screen this compound against a panel of 456 clinically-relevant kinases. Indole-3-acetate showed a very specific physical interaction with a single kinase, the Vascular Endothelial Growth Factor Receptor-2 (VEGFR-2) (Figure 35a).

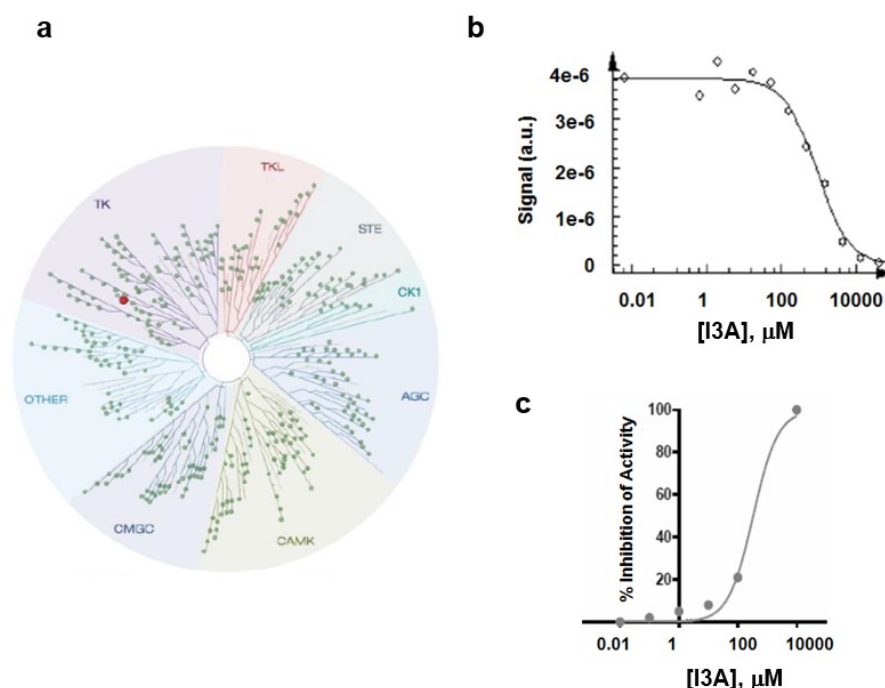


Figure 35. Indole-3-acetate binds VEGFR-2 and functionally inhibits its phosphorylation activity in a high-throughput kinase screening.

a) VEGFR-2 was the only target identified for indole-3-acetate in the single dose-kinome screening (10000 nM). Preliminary positive hits with potential binding  $<35\%$  are represented by a red dot, whilst non-binding kinases are represented by a green dot. Kinase group names: TK, Tyrosine Kinases; TKL, Tyrosine Kinase-like; STE, STE kinase group; CK1, Cell Kinase 1; AGC, Protein kinase A, G and C families; CAMK, Calcium and Calmodulin-regulated kinases; CMGC, CMGC kinase group. b) Confirmation of a physical interaction between indole-3-acetate and VEGFR-2 ( $K_D=930 \mu\text{M}$ ); c) Confirmation of the functional inhibition of VEGFR-2 phosphorylation activity by indole-3-acetate ( $IC_{50}=333.7 \mu\text{M}$ ). a.u.: arbitrary units.

No other physical interactions between indole-3-acetate and other kinases were identified. The binding between indole-3-acetate and VEGFR-2 was confirmed by a multiple-dose binding curve, defining a dissociation constant ( $K_D$ ) of 890  $\mu\text{M}$  (Figure 35b). As the  $K_D$  only assesses the physical interaction (*i.e.* binding), the functional effect was tested by quantifying VEGFR-2 phosphorylation in the presence of ATP and multiple doses of indole-3-acetate. Increasing doses of indole-3-acetate progressively inhibited VEGFR-2 phosphorylation activity, defining a half-maximal inhibitory constant ( $IC_{50}$ ) of 333.7  $\mu\text{M}$  (Figure 35c), and therefore confirming a functional effect.

### **3.2. Investigating the interaction between indole-3-acetate and VEGFR-2 by molecular docking analysis**

Next, the mechanism underlying the inhibitory effect of indole-3-acetate on VEGFR-2 was explored. First, the preferred binding sites of both ATP and indole-3-acetate on VEGFR-2 molecule were predicted, determined as the 10 best docking solutions based on the respective ChemPLP scores (Table 9). The best docking solution for ATP (solution number 6, ChemPLP score = 52.00) was predicted to form hydrogen bonds with three different residues of the VEGFR-2 molecule (Cys-919, Glu-885 and Asp-1046) (Table 9). A spatial visualisation of this solution is shown in Figure 36a-b.

For indole-3-acetate, the highest scoring docking solution (solution 8, ChemPLP = 49.89) predicted the formation of a bond between one of the oxygen atoms of indole-3-acetate and the backbone nitrogen atom of the Asp-1046 residue of VEGFR-2 (Table 9 and Figure 36c-d). As previously shown, ATP interacts with the same residue. The predicted length of the bond formed between indole-3-acetate and Asp-1046 (2.74 Å) is shorter than the sum of the van der Waals radii of both atoms, thus suggesting a hydrogen bond.

Table 9. Best docking solutions for ATP and indole-3-acetate, based on molecular docking predictions (ChemPLP scores)

| ATP              |          |                                                       |
|------------------|----------|-------------------------------------------------------|
| Solution         | ChemPLP* | Residues                                              |
| 1                | 46.11    | Cys-919, Glu-917, Glu-885, Asp-1046, Lys-868, Leu-840 |
| 2                | 50.58    | Cys-919, Glu-917, Glu-885, Asp-1046, Lys-868, Asn-923 |
| 3                | 49.74    | Cys-919, Glu-917, Glu-885, Asp-1046, Lys-868, Asn-923 |
| 4                | 10.52    | Asn-923, Cys-1045, Lys-868, Val-914, Glu-885          |
| 5                | 44.94    | Cys-919, Glu-917, Glu-885, Asp-1046, Lys-868, Asn-923 |
| 6                | 52.00    | Cys-919, Glu-885, Asp-1046                            |
| 7                | 51.84    | Cys-919, Gly-885, Asp-1046, Asn-923                   |
| 8                | 44.46    | Cys-919, Glu-885, Asp-1046                            |
| 9                | 47.94    | Cys-919, Glu-917, Glu-885, Asp-1046, Lys-868, Leu-840 |
| 10               | 49.10    | Cys-919, Glu-917, Asp-1046, Lys-868, Leu-840          |
| Indole-3-acetate |          |                                                       |
| Solution         | ChemPLP  | Residues                                              |
| 1                | 45.58    | Cys-919                                               |
| 2                | 46.40    | Cys-919                                               |
| 3                | 75.86    | Asp-1046                                              |
| 4                | 47.09    | Cys-919                                               |
| 5                | 45.96    | Cys-919                                               |
| 6                | 45.93    | Asp-1046                                              |
| 7                | 45.88    | Asp-1046                                              |
| 8                | 49.89    | Asp-1046                                              |
| 9                | 46.34    | Cys-919                                               |
| 10               | 49.82    | Asp-1046                                              |

\*Highest ChemPLP scores represent higher binding affinity. Highest solution numbers and CHEMPLP scores are shown in bold. Residues from the best solutions for each case are also shown in bold whenever present in other solutions.

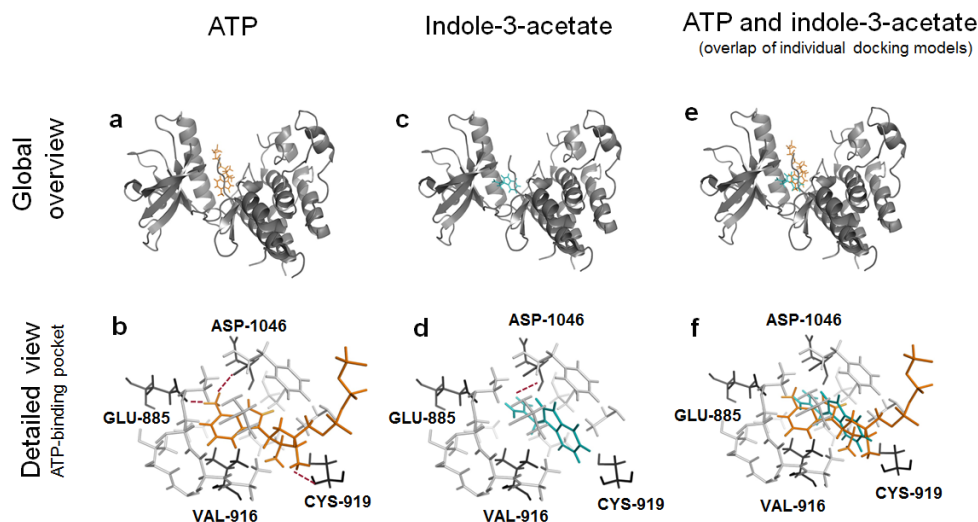


Figure 36. Indole-3-acetate binds to the ATP-binding pocket in molecular docking analysis. a-b) Preferred binding position for ATP molecule (orange) on the VEGFR-2 molecule (grey). c-d) Preferred predicted binding site of indole-3-acetate (turquoise) is on the ATP-binding cavity of VEGFR-2. e-f) Overlap of the two individual models reveals a close spatial relationship between the preferred docking positions of both molecules. In the detailed view of the ATP-binding pocket, VEGFR-2 residues are shown in dark grey and labelled; the hydrogen bonds established are represented as red dashed lines.

Of note, the ten best docking solutions for indole-3-acetate on VEGFR-2 predicted the formation of the hydrogen bond with residues identified as part of the ATP-binding pocket (either Asp-1046 or Cys-919). The overlap of the preferred docking positions of ATP and indole-3-acetate highlights the existence of a close spatial relationship between both (Figure 36e-f).

Next, it was assessed whether indole-3-acetate docking might be able to prevent subsequent ATP binding in the VEGFR-2 molecule. In order to test this hypothesis, the merged structure of VEGFR-2 and indole-3-acetate was used to model the best docking position for ATP. In this case, ATP was not predicted to form hydrogen bonds with any of the residues normally involved in the ATP-binding pocket (Table 10). As visualised in Figure 37, when modelling this complex (indole-3-acetate + VEGFR-2), ATP binding is displaced from the usual catalytic site.

Table 10. Best docking solutions for ATP after previous docking of indole-3-acetate in the VEGFR-2 molecule

| Solution number | ChemPLP score | Residues                           |
|-----------------|---------------|------------------------------------|
| 1               | 40.83         | Gln-847, Lys-871, Gly-846, Gly-841 |
| 2               | 37.37         | Lys-838, Leu-840                   |
| 3               | 43.28         | Lys-838, Leu-840                   |
| 4               | 44.37         | Gln-847, Lys-871, Gly-841          |
| 5               | 34.65         | Lys-383, Leu-840                   |

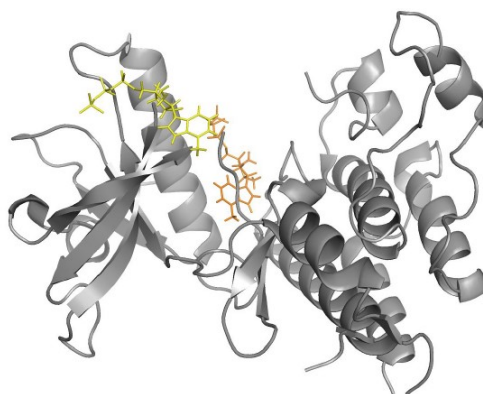


Figure 37. Indole-3-acetate prevents ATP binding in the ATP-binding pocket of VEGFR-2. Superimposition of two ATP docking models: Preferred ATP position in the absence of indole-3-acetate, in the ATP-binding pocket of VEGFR-2 (ATP molecule in orange; VEGFR-2 molecule in grey); b) After docking indole-3-acetate in its highest scoring binding site, the optimal ATP binding site moves out of the ATP-binding cavity (ATP molecule in yellow).

In summary, *in silico* docking simulations suggest that both ATP and indole-3-acetate bind in the ATP-binding pocket of VEGFR-2, and both establish a hydrogen bond with the Asp-1046 residue. Furthermore, indole-3-acetate pre-docking on the VEGFR-2 molecule is predicted to prevent ATP to bind in its usual pocket, thus suggesting an ATP-competitive inhibitory mechanism.

### 3.3. Optimisation of experimental conditions for phosphorylation assays

Before validating the inhibitory effect of indole-3-acetate on VEGFR-2 *in vitro*, the experimental conditions were optimised. Human Umbilical Vascular Endothelial cells (HUVEC), a cell line that constitutively expresses this tyrosine kinase receptor (Arnaoutova and Kleinman, 2010), were used in this experiment. The non-toxic range of indole-3-acetate was first tested by assessing the impact of the compound on cell viability. At the concentration of 1000  $\mu\text{M}$ , indole-3-acetate significantly reduced cell number (-47.1% *vs* control,  $p < 0.001$ ) (Figure 38).

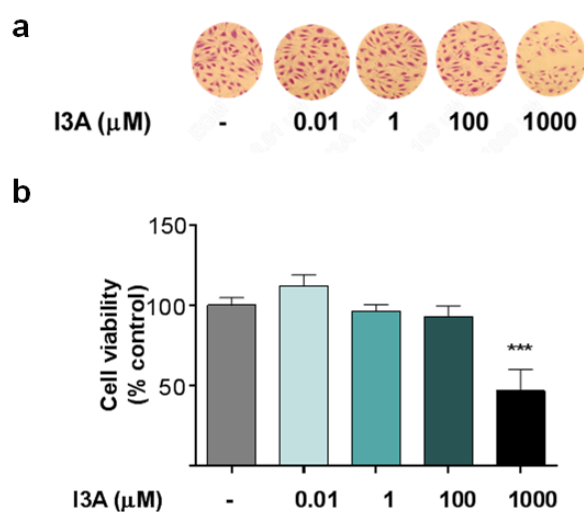


Figure 38. Indole-3-acetate affects HUVEC viability at 1000  $\mu\text{M}$ .

After 48 hours of treatment with indole-3-acetate, cells were fixed. Viability was assessed by Crystal Violet staining (Sigma Aldrich ®) and measured at 450 nm. a) Representative pictures of Crystal Violet-stained HUVEC were taken with 40x magnification. b) Quantitative results are shown as means  $\pm$  s.e.m (n=3 independent experiments, 4 replicates each). One-way analysis of variance was performed ( $p=0.0073$ ) and *post-hoc* comparisons between groups were performed with Dunnett's multiple comparisons test. \*\*\* $p=0.001$

For subsequent assays, HUVEC cells were treated with concentrations ranging from 0.01 to 100  $\mu$ M, for which no toxic effect was observed.

Next, the optimal time point to study VEGFR-2 phosphorylation was investigated. The main ligand of VEGFR-2 is the Vascular Endothelial Growth Factor A (VEGF<sub>A</sub>), which was used as agonist in the following experiments (Olsson, et al., 2006). Total VEGFR-2 expression decreased in a time-dependent manner after stimulation with VEGF<sub>A</sub>; at 10 seconds post-stimulation a reduction of 9.3% was observed ( $p<0.01$ ), followed by a progressive reduction of up to -51.7% at 240 minutes ( $p<0.0001$ ) (Figure 39a). This effect was previously described and is reportedly linked to proteolytic phenomena and removal of endosome-associated cytoplasmic domains following VEGFR-2 activation and ubiquitination (Bruns, et al., 2010). As expected, VEGF<sub>A</sub> induced a rapid increase in the Phospho(Tyr<sup>1175</sup>)-VEGFR-2 fraction, with an increase of 45.0% after 10 seconds of stimulation ( $p<0.001$ ); this increase was observed up to 60 minutes after stimulation (Figure 39b). Regarding the Phospho(Tyr<sup>1175</sup>)/total VEGFR-2 ratio, a rapid increase was also observed after 10 seconds of stimulation (+59.8%,  $p<0.01$ ) which remained significant up to 240 minutes (+117.1%,  $p<0.0001$ ) (Figure 39c).

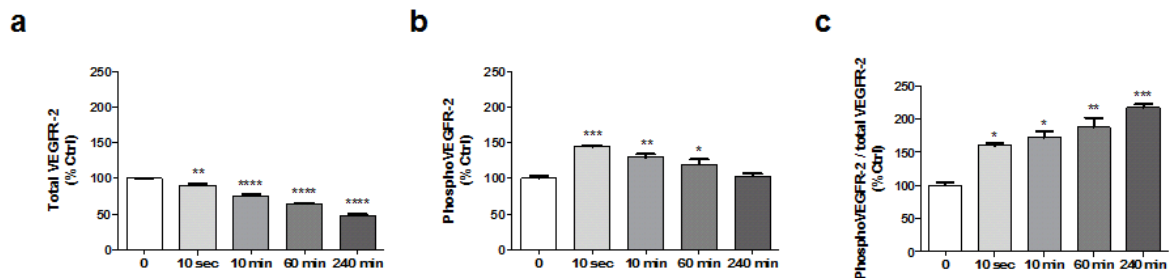


Figure 39. Kinetics of total and Phospho(Tyr1175)-VEGFR-2 expression after stimulation.

HUVEC cells were starved in basal medium for 2 hours and subsequently stimulated with VEGF<sub>A</sub> for 15 minutes. Cell lysates were collected at 10 seconds and at 10, 60 and 240 minutes. Total VEGFR-2 and Phospho(Tyr1175)VEGFR-2 were quantified by ELISA in the cell lysates. a) Total VEGFR-2 expression; b) Phospho(Tyr<sup>1175</sup>) VEGFR-2 quantification; c) Phospho(Tyr<sup>1175</sup>)-VEGFR-2/total VEGFR-2 ratio quantification. Results are shown as mean  $\pm$  s.e.m. (% of control) of one experiment (3 replicates). One-way analysis of variance was performed. Comparisons between groups were performed using Dunnett's *post-hoc* testing. \* $p<0.05$ ; \*\* $p<0.01$ ; \*\*\* $p<0.001$ ; \*\*\*\* $p<0.0001$  compared to non-stimulated cells.

The earlier time points (10 seconds and 10 minutes) were therefore considered optimal for phosphorylation assays as they show a) a significant increase in phosphorylated/total VEGFR-2 ratio; b) maximal values for the phosphorylated fraction quantification (thus providing an optimal sensitivity for the assay) and c) minimal VEGFR-2 loss, therefore minimizing artefactual increases of the phosphorylated/total VEGFR-2 ratio. For technical feasibility reasons, the 10 minutes time point was chosen to test the effect of different doses of indole-3-acetate on VEGFR-2 phosphorylation.

### 3.4. Effect of indole-3-acetate on VEGFR-2 phosphorylation

After optimisation, the effect of indole-3-acetate on VEGFR-2 phosphorylation was assessed. The experimental design is shown in Figure 40a.

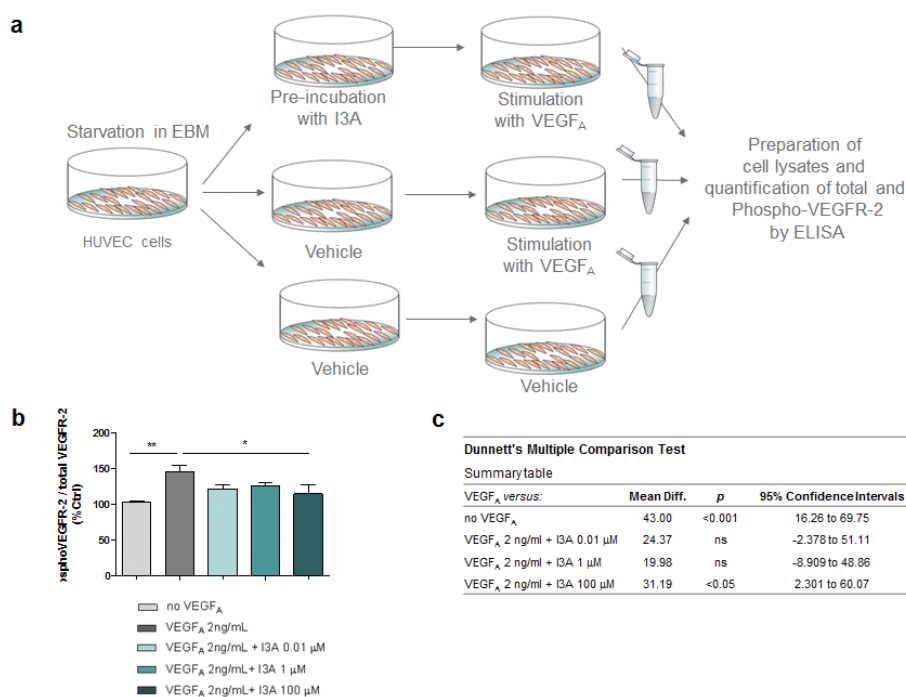


Figure 40. Effect of indole-3-acetate on VEGFR-2 phosphorylation.

a) Experimental design. HUVEC cells were starved in basal medium for 2 hours, treated with vehicle or indole-3-acetate at several doses for 15 minutes and stimulated with VEGF<sub>A</sub> for 10 minutes. Cell lysates were collected and total and phosphorylated (Tyr<sup>1145</sup>) VEGFR-2 were quantified by ELISA (Cell Signalling Technology). b) Phosphorylated VEGFR-2/total VEGFR-2 ratio quantification. Results are shown as mean  $\pm$  s.e.m. (% of control) of four independent experiments. Parametric one-way analysis of variance was performed ( $p < 0.0092$ ). c) Comparisons between groups were performed using Dunnett's *post-hoc* multiple comparisons test. \* $p < 0.05$ ; \*\*\* $p < 0.001$ . EBM: endothelial basal medium; I3A – indole-3-acetate; VEGF<sub>A</sub> – Vascular Endothelial Growth Factor<sub>A</sub>.

As expected, stimulation with VEGF<sub>A</sub> resulted in a 51.0% increase of the phosphorylated (Tyr<sup>1154</sup>)/total VEGFR-2 ratio, when compared to non-stimulated cells. Pre-treatment with indole-3-acetate at 100 µM significantly reduced VEGFR-2 phosphorylation (-29.2% compared to VEGF<sub>A</sub> alone,  $p < 0.05$ ) (Figure 40b-c).

### **3.5. Effect of indole-3-acetate on VEGFR-2 downstream signalling: AKT phosphorylation and NO production**

Based on the previous results suggesting that indole-3-acetate inhibits VEGFR-2, it was hypothesised that it would equally inhibit the kinase downstream signalling. Amongst the various signalling cascades activated by VEGFR-2 phosphorylation, the phosphoinositide-3-kinase Serine/AKT (PI3K-AKT) is particularly relevant for vascular health (Figure 41a). This signalling pathway regulates key physiological processes in vascular endothelial cells, including cell proliferation, differentiation and migration (Kang, et al., 2013). AKT also activates the endothelial Nitric Oxide Synthase (eNOS) and consequently the production of NO (Trouillon, et al., 2010), a compound that exhibits both vasodilator and pro-angiogenic properties.

In these experiments, the effect of indole-3-acetate on AKT phosphorylation and NO production was evaluated. As expected, VEGF<sub>A</sub> induced a significant increase on AKT phosphorylation (+21.1%,  $p = 0.033$ ), when compared to non-stimulated cells (Figure 41b). Consistently with its effect as a VEGFR-2 inhibitor, indole-3-acetate was able to significantly counteract VEGF<sub>A</sub>-induced phosphorylation of AKT at the concentration of 100 µM ( $p = 0.026$ ), using one-tail statistics (Figure 41b). The linear regression model shows that increasing concentrations of indole-3-acetate progressively reduce AKT phosphorylation ( $p = 0.025$ ) (Figure 41c).

The effect of indole-3-acetate on NO production was also evaluated. At 100 µM, indole-3-acetate inhibited VEGF<sub>A</sub>-stimulated production of NO in 20.4%, when compared to the vehicle ( $p = 0.037$ , one-tail t-test) (Figure 41d). In a linear regression

model, it is possible to observe that increasing concentrations of indole-3-acetate progressively reduce NO production ( $p=0.0157$ ) (Figure 41e).

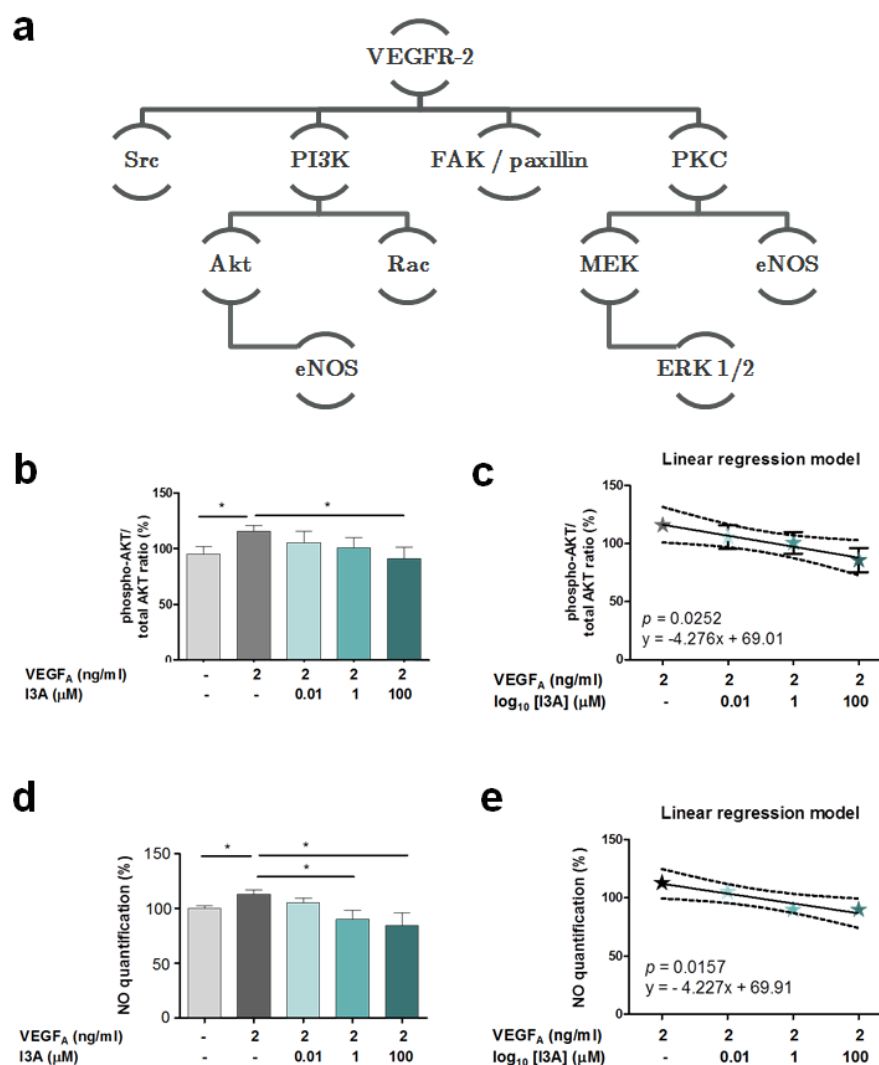


Figure 41. Effect of indole-3-acetate on VEGFR-2 downstream signalling effectors: AKT phosphorylation and NO production. a) Overview of VEGFR-2 downstream signalling (adapted from (Olsson, et al., 2006)). Results for AKT phosphorylation (b) and NO production (d) are shown as mean  $\pm$  SEM (% control) of three independent experiments; normality was assessed by D'Agostino & Pearson Omnibus test and comparisons between groups were performed using one-tail t-test. Linear regression models for AKT phosphorylation (c) and NO production (d); results are shown as mean  $\pm$  95% Confidence Intervals. \* $p < 0.05$ ; \*\*\* $p < 0.001$ . I3A – indole-3-acetate; VEGFA – Vascular Endothelial Growth FactorA

## 4. Discussion

### 4.1. High-throughput kinase screening identified indole-3-acetate as an inhibitor of the VEGFR-2

In this chapter, a high-throughput strategy was used to identify targets of indole-3-acetate in the human kinome. For the first time, indole-3-acetate was identified as an inhibitor for VEGFR-2. The physical interaction was confirmed by the calculation of the dissociation constant ( $K_D$ ) using an 11-point dose-response curve. As these methods only confirm compound-kinase binding (*i.e.* physical interaction), the functional effect was also confirmed by verification of the inhibition of the phosphorylation activity and calculation of the respective  $IC_{50}$ .

VEGFR-2 is a tyrosine kinase receptor crucial for endothelial vascular cell function and angiogenesis, which main ligand is the Vascular Endothelial Growth Factor A (VEGF<sub>A</sub>) (Olsson, et al., 2006) (*A detailed description of the cellular effects of VEGFR-2 during the angiogenic process will be provided in Chapter 7, where these outcomes will be investigated*).

Of note, indole-3-acetate showed a very specific affinity for VEGFR-2. No other physical interactions were identified between indole-3-acetate and other receptors from the VEGF or Receptor Tyrosine Kinase (RTK) families, nor with any other kinases. This is a particularly important characteristic, as achieving a highly selective effect has proven to be a major challenge regarding the use of kinase inhibitors as a therapeutic tool (Fabian, et al., 2005; Fedorov, et al., 2007). VEGFR-2 inhibitors have been used as a therapeutic tool in medical conditions characterised by an upregulation of angiogenesis (Ellis and Hicklin, 2008); unfortunately, these molecules often exhibit undesired side-effects and cellular toxicity resulting from their lack of specificity. Gefitinib, a VEGFR-2 inhibitor, has failed in clinical development due to intolerable side effects resulting from its lack of specificity (Huang, et al., 2012). Other VEGFR-2 inhibitors show cross-reactivity towards targets involved in inflammatory responses

(DiMauro, et al., 2008; Weiss, et al., 2008). Contrarily to other VEGFR-2 inhibitors, indole-3-acetate seems to have a very selective effect and, presumably, less potential for side effects.

## **4.2. Molecular docking suggests indole-3-acetate inhibits VEGFR-2 through an ATP-competitive mechanism**

Kinase inhibitors are known to exert their effects by four different mechanisms: a) competing with adenosine triphosphate (ATP), b) competing with the substrate, c) competing with both ATP and the substrate or d) binding outside the active site and inducing conformational changes that impact the active site activity (allosteric inhibition) (Posner, et al., 1994).

In molecular docking analyses, indole-3-acetate was predicted to form a hydrogen bond with the backbone nitrogen atom of the Asp-1046 residue on the catalytic site of VEGFR-2. Asp-1046 is one of the elements of the hydrogen donor-acceptor pair in the catalytic site of VEGFR-2 (Huang, et al., 2012) and facilitates the transfer of the  $\gamma$ -phosphate group (Kornev, et al., 2006). Therefore, by interacting with Asp-1046, indole-3-acetate inhibits VEGFR-2 activation by potentially impacting both mechanisms.

Molecular docking simulations do not suggest the existence of a concomitant allosteric inhibitory component. The 10 best docking solutions predicted the formation of a hydrogen bond between indole-3-acetate and residues of the catalytic domain (either Asp-1046 or Cys-919, a cysteine residue located in the hinge). None of these models predicted a physical interaction with previously described allosteric sites (*e.g.* “HYD II” or “Phe pocket” (Weiss, et al., 2008)) or other residues outside of the catalytic site. Although these molecular docking results support an inhibitory effect of indole-3-acetate on VEGFR-2, the main limitation of these findings is that they are based on *in silico* modelling and prediction. Therefore, these results provide incentive for further *in vitro* validations.

### 4.3. Indole-3-acetate inhibits VEGFR-2 phosphorylation in HUVEC

In this chapter, it was shown that indole-3-acetate is able to impair the phosphorylation of VEGFR-2 in an *in vitro* model.

The HUVEC cell line was selected for these validation studies, since it constitutively expresses VEGFR-2 and is a classic model for the study of this receptor (Arnaoutova and Kleinman, 2010). VEGFR-2 is a tyrosine kinase receptor crucial for endothelial vascular cell function and angiogenesis, whose main ligand is the Vascular Endothelial Growth Factor A (VEGF<sub>A</sub>) (Olsson, et al., 2006). Although there are multiple isoforms of VEGF<sub>A</sub>, the most common in humans, both quantitatively and qualitatively, is VEGF<sub>165</sub> (Mac Gabhann and Popel, 2008) (Shibuya, 2011). Upon VEGF<sub>A</sub> binding, VEGFR-2 undergoes dimerisation, followed by autophosphorylation of the tyrosine residues on the receptors intracellular domain (Olsson, et al., 2006). VEGFR-2 has several tyrosine phosphorylation sites bound to VEGF<sub>A</sub>, either in the kinase-insert domain (Tyr<sup>951</sup> and Tyr<sup>996</sup>), in the kinase domain (Tyr<sup>1054</sup> and Tyr<sup>1059</sup>) and in the C-terminal tail (Tyr<sup>1214</sup> and Tyr<sup>1175</sup>) (Takahashi and Shibuya, 2005). Each phosphotyrosine site recruits distinct sets of adaptor proteins, leading to distinct downstream signalling pathways (Olsson, et al., 2006; Tan, et al., 2013). Amongst these phosphosites, Tyr<sup>1175</sup> is one of the major VEGF<sub>A</sub>-dependent autophosphorylation sites in humans (Takahashi, et al., 2001). Tyr<sup>1175</sup> plays an essential role in vascular function and development; mice that express a mutated Tyr<sup>1173</sup> (the homologous residue for Tyr<sup>1175</sup> in murine) die at early embryonic stages due to vascular defects similar to the ones observed in VEGFR2<sup>-/-</sup> mice.

In line with previous literature (Shibuya, 2011), VEGF<sub>A</sub> induced VEGFR-2 phosphorylation on the Tyr<sup>1175</sup> residue. It is important to note that the increase on VEGFR-2 phosphorylation observed, although significant, was not consistently observed in other experiments performed using the same experimental conditions.

Regarding indole-3-acetate, it was shown to impair VEGF<sub>A</sub>-induced phosphorylation of the same residue. These findings are consistent with previous results

from the kinome screen and molecular docking analysis, and validate *in vitro* the inhibitory effect of indole-3-acetate on VEGFR-2. As future work, these results could be potentially strengthened by testing higher concentrations of agonist (as well as different combinations of concentrations of both VEGFR-2 and indole-3-acetate) and by trying alternative normalisation techniques (*i.e.* normalisation to total protein content).

#### **4.4. Indole-3-acetate supresses VEGF<sub>A</sub>-induced AKT phosphorylation and NO production**

In this chapter, it was demonstrated that indole-3-acetate impacts signalling pathways downstream of (Tyr<sup>1145</sup>)VEGFR-2, including AKT phosphorylation and NO production.

It has been previously documented that VEGFR-2 phosphorylation in the Tyr<sup>1145</sup> residue activates several downstream signalling pathways (Shibuya, 2011). In particular, phosphorylated (Tyr<sup>1145</sup>) binds to two SH2-domain-containing adaptor molecules: 1) phospholipase C $\gamma$  (PLC $\gamma$ ) and 2) shb (Takahashi, et al., 2001) (Holmqvist, et al., 2004). The PLC $\gamma$ /PKC pathway activates several downstream signalling effectors, including the endothelial Nitric Oxide synthase (eNOs) and the MEK/MAPK and Raf/Ras pathways (Olsson, et al., 2006) (Figure 41a). The shb/PI3K pathway results in the downstream activation of AKT (Holmqvist, et al., 2004). The PI3K-AKT pathway is particularly relevant for vascular health; it is involved in a range of signalling pathways that regulate cell proliferation, differentiation and migration (Kang, et al., 2013). By activating the endothelial Nitric Oxide synthase (eNOs), AKT also induces the production of NO, a compound that exhibits both vasodilator and pro-angiogenic properties (Trouillon, et al., 2010).

As previous results indicate that indole-3-acetate is a VEGFR-2 inhibitor, testing was performed for the possibility of a relationship in a single direction (one-tail t-test). Strategies to improve these results include a) to increase the number of experiments, b) to optimise the combination of the concentrations of VEGFR-2 and indole-3-acetate

used and c) to investigate the optimal time points to access AKT phosphorylation and NO production. Additionally, in order to have a more comprehensive overview of the impact of indole-3-acetate on (Tyr<sup>1145</sup>)VEGFR-2 downstream signalling, it would be important to clarify its putative inhibitory effect in other pathways (*i.e.* FAK/paxillin, Src and PKC/MEK/ERK).

## 5. Conclusion

In summary, indole-3-acetate was proposed, for the first time, to be an inhibitor of VEGFR-2 - a key receptor for vascular endothelial function and angiogenesis.

A high-throughput kinase screening was used to identify the human targets of indole-3-acetate in the human kinome, and revealed a very specific effect of indole-3-acetate as VEGFR-2 inhibitor. Using molecular docking analysis, it was showed that indole-3-acetate could preferentially dock in the ATP-binding site of VEGFR-2, thus exerting a potential competitive effect. The effect of indole-3-acetate was further validated in cell-based assays in HUVEC, which demonstrated that pre-incubation with indole-3-acetate impaired (Tyr<sup>1145</sup>)VEGFR-2 phosphorylation. Upon ligand binding, VEGFR-2 undergoes phosphorylation and activates signalling cascades involved in vascular function and key steps of the angiogenic process (*i.e.* AKT/NO production). Consistently with its role as a VEGFR-2 inhibitor, these preliminary results suggest indole-3-acetate inhibits VEGF<sub>A</sub>-induced phosphorylation of downstream signalling effectors (AKT), as well as the production of its respective vasoactive end-product (NO), in one-tail analyses.

By showing that indole-3-acetate is able to physically bind to VEGFR-2 and inhibit its phosphorylation activity and its downstream signalling, a novel role as a microbial metabolite with anti-angiogenic properties emerges for this compound. *In vitro* and *in vivo* validation assays are required to validate this hypothesis.







## Chapter 7

# Effect of indole-3-acetate on angiogenesis *in vitro*

### 1. Introduction

In the previous chapter (Chapter 6), a combined strategy including a high-throughput kinome scan, molecular docking modelling and cell-based assays identified indole-3-acetate as a putative competitive inhibitor of VEGFR-2.

VEGFR-2 is the main receptor involved in angiogenesis, the physiological process of growth of new blood vessels (Potente, et al., 2011; Shibuya, 2011). In the strictest sense, angiogenesis designates vessel sprouting from pre-existing vascular structures; this process is distinct from *vasculogenesis*, which refers to the formation of primitive blood vessels from angioblasts (mesoderm cell precursors) (Potente, et al.,

2011; Risau and Flamme, 1995). Angiogenesis is a fundamental process in human development, where the growth of new blood vessels provides the growing organs with the necessary oxygen supply, as well as with trophic signals that modulate morphogenesis (Carmeliet, 2005). In adulthood, endothelial cells still preserve their ability of dividing in response to pro-angiogenic stimuli in physiological processes (ovarian cycle and placental development during pregnancy). Of note, abnormal angiogenesis has been reported in a range of human diseases, including cancer, inflammatory diseases and, importantly, in several cardiometabolic disorders (*e.g.* atherosclerosis (Van Belle, et al., 1997), ischemic heart disease (Jesmin, et al., 2005; Shiojima, et al., 2005), obesity (Fukumura, et al., 2003; Rupnick, et al., 2002) diabetes (Caldwell, et al., 2005; Rivard, et al., 1999; Waltenberger, 2001) and arterial hypertension (Kubis, et al., 2002; Sane, et al., 2004)).

Angiogenesis is a multistep process, which includes specific hallmarks: 1) cell activation and matrix remodelling, 2) cell proliferation, 3) cell migration and 4) tube formation (Adams and Alitalo, 2007; Carmeliet and Jain, 2011) (Figure 42).

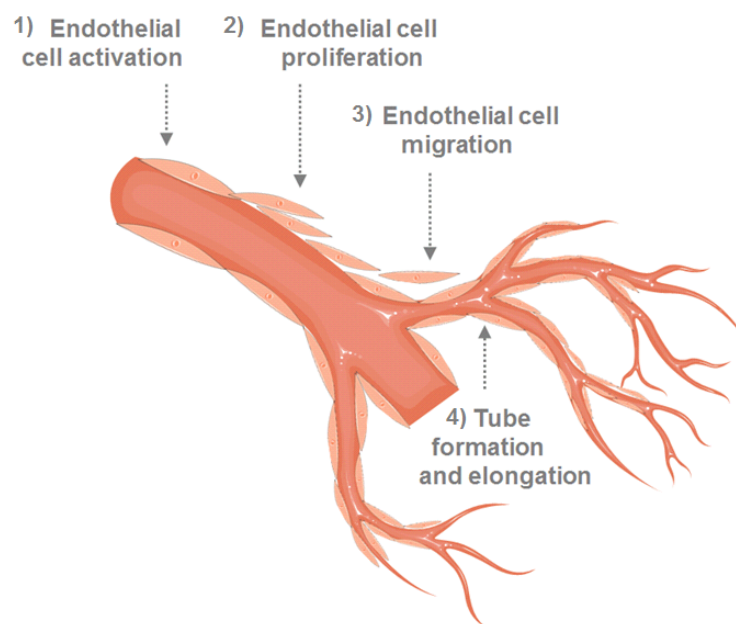


Figure 42. Hallmarks of angiogenesis.

The events leading to the formation of new vessels include a) endothelial cell activation and dissolution of the basement membrane; b) proliferation of endothelial cells; c) cell migration towards angiogenic cues, and d) tube elongation and formation of branches and loops by convergence of the newly-formed sprouts. Adapted from (Adams and Alitalo, 2007; Carmeliet and Jain, 2011) and designed using Servier Medical Art elements.

The VEGF<sub>A</sub>/VEGFR-2 system impacts most of the key steps of this process (Olsson, et al., 2006; Shibuya, 2011) (Figure 43).

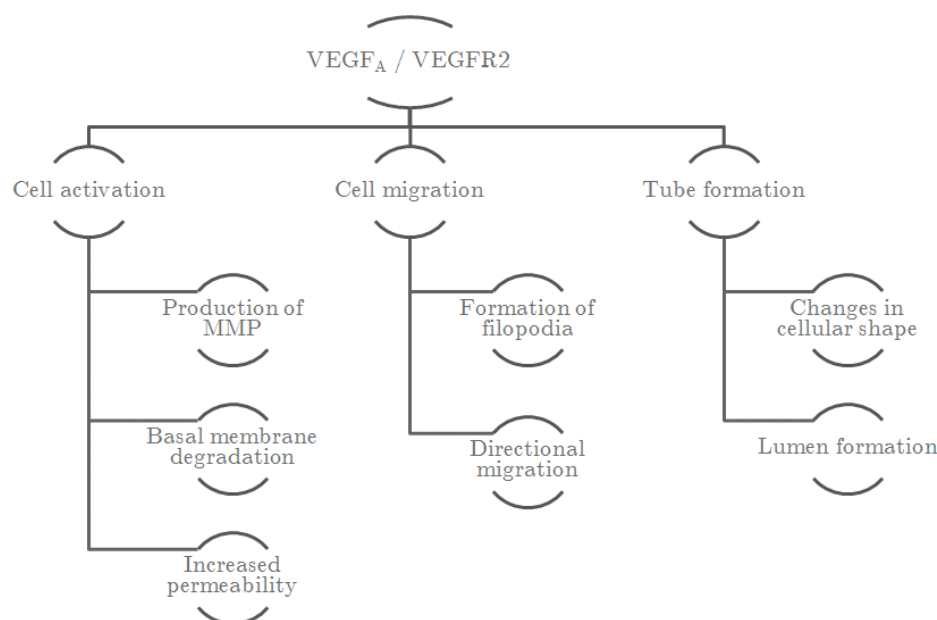


Figure 43. Role of VEGF<sub>A</sub>/VEGFR-2 in key steps of angiogenesis.

Cell activation is the first step of the angiogenic process (Eble and Niland, 2009; Herbert and Stainier, 2011). In the absence of pro-angiogenic stimuli, endothelial cells remain in a quiescent status, regulated by a low-level autocrine VEGF<sub>A</sub> signalling (Herbert and Stainier, 2011). Increased levels of pro-angiogenic stimuli (amongst which VEGF<sub>A</sub> is one of the most important) activate endothelial cells to produce matrix metalloproteinases that degrade the basement membrane (Blasi and Carmeliet, 2002; Eble and Niland, 2009; Herbert and Stainier, 2011). The degradation of the basement matrix is crucial to overcome the spatial restriction and therefore allow endothelial cell migration (Potente, et al., 2011). Subsequently, VEGF<sub>A</sub> enhances the permeability of the endothelial layer and promotes the extravasation of plasmatic proteins, thus creating a scaffold into which endothelial cells can migrate (Eble and Niland, 2009; Herbert and Stainier, 2011) (Figure 43).

VEGF<sub>A</sub> does not have a relevant effect as a mitogen in endothelial cells (Olsson, et al., 2006), but it does regulates endothelial cell migration (Figure 43). *Via* VEGFR-

2 activation, VEGF<sub>A</sub> induces the formation of long and abundant filopodia in endothelial cells and consequently the adoption of a motile phenotype (Gerhardt, et al., 2003) (Potente, et al., 2011). VEGFR-2 receptors present in filopodia are able to sense VEGF<sub>A</sub> and other chemoattractant stimuli; hence, they guide cell migration and orientate the growth of the vessel sprout (Potente, et al., 2011). Finally, the formation of tubular structures occurs, supported by endothelial cells that elongate the vascular sprout and establish adherent and tight junctions, thus ensuring the stability of the newly-formed vessel (Potente, et al., 2011). After defining its basal-apical polarity, endothelial cells contour their apical membrane with negatively charged glycoproteins; this creates an apical repulsive signal that results in the creation of the lumen (Strilic, et al., 2009). Initiation of blood flow, pericyte attachment and formation of a basement membrane provide further stabilisation of the structure (Potente, et al., 2011).

During angiogenesis, endothelial cells experience profound phenotypical changes that require the remodelling of the actin cytoskeleton (*i.e.* formation of filopodia, cell migration, formation of the vascular lumen) (Gerhardt, et al., 2003; Lamalice, et al., 2007). Downstream of VEGFR-2, the AKT/endothelial NO synthase pathway has been shown to be critically involved in these mechanisms. AKT inhibition impairs VEGF<sub>A</sub>-stimulated remodelling of the actin cytoskeleton and cell migration (Morales-Ruiz, et al., 2000; Noiri, et al., 1997; Ziche, et al., 1994).

Previous results suggest that indole-3-acetate is able to inhibit the phosphorylation of VEGFR-2 and its downstream signalling effector AKT; therefore, it was hypothesised that indole-3-acetate would exhibit anti-angiogenic properties. In this chapter, a cell-based approach was used to examine the putative inhibitory effect of indole-3-acetate in key steps of angiogenesis *in vitro*. Specific aims include:

- a) To assess the impact of indole-3-acetate on HUVEC cell migration;
- b) To assess the impact of indole-3-acetate on HUVEC tube formation.

## **2. Materials and Methods**

### **2.1. Cell culture**

HUVEC were obtained from BD Biosciences. In brief, cells were grown in collagen I-coated flasks and maintained in an atmosphere of 5% CO<sub>2</sub> at 37°C. After reaching 80% confluence, cells were trypsinised and split at the ratio of 1:2 to 1:6. Cells between passage number 3 and 10 were used for the experiments described.

### **2.2. Cell migration assay**

HUVEC cells were trypsinised and seeded in a 24-well Radius Cell Migration Assay plate (Cell Biolabs). Each well of the plate contained a 0.68 mm biocompatible hydrogel spot (Radius™ Gel) that prevents HUVEC cells attachment. Five hours after seeding, cells reached confluency and the hydrogel spot was removed. Cells were then pre-treated with either indole-3-acetate (0.01, 1 or 100 µM) or vehicle for 15 minutes, followed by stimulation with endothelial growth medium (EGM) (prepared with endothelial basal medium (EBM-2, Lonza) and endothelial cell growth supplements (EGM-2 Lonza Bulletkit), and supplemented with VEGF<sub>A</sub> at the concentration of 2ng/mL (Sigma Aldrich, LCC)). Photographs were taken hour for 6 hours using a Zeiss Axiovert 200M Live Cell Imaging Microscope (Zeiss, Jena, Germany). Migration was quantified as the percentage of the area recovered by migrating cells after hydrogel spot removal (full details in Chapter 2. “General Methodology”). Image Processing and Analysis in Java (ImageJ) software was used to perform area measurements (Schneider, et al., 2012).

### **2.3. Tube formation assay**

The Millicell Micro-Angiogenesis Inhibition Assay (EMD Millipore) was used to assess capillary tube formation. Matrigel was added to a Micro-Angiogenesis slide and

allowed to solidify at 37°C for 1 hour. HUVEC were seeded at a density of  $2 \times 10^5$  cells/well and treated either with a) vehicle, b) endothelial growth medium (EGM, prepared with EBM-2 and endothelial cell growth supplements (EGM-2 Lonza Bulletkit), and supplemented with VEGF<sub>A</sub> (Sigma Aldrich), at the concentration of 2ng/mL or c) EGM (as previously described) plus various concentrations of indole-3-acetate (0.01, 1 or 100  $\mu$ M)). After 16 hours of incubation, cells were labelled with 2  $\mu$ M of Calcein-AM for 15 minutes for 30 minutes at 37°C in 5% CO<sub>2</sub>. Photographs were taken using a 484 nm excitation and 520 nm emission filter on an Axiovert 200 microscope equipped with a 10x objective. Confocal microscopy photographs were taken simultaneously. The quantification of the network parameters was evaluated by the Angiogenesis Analyser plug-in, ImageJ software (Carpentier, 2012).

## 2.4. Statistical analysis

Normality was tested using the D'Agostino-Pearson Omnibus normality test, and parametric or non-parametric testing chosen accordingly. For cell-based assays, parametric or non-parametric one-way ANOVA was used, followed by *post hoc* testing ( $p < 0.05$  considered to be statistically significant). The number of replicates for each experiment is indicated in the individual figure legends. Data are displayed as mean  $\pm$  standard error of the mean (s.e.m) in all figures.

## 3. Results

### 3.1. Effect of indole-3-acetate on cell migration

In order to study the effect of indole-3-acetate on key steps of angiogenesis, we first investigated the effect of the compound on cell migration, one of the most critical hallmarks of the process. In brief, HUVEC cells were plated in a Radius™ 24-well plate, using EGM (Figure 44a). Each well contained a biocompatible hydrogel spot that prevented cell attachment. For 5 hours, cells were allowed to reach confluence by firmly attaching outside of the spot. After hydrogel spot removal (Figure 44b), HUVEC cells were pre-treated with either indole-3-acetate (0.01, 1 or 100  $\mu$ M) or vehicle for 15 minutes, followed by stimulation with EGM (Figure 44c-d). EGM was prepared with endothelial basal medium (EBM-2) and endothelial cell growth supplements (EGM-2 Lonza Bulletkit), and supplemented with VEGF<sub>A</sub> (Sigma Aldrich), at the concentration of 2ng/mL).

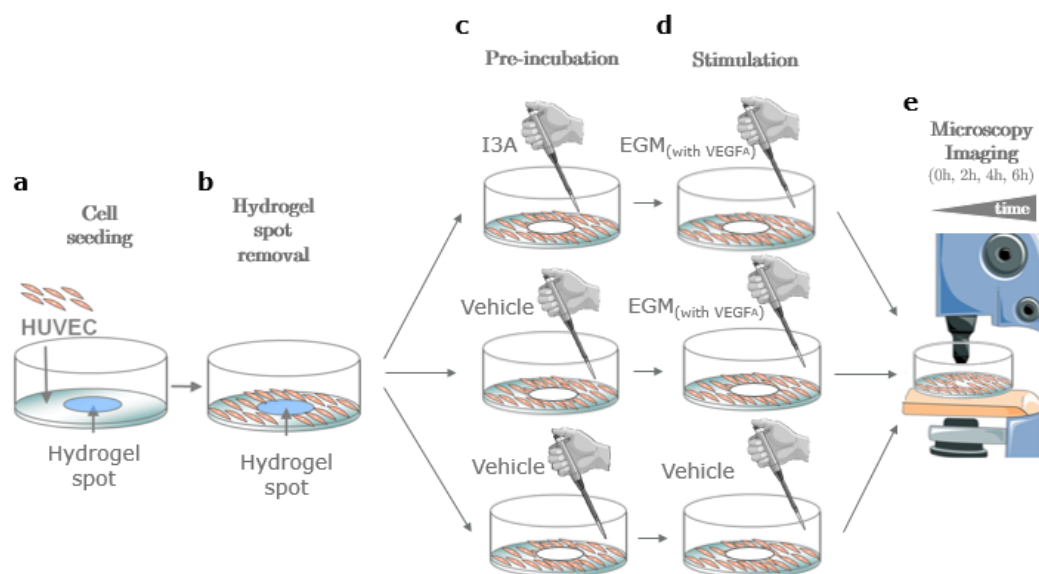


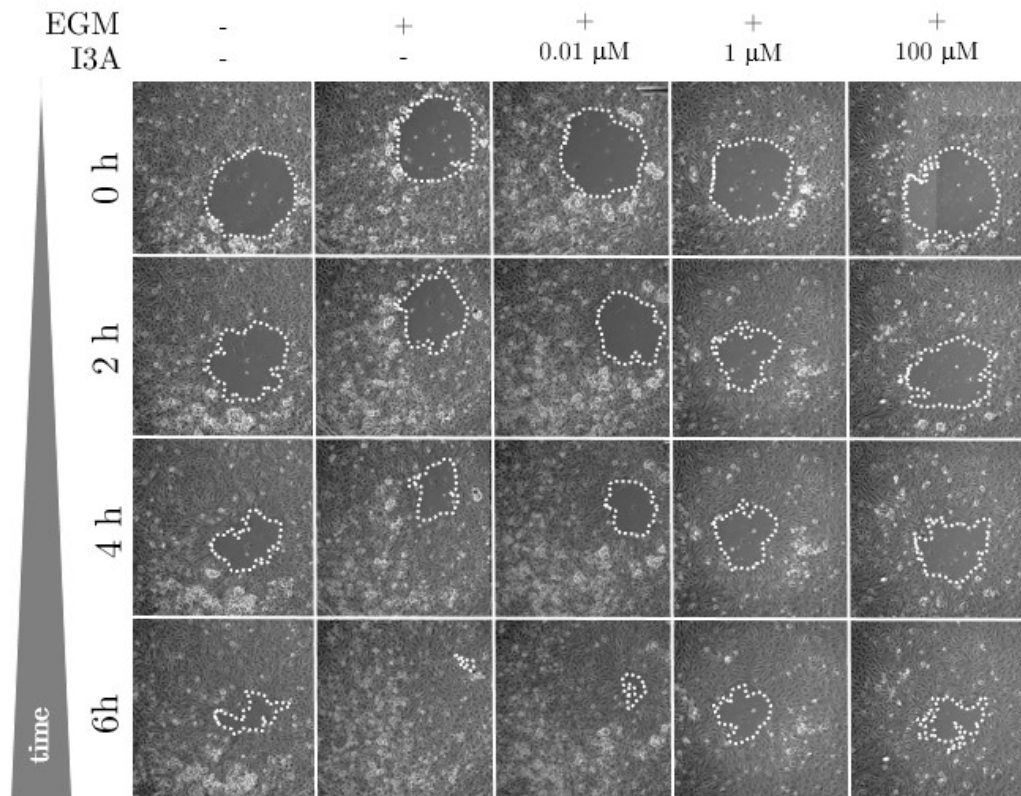
Figure 44. Assessment of HUVEC migrating activity *in vitro*.

An exclusion zone assay (Radius™, Cell Biolabs) was used. Experimental steps include a) HUVEC seeding; b) hydrogen spot removal; c-d) HUVEC pre-incubation with either indole-3-acetate or vehicle for 10 minutes, followed by stimulation with endothelial growth medium (EGM was prepared with basal medium (EBM-2) and endothelial growth factors, and supplemented with VEGF<sub>A</sub>, as described). Cells were allowed to migrate for 6h into the newly revealed area. e) After 6 hours, confocal microscopy images were taken using a FITC filter. Designed using Servier Medical Art elements. EBM-2: Endothelial Basal Medium; EGM: Endothelial Growth Medium; I3A: Indole-3-acetate; VEGF<sub>A</sub>: Vascular Endothelial Growth Factor A.

Cells were allowed to migrate for 6h and photographs were taken at 0h, 2h, 4h and 6h) (Figure 44e). The proportion of the area recovered was measured and used as a proxy for migration potential (*For further details, see section 2.2 of this Chapter*). As expected, stimulation with EGM increased cell migration, when compared to non-stimulated cells, at both 4h (54.4% vs 41.6%,  $p=0.037$ ) and 6h (55.6% vs 69.4%,  $p=0.038$ ) (Figure 45a-b).

Pre-treatment with indole-3-acetate at 100  $\mu\text{M}$  was able to counteract this effect. Cells pre-treated with indole-3-acetate (100  $\mu\text{M}$ ) exhibited a significant reduction of their migrating activity at both 4h and 6h, when compared to cells pre-treated with vehicle prior to stimulation (27.4% vs 54.2%,  $p=0.008$  and 40.1% vs 69.4%,  $p=0.019$ , respectively). Pre-treatment with lower concentrations of indole-3-acetate (0.01 and 1  $\mu\text{M}$ ) however, did not have a significant effect on cell migration at any of the time points evaluated ( $p>0.05$ ) (Figure 45a-b).

**a**



**b**

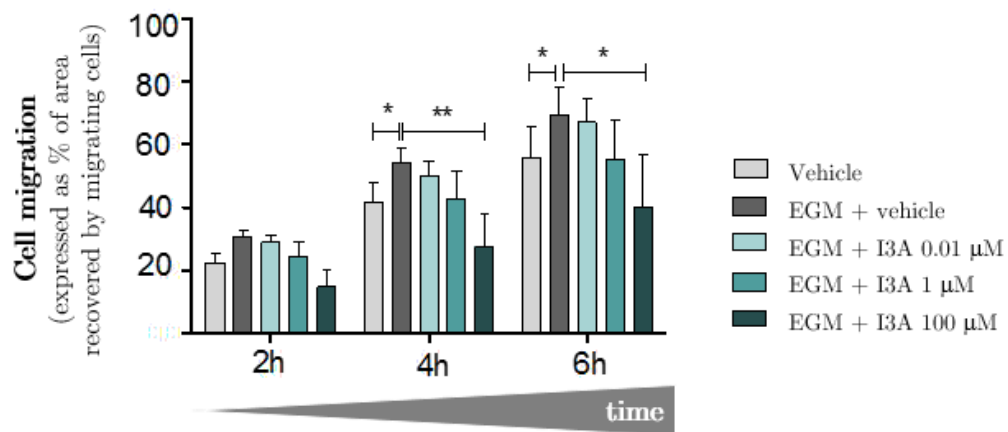


Figure 45. Indole-3-acetate inhibits cell migration in HUVEC.

a) Representative microscopic images (10x magnification). The white dashed line delimits the exclusion zone area. b) Migrating activity, expressed as the percentage of the initial area (0h) recovered by migrating cells. Results are shown as mean  $\pm$  s.e.m. of two independent experiments. One-way ANOVA was performed and Fisher's *post-hoc* testing was used for comparisons between groups. \* $p < 0.05$ ; \*\* $p < 0.01$ ; EGM: Endothelial Growth Medium; I3A - indole-3-acetate.

### 3.2. Effect of indole-3-acetate on tube formation

In this experiment, the effect of indole-3-acetate on endothelial cell tube formation was evaluated. In brief, endothelial cells were seeded in a gelled basement membrane extract (Matrigel), which provides a structure for cells to attach, migrate towards each other and finally align in order to form tubular structures (Arnaoutova, et al., 2009). HUVEC cells were then stimulated with EGM (prepared with EBM-2 and endothelial cell growth supplements (EGM-2 Lonza Bulletkit), and supplemented with VEGF<sub>A</sub> (Sigma Aldrich), at the concentration of 2ng/mL), with or without indole-3-acetate (Figure 46a). As recommended by the manufacturer, a known inhibitor of tube formation (sulforaphane) was used to confirm that the experimental system was responsive to inhibitory stimuli (Arnaoutova and Kleinman, 2010).

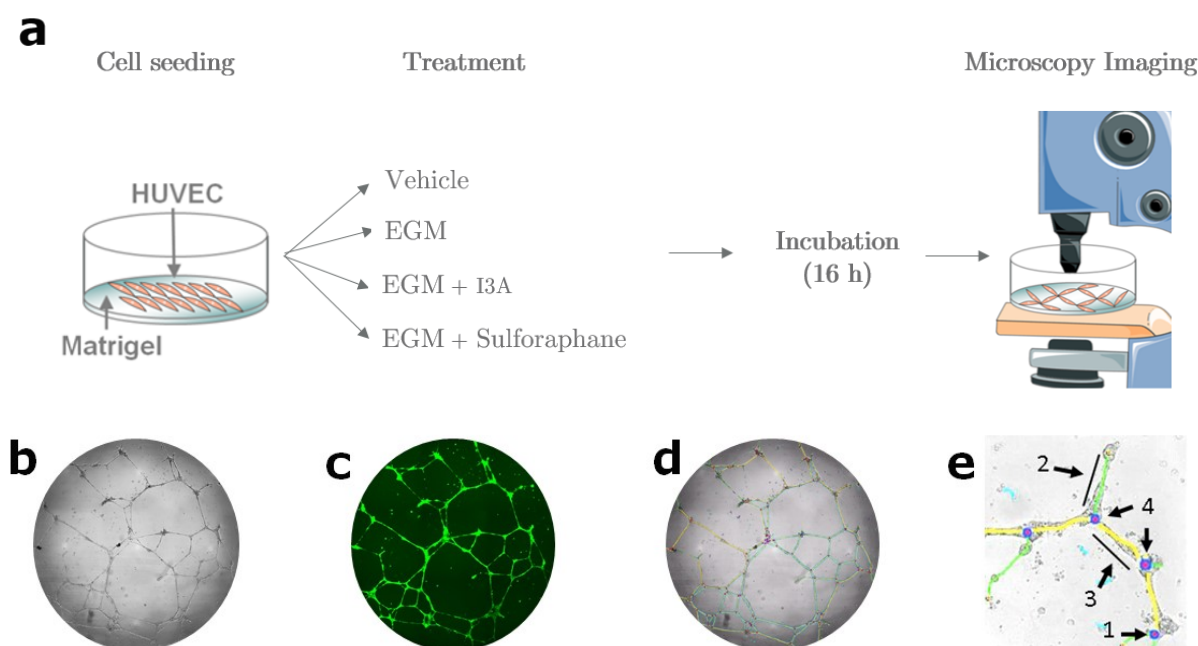


Figure 46. Assessment of tube formation by HUVEC *in vitro*.

a) Experimental design. b-e) HUVEC stimulated with EGM were used in this figure as an example to describe the constitutive elements of the network. Confocal microscopy (b) and fluorescence microscopy, after Calcein-AM staining (c). d) Confocal image processed by ImageJ Angiogenesis Analyser. e) Detailed description of the main variables quantified by ImageJ Angiogenesis Analyser: *nodes* (1) are identified as pixels with at least 3 neighbours and give rise to *branches* (terminal portions of the tree) (2). A *segment* (3) is a non-terminal portion of the tree, delimited by two *junctions* (4). The definition of *piece* includes both terminal and non-terminal portions of the tree, *i.e.* branches (2) and segments (3). EGM: Endothelial Growth Medium (as previously described); I3A: Indole-3-acetate.

After 16h of incubation, HUVEC cells formed stable tubular structures. Calcein-AM staining was then performed and microscopic images were taken and processed using ImageJ Angiogenesis Analyser software (Carpentier, 2012), automatically quantifying the constitutive elements of the tubular network (Figure 46d-e). The constitutive elements describe the network regarding a) the number of connecting structures (number of nodes, junctions, branches and pieces) and b) the length of the tubular network (total length, total branches length and, total branching length [sum of length of pieces]).

### 3.2.1. Exploratory approach using unsupervised multivariate methods

In order to understand the contribution of the individual variables to the model, all variables (*i.e.* constitutive elements of the network) were used to build a Principal Component Analysis (PCA) model, which explained 94.4% of the variability (2 Principal Components,  $R^2 = 0.944$ ). In the scatter scores plot, it is possible to observe a clear separation between groups, as well as the effect of increasing doses of indole-3-acetate (yellow arrow) in the tubular network (Figure 47a). The loadings score plot shows the global contribution of the variables to the model (Figure 47b).

In order to identify which variables contribute most to the model, the cumulative  $R^2$  of the individual contributions to the PCA model were plotted (Figure 47c). Five variables exhibited a  $R^2 > 0.90$ , including number of nodes, number of junctions, number of branches, total length, total branching length and total branches length (Figure 47c). The impact of indole-3-acetate on these variables will be further explored in the next section using univariate statistics. The effect of indole-3-acetate is also observed in an O2PLS-DA model ( $R^2_{Y_{\text{hat}}} = 0.338$ ,  $Q^2_{Y_{\text{hat}}} = 0.139$ ) (Figure 47d).

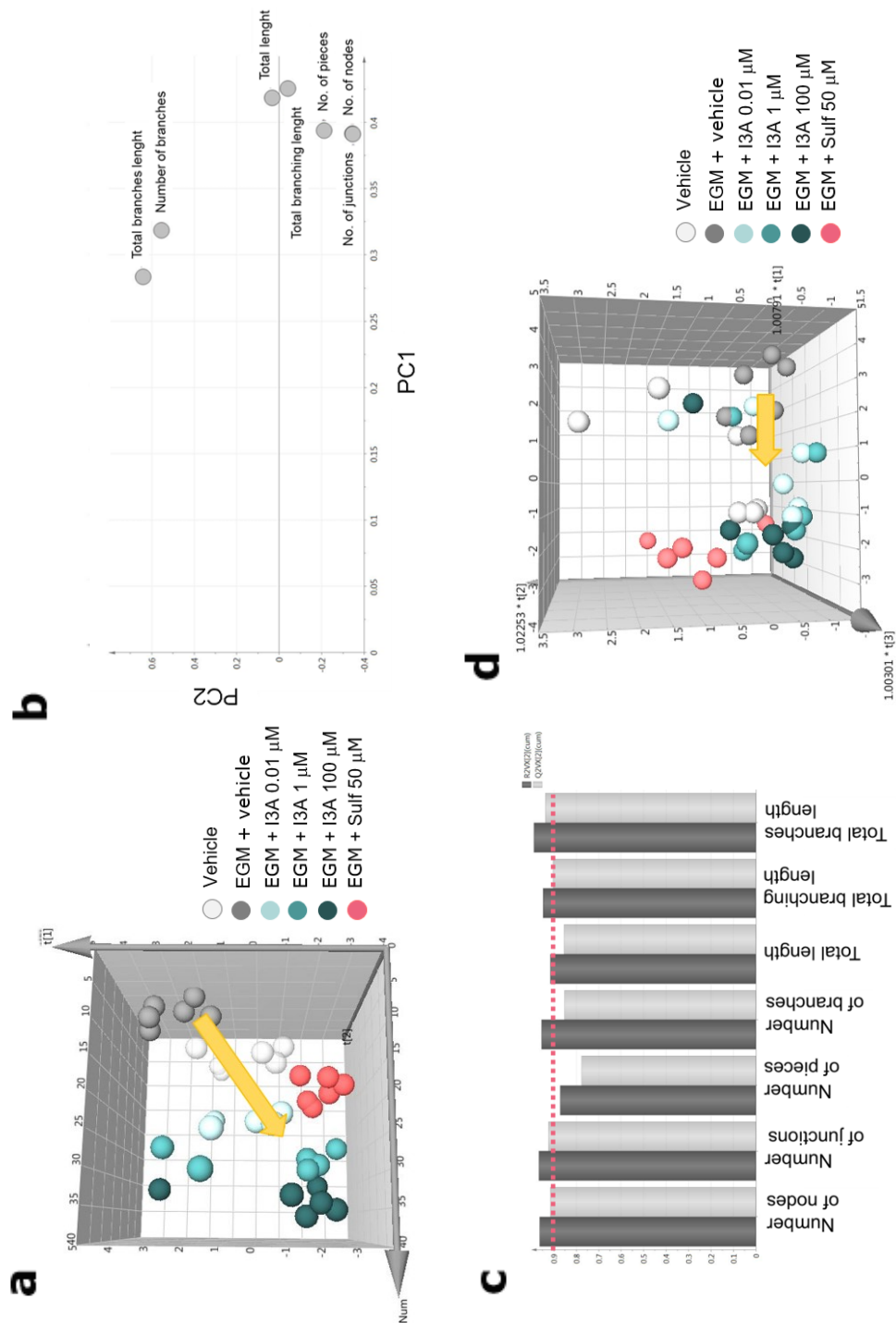


Figure 47. Unsupervised multivariate analysis by PCA and O2PLS-DA: an exploratory approach to study the effect of indole-3-acetate on tube formation *in vitro*. a) PCA 3D Scores plot; b) PCA loadings plot; c) Contribution of individual variables to the PCA. The yellow dashed represents the threshold for  $R^2=0.90$ . d) O2PLS-DA model.

### 3.2.2. Univariate analysis of the constitutive elements of the network

In order to quantify the effect of indole-3-acetate on tube formation, the above-mentioned variables were individually measured and univariate analysis was performed. HUVEC cells formed tubular networks in all tested conditions, except when treated with sulforaphane, a known angiogenesis inhibitor; in this case, HUVEC cells remained rounded as solitary cells, thus validating the responsiveness of this assay to inhibitory stimuli (Figure 48a).

As expected, stimulation with EGM significantly increased the number of connecting structures of the tubular network; the number of nodes, junctions and pieces increased by 59.7%, 80.2% and +52.6%, respectively ( $p < 0.05$  for all) (Figure 48b-d). Indole-3-acetate was able to counteract these effects, both at 1 and 100  $\mu\text{M}$  ( $p < 0.05$ ); cells with indole-3-acetate at 100  $\mu\text{M}$  exhibited a significant reduction on the number of nodes (-52.4%,  $p < 0.01$ ), number of junctions (-51.5%,  $p < 0.01$ ) and number of pieces (-49.2%,  $p < 0.01$ ), when compared to control (Figure 48b-d). For all these parameters, treatment with lowest doses of indole-3-acetate (0.01  $\mu\text{M}$ ) did not show a significant effect ( $p < 0.05$ ) (Figure 48b-d). Indole-3-acetate was also able to reduce parameters regarding the length of the tubular network; but, in this case, a significant effect was only observed at the highest dose tested (100  $\mu\text{M}$ ) (Figure 48e-g). Cells treated with indole-3-acetate (100  $\mu\text{M}$ ) exhibited a reduction on the total length, total branching length and total branches length of the tubular network (-37.3%, -50.9% and -55.9%, respectively; all  $p < 0.05$ ). Treatment with lower doses of indole-3-acetate (0.01 and 1  $\mu\text{M}$ ), however, did not have a significant effect (Figure 48e-g).

Altogether, these results demonstrate that indole-3-acetate is able to inhibit the ability of HUVEC cells to form tubular structures *in vitro*, as quantitatively demonstrated by its negative impact on the number of connections and on the length of the tubular network (Figure 48h).

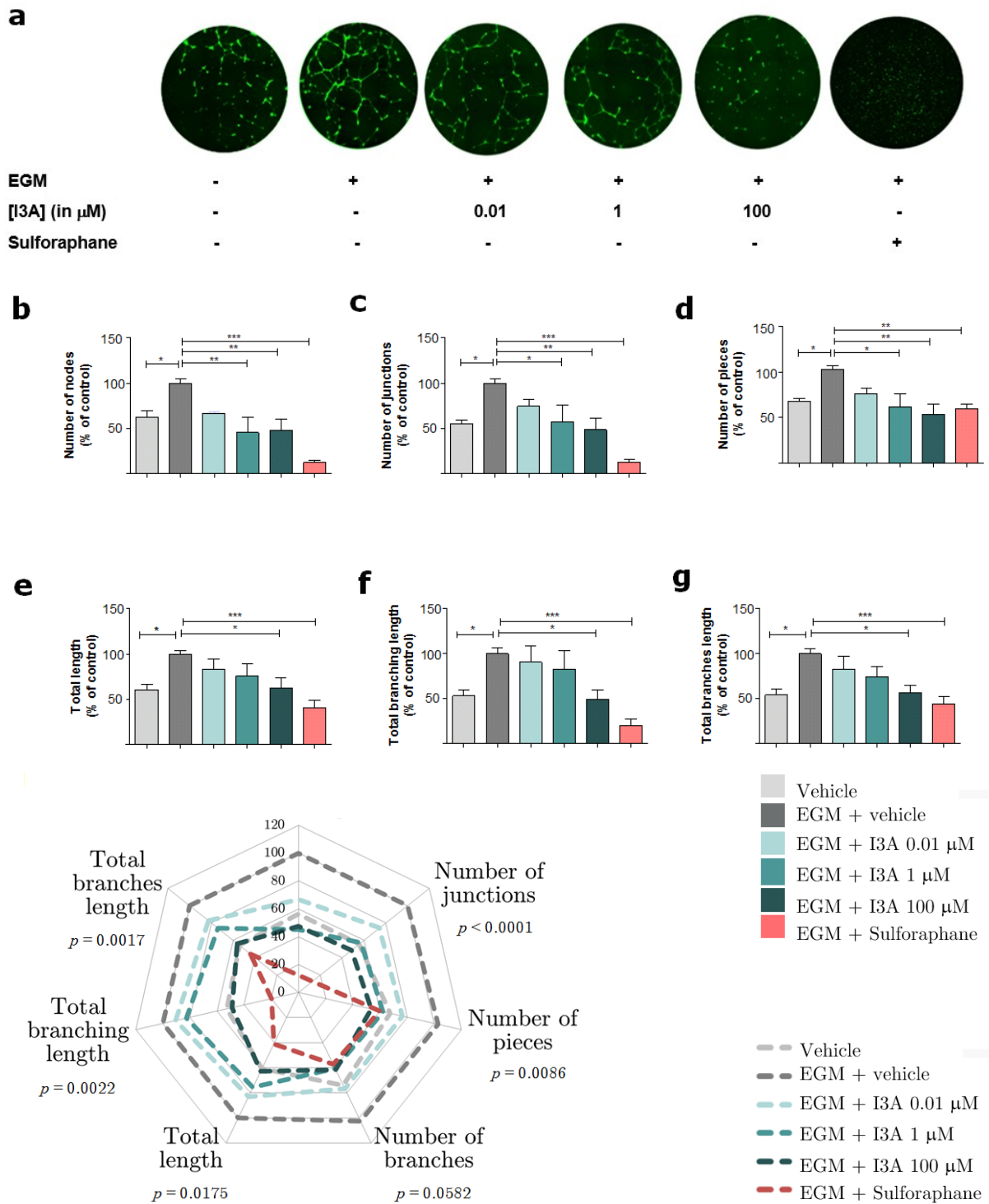


Figure 48. Indole-3-acetate inhibits tube formation in HUVEC.

a) HUVEC cells were seeded in Matrigel-coated wells and treated as previously described. b-g) Results are shown as mean  $\pm$  s.e.m. of 4 independent experiments. One-way ANOVA was performed and Dunnett's *post-hoc* testing was used for comparisons between groups. h) Radar plot providing an overview of the mean values for each one of the variables and respective *p* values for one-way ANOVA \**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001. I3A - indole-3-acetate; EGM - Endothelial Growth Medium; Sulf - Sulforaphane.

## 4. Discussion

In the previous chapter, molecular docking analyses suggested that indole-3-acetate is able to physically bind to VEGFR-2, and inhibit both its phosphorylating activity and its downstream signalling (AKT/NO). Altogether, these results suggested that indole-3-acetate could exhibit anti-angiogenic properties. In this chapter, this effect of indole-3-acetate was tested *in vitro*.

In this chapter, well-established cell-based assays were used to demonstrate the inhibitory effect of indole-3-acetate on HUVEC cells on cell migration and tube formation. As previously shown, indole-3-acetate at 1000  $\mu\text{M}$  significantly reduces cell viability (see Chapter 6); also, the maximum concentration of indole-3-acetate detected in blood is in the micromolar range (Calaf, et al., 2011; Duranton, et al., 2012). Therefore, in order to best mimic physiological conditions, the concentrations used in these experiments ranged from 0.01 to 100  $\mu\text{M}$ .

### 4.1. Indole-3-acetate inhibits cell migration *in vitro*

As previously described, angiogenesis is a multistep process, that includes a) cell activation and basement membrane disruption, b) cell migration and proliferation and c) formation of tubular structures. The preferred initial strategy to investigate the putative anti-angiogenic properties of a given compound is to test its effects in one or several of these steps *in vitro* (Staton, et al., 2004) (Auerbach, et al., 2003).

In the first study, the effect of indole-3-acetate on cell migration was assessed using a “cell exclusion zone assay”. Cell exclusion zone assays are a novel approach to overcome the disadvantages of the classic scratch assays or transmembrane assays (Kam, et al., 2008; Mastuyugin, et al., 2005; Staton, et al., 2009; Staton, et al., 2004). Cell exclusion zone assays do not induce cell damage (Poujade, et al., 2007) and provide an accurate and precise area to study cell migration, thus offering robust and reproducible data (Hulkower and Herber, 2011). Other advantages of this novel method include the possibility to use a more physiological migrating surface (*e.g.* collagen I

coating) and the possibility to perform real-time measurements (Hulkower and Herber, 2011).

In this experiment, it was established for the first time that indole-3-acetate (100  $\mu\text{M}$ ) significantly reduces the migrating activity of HUVEC; these results are consistent with its putative role as a VEGFR-2 inhibitor. Several *in vitro* studies have demonstrated that other known VEGFR-2/VEGF<sub>A</sub> inhibitors (e.g. bevacizumab) exert their anti-angiogenic properties, at least partially, by significantly impairing cell migration (Hein and Graver, 2013). Interestingly, other indolic compounds (indole-3-carbinol and indoxyl sulfate) were also shown to reduce cell migration in endothelial cells (Wu, et al., 2005) (Dou, et al., 2004). However, there is no published evidence suggesting that these effects are mediated through VEGFR-2 inhibition.

Although a critical event, cell migration is only one step amongst several required for the formation of new vascular sprouts. Therefore, in order to demonstrate the anti-angiogenic effect of a given compound *in vitro* it is recommended to assess its effect in more than one mechanistic step (Auerbach, et al., 2003).

## 4.2. Indole-3-acetate inhibits tube formation *in vitro*

In this study, the effect of indole-3-acetate on tube formation was assessed. Tube formation is a complex phenotype, that requires endothelial cells to synthesise proteases, to digest the matrix (*e.g.* Matrigel), to migrate towards each other and, finally, to align in order to form tubular structures - thus mimicking a pseudo-capillary network *in vitro* (Arnaoutova, et al., 2009; Arnaoutova and Kleinman, 2010; Grant, et al., 1991; Haralabopoulos, et al., 1994). Hence, the tube formation assay complements the previous cell migration assay and provides a more comprehensive assessment of angiogenesis *in vitro* (Arnaoutova, et al., 2009; Arnaoutova and Kleinman, 2010; Haralabopoulos, et al., 1994). Although being the most specific assay to test angiogenesis *in vitro*, some considerations shall be discussed regarding the tube formation assay. First, uneven surfaces affect the response of the tubular network to pro-angiogenic factors; therefore, the setting up of the assay shall be performed

carefully, ensuring an even coating in each well (Liu, et al., 2002). In this study, these potential biases were minimised by excluding all the wells in which coating was uneven and/or bubbles were present. Second, it is important to assure an unbiased quantitative evaluation of the vascular network, ensuring that tubular structures are adequately distinguished from the background and/or undifferentiated proliferating islands (Auerbach, et al., 2003). To overcome this pitfall, an automated tool (Angiogenesis Analyser for Image J) was used in this study to automatically extract unbiased characteristic information of the network (Carpentier, 2012).

In this study, it was shown for the first time that indole-3-acetate (at 100  $\mu$ M) is able to significantly reduce several quantitative parameters of the pseudo-capillary network formed by HUVEC *in vitro*. In particular, the compound significantly reduced the length of the network (total length, the total branching length and total branches length of the tubular network) and the number of connecting structures (nodes, junctions and pieces), when compared to vehicle-treated cells. Although the same decreasing effect on the number of connections was observed when treating HUVEC with a lower dose (1  $\mu$ M), no significant effects were observed on tubular length parameters. These results are consistent with the previously reported effects of indole-3-acetate on cell migration, where a significant inhibitory effect was observed at 100  $\mu$ M. By demonstrating that indole-3-acetate is able to significantly inhibit HUVEC tube formation *in vitro*, these results complement the previously described inhibitory effect on cell migration - and reinforce its putative role as a gut microbial metabolite with anti-angiogenic properties.

Although previous pharmacological inhibition assays and molecular docking analyses suggest that I3A anti-angiogenic effect might be directly mediated through a competitive inhibition of VEGFR-2, further experiments are required to explore the contribution of this and other putative receptors. Small interfering RNA (siRNA) / short hairpin RNA (shRNA) experiments targeting other putative receptor involved (*i.e.* AHR) might be useful to clarify and/or exclude other underlying mechanisms involved. The impact of indole-3-acetate on VEGFR-2 would be unequivocally demonstrated by a cell-based model with mutagenesis on specific amino acid residues of the receptor that bind the metabolite. In this case, the absence of anti-angiogenic

effect in the presence of indole-3-acetate would validate that the compound exerts its effect through VEGFR-2 inhibition.

Other indolic compounds (indole-3-carbinol and indoxyl sulfate) were previously shown to inhibit tube formation *in vitro*; however, as previously mentioned, no target was proposed for their inhibitory effects (Wu, et al., 2005). Due to the structural similarities between molecules, it would be relevant to also screen these molecules as potential VEGFR-2 inhibitors.

### 4.3. Angiogenesis assays *in vitro*: advantages and pitfalls

Angiogenesis assays *in vitro* present several advantages: they can be performed rapidly and generate data that allow the quantitative analysis of the angiogenic process in cell-based models (Auerbach, et al., 2003). An approach to maximize the information yielded by *in vitro* assays is performing several complementary assays *in vitro*, as described in this Chapter (Staton, et al., 2009; Staton, et al., 2004). Furthermore, the value of *in vitro* assays can be improved by safeguarding that the conditions used are as close as possible to the ones observed *in vivo* (Auerbach, et al., 2003). In order to accomplish this, in these studies, physiological concentrations of the compound and specific coatings were used (*e.g.* Collagen-I, Matrigel).

Although *in vitro* assays provide critical information on the inhibitory effect of indole-3-acetate on angiogenesis, it is important to note that these assays represent but a first step to validate the anti-angiogenic properties of this compound. Regardless of the adjustments performed, *in vitro* assays are carried out in conditions that, globally, are very different from physiological. Consequently, irrespective of the amount of information generated by *in vitro* tests, *in vivo* tests are an absolute requirement for the accurate evaluation of regulators of angiogenesis.

## 5. Conclusion

This chapter demonstrated that indole-3-acetate, at the concentration of 100  $\mu\text{M}$ , is able to significantly reduce the migrating activity of HUVEC. Consistently, indole-3-acetate also reduced tube formation, by decreasing both the length of the network and the number of connections established observed. Altogether, these results demonstrate that indole-3- acetate at 100  $\mu\text{M}$  is able to impair the development of the pseudo-capillary network *in vitro*.

These results are consistent with the previous chapter's findings suggesting that indole-3-acetate is a VEGFR-2 inhibitor, inhibiting both the phosphorylation of the receptor and downstream signalling pathways relevant for the angiogenic process (Chapter 6). Altogether, these results support the role of indole-3-acetate as a gut microbial metabolite with anti-angiogenic properties. However, *in vivo* tests are necessary to validate these findings, as will be explored in the next chapter.







## Chapter 8

# Effect of indole-3-acetate on angiogenesis *in vivo*

### 1. Introduction

Angiogenesis is a key process for human development and health – and its dysregulation is associated with a range of cardiometabolic disorders (*e.g.* atherosclerosis, ischemic heart disease and post-ischemic revascularisation, obesity, diabetes and arterial hypertension) (Caldwell, et al., 2005; Fukumura, et al., 2003; Jesmin, et al., 2005; Kubis, et al., 2002; Rivard, et al., 1999; Rupnick, et al., 2002;

Sane, et al., 2004; Shiojima, et al., 2005; Van Belle, et al., 1997; Waltenberger, 2001) (Brill, et al., 2005)).

In the previous chapter, it was demonstrated that indole-3-acetate is able to inhibit key steps of angiogenesis *in vitro* (*i.e.* cell migration and tube formation). Although *in vitro* assays are the first validating step when assessing the pro- or anti-angiogenic effect of a given compound (Auerbach, et al., 2003), it is not clear to which extent they predict responses *in vivo* (Arnaoutova, et al., 2009; Staton, et al., 2009). *In vivo*, angiogenesis involves primarily endothelial cells but is further controlled by the surrounding microenvironment (Rahat, et al., 2014). The relationship and cross-talk between endothelial cells and immune cells, fibroblasts, pericytes and the extracellular matrix, affects gene expression and phenotypic distinction - and, finally, determines whether endothelial cells survive and proliferate, or enter apoptosis (Jung, et al., 2002). *In vivo* tests are, therefore, an absolute requirement for the accurate evaluation of regulators of angiogenesis (Staton, et al., 2009; Tahergorabi and Khazaei, 2012).

*In vivo* angiogenesis models are based on the natural development of angiogenesis in different biological matrices, including the rat cornea ("*corneal angiogenesis assay*"), the chick extraembryonic membrane ("*chick chorioallantoic membrane (CAM) assay*") or in subcutaneous plugs of mice sarcoma extract ("*Matrigel plug assay*") (Tahergorabi and Khazaei, 2012). Amongst the *in vivo* assays to study angiogenesis, the Matrigel subcutaneous plug assay is a fast and effective approach to screen stimulators or inhibitors of angiogenesis (Fridman, et al., 2012; Tahergorabi and Khazaei, 2012). Furthermore, this assay allows a direct validation of the observed inhibitory effect of indole-3-acetate on tube formation *in vitro*, an experiment also performed in Matrigel.

Matrigel consists of a mixture of extracellular matrix and growth factors obtained from Engelbreth-Holm-Swarm mice sarcoma (Passaniti, et al., 1992). Pro- or anti-angiogenic compounds can be added to cold liquid Matrigel (<4°C), which is subsequently injected subcutaneously in the dorsal or ventral region of the mice (Passaniti, et al., 1992). At body temperature, Matrigel becomes a solid gel (the

“*Matrigel plug*”), into which endothelial cells migrate, thus forming vessel-like structures (Baker, et al., 2006; Malinda, 2001). Importantly, as the Matrigel plug is injected into the subcutaneous tissue (“*hypodermis*”), its close contact with fibroblasts, immune cells, adipose cells and macrophages mimics closely the microenvironment of physiological angiogenesis *in vivo* (Tahergorabi and Khazaei, 2012).

In this chapter, we tested the effect of indole-3-acetate *in vivo* using the subcutaneous Matrigel plug. In particular, we investigated the impact of indole-3-acetate on a) plug weight (as an indirect measure of vascularisation (Bhat, et al., 2013)) and b) on the area of the endothelial vascular network developed within the plug.

## 2. Materials and Methods

### 2.1. Animal studies

C57BL/6J mice (Charles River) were weaned at 3 weeks, housed in groups of 6 animals, maintained under standard breeding conditions (12 h/12h light-dark cycle) and fed *ad libitum* a chow diet (B&K Universal, Hull, U.K.) At day 0, 8-week-old mice were anaesthetised with propofol to undergo Matrigel subcutaneous injection (ForeneH, Abbott). All procedures were carried out in accordance with the U.K. Home Office guidelines on animal welfare.

### 2.2. Matrigel *in vivo* plug assay

Matrigel was prepared with 20 U/mL of heparin, 150 ng/mL of mouse recombinant VEGF<sub>A</sub> (VEGF<sub>A</sub>165, Sigma Aldrich Co, LCC), with or without indole-3-acetate (100 µM). 300 µL of liquid Matrigel with or without VEGF<sub>A</sub> (100 µM) were subcutaneously injected into the right and left flanks of each mouse, respectively. Matrigel was allowed to polymerize into a solid gel by body temperature. After 7 days, the animals were sacrificed by cervical dislocation and the Matrigel plugs were

harvested, weighted and photographed. Following fixation in 10% buffered formalin, the plugs were embedded in paraffin. Sections from each sample were cut at a thickness of 5  $\mu\text{m}$  using a standard rotary microtome (Leica Biosystems).

### **2.3. Immunohistochemistry**

Four of 5  $\mu\text{m}$  sections were transferred onto pre-cleaned glass slides (Mercedes Medical, Sarasota, FL, USA). Dried slides were incubated at 65°C for 1h and used for haematoxylin and eosin (H&E) staining or immunohistochemistry to observe the nuclei of infiltrating cells. For immunohistochemistry, sections were deparaffinised and blocked for 30min with 10% normal goat serum. Slides were incubated with primary anti-CD31 antibody (Abcam, UK) (dilution 1:50) overnight at 4°C. The secondary antibody (Alexa Fluor 488-AffiniPure Goat Anti-Armenian Hamster, Jackson ImmunoResearch, PA, USA) was incubated for 1h at room temperature. Cell nuclei were counterstained with 4'-6-Diamidino-2-phenylindole (DAPI). DAPI in mounting medium, before assembling the slide.

### **2.4. Imaging protocol**

CD31 (or platelet endothelial cell adhesion molecule-1 (PECAM-1)), a marker for endothelial cells, was used to quantify angiogenesis (Muruganandham, et al., 2006; Ragel, et al., 2007). CD31<sup>+</sup> area was quantified by considering three random sections of the peripheral region of each Matrigel plug, at 40x magnification. Fluorescent images were obtained using a Nikon Eclipse E600 fluorescent microscope coupled to a RGB-MS-C micro colour camera. Image J software was used for image processing and quantification, according to the protocol described in Figure 49.

**a**

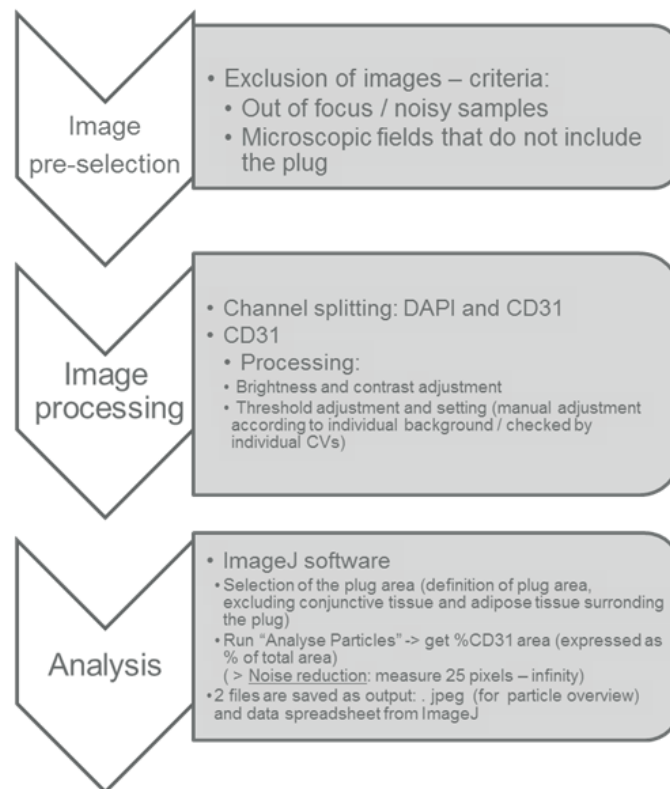
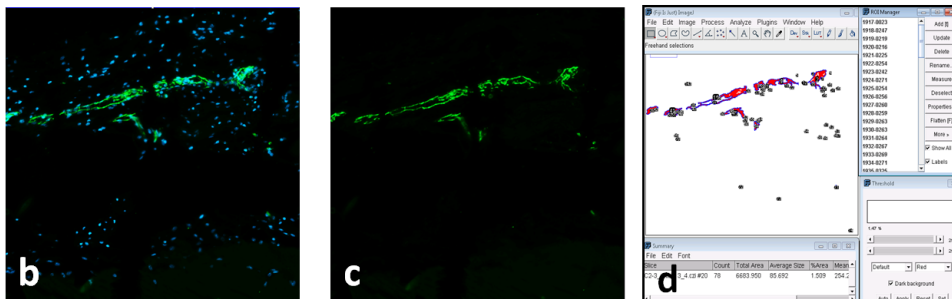


Figure 49. Microscopic imaging and analysis protocol for the quantification of the CD31<sup>+</sup> area. a) Protocol for imaging analysis. b) Details of the staining for DAPI (blue) and CD31 (green). c) CD31 staining and d) percentage of CD31<sup>+</sup> -stained area / total,, as assessed by ImageJ quantification (Angiogenesis Analyser plug-in) (Carpentier, 2012). All images were captured at 40x magnification.

## 2.5. Statistical analysis

Normality was tested using D'Agostino & Pearson Omnibus normality test, and parametric or non-parametric testing chosen accordingly. This study included three independent experiments, with a total number of nine mice. Paired t-test was used to compare groups. Two-way ANOVA was used to quantify the percentage of variation explained by indole-3-acetate and individual effect. Data are displayed as mean  $\pm$  s.e.m in all figures. A  $p$ -value of 0.05 was used for statistical significance.

## 3. Results

In brief, 300  $\mu$ L of liquid Matrigel (150 ng/mL of mouse recombinant VEGF<sub>A</sub>, 20 U/mL of heparin (Sigma Aldrich Co, LCC) with indole-3-acetate (100  $\mu$ M) or vehicle were subcutaneously injected, respectively, into the right and left flanks of each C57BL/6 mouse (Figure 50a). After 7 days, the animals were sacrificed and Matrigel plugs were harvested, weighted and fixed. Figure 50 describes the experimental design and the main histologic characteristics observed in a control plug. During this 7-day experiment, mouse endothelial cells migrated into the plug (Figure 50b) and formed new vascular structures, which can be observed in further detail in Figure 50c-d.

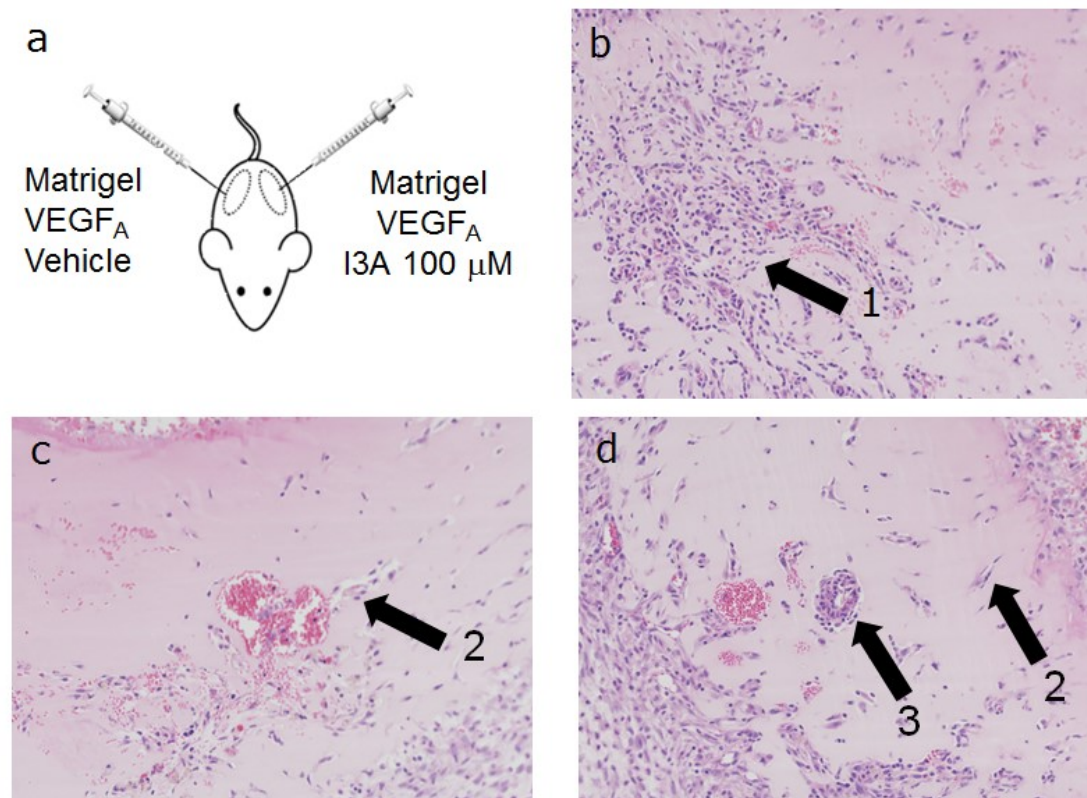


Figure 50. Matrigel subcutaneous plug assay (control plug).

a) Experimental design using dual implantation for paired statistics. b-d) Microscopic details of the vascularization network at day 7 in the control plug (H+E staining, 20x magnification): endothelial cells infiltrate the Matrigel plug (1), forming new vascular structures, which are observed in both longitudinal (2) and cross sectional (3) perspectives.

### 3.1. Effect of indole-3-acetate on plug weight

Plug weight is an indirect measure of angiogenesis, subsequent to the increase on the number of vascular structures and intraluminal blood in the plug (Bhat, et al., 2013). In this study, plugs with VEGF<sub>A</sub> alone and VEGF<sub>A</sub> + indole-3-acetate weighted  $0.185 \pm 0.042$  g and  $0.132 \pm 0.037$  g, respectively. Indole-3-acetate plugs were on average 26.5% lighter than their respective controls ( $p=0.022$ ) (Figure 51).

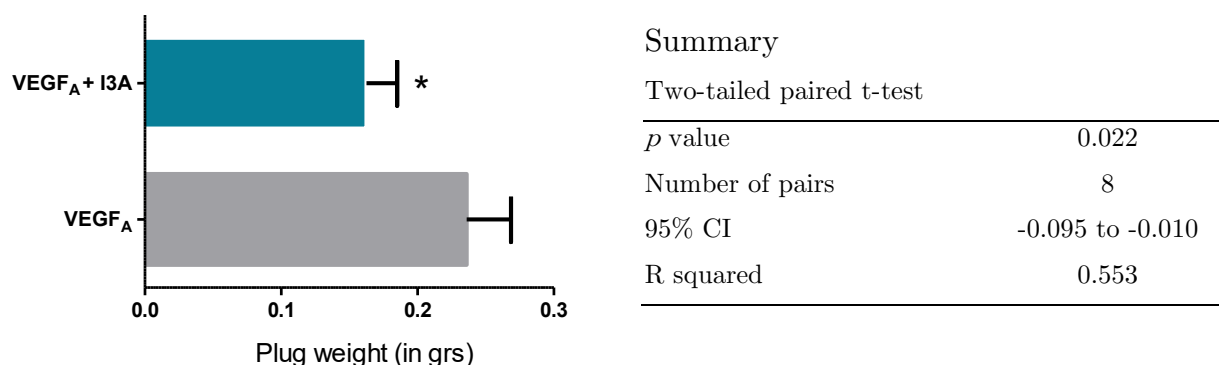


Figure 51. Plug weight is lower in plugs treated with indole-3-acetate. Plug weight values are expressed as mean  $\pm$  s.e.m. of three independent experiments ( $n = 3$ ). A paired two-tail t-test was performed for comparisons between groups. \* $p < 0.05$ .

### 3.2. Effect of indole-3-acetate on the vascular area

Macroscopically, the control plugs (containing VEGF<sub>A</sub> only) displayed a darker red colour, denoting increased blood content (Koo, et al., 2015)), when compared with plugs containing VEGF<sub>A</sub> and indole-3-acetate and implanted in the same mouse (Figure 52a). Regardless of inter-individual variability, for each mouse, the abundance of red colour was generally higher in the control plug (Figure 52a). Representative microscopy imaging after H+E staining of both conditions also suggests a higher development of the vascular network in the control plugs (VEGF<sub>A</sub> only) (Figure 52b-e).

In order to quantify the vascular area, immunohistochemistry for the Cluster of Differentiation 31 (CD31) was performed. CD31 (or platelet endothelial cell adhesion molecule-1 (PECAM-1)) is abundantly present in the surface of endothelial cells, and has been extensively used to quantify angiogenesis (Muruganandham, et al., 2006; Ragel, et al., 2007). Stitched images provide a global overview of the CD31<sup>+</sup> area in the plugs (Figure 53a-b); within the control plug, it is possible to observe several convoluted CD31<sup>+</sup> structures (Figure 53a); conversely, in the plug treated with indole-3-acetate, these structures are less evident (Figure 53b).

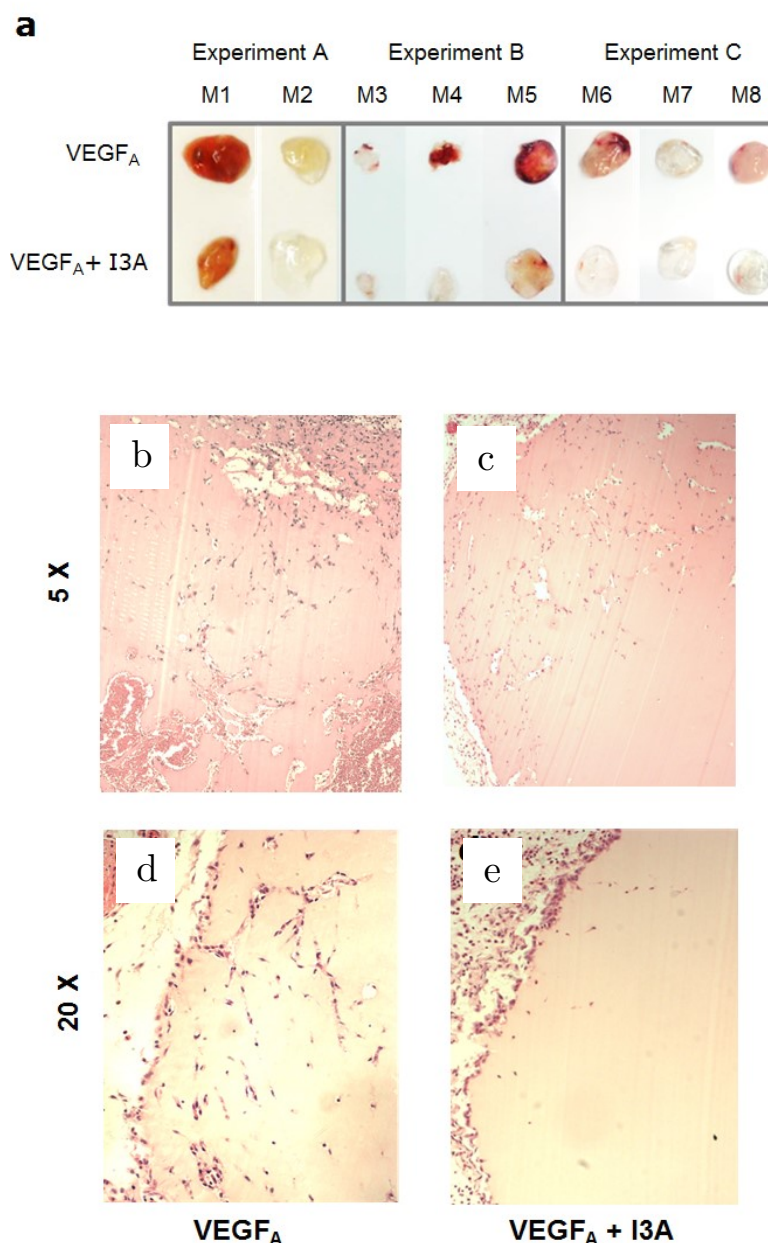


Figure 52. Macroscopic and microscopic overview of the Matrigel plugs.

a) Macroscopic aspect of all the plugs included in the analysis. Individual mice are described as M1 to M8. b-d) Representative pictures for one mouse are shown (H+E staining). b) The control exhibits a well-developed network, with newly-formed vessels converging into vascular pools (b, right and left inferior corners). c) A poorly-developed vascular network is observed in the plug containing indole-3-acetate, and no blood pools are detected. At a higher magnification (20x), the peripheral region of the control plug exhibits a higher density of migrating cells (d), when compared to the plug containing VEGF<sub>A</sub> + I3A (e). I3A: indole-3-acetate; VEGF<sub>A</sub>: Vascular Endothelial Growth Factor<sub>A</sub>.

The CD31<sup>+</sup> area was measured using the Image J software and expressed as the % of the total area of microscopic field (see representative images in Figure 53c). In

average, the CD31<sup>+</sup> area in the plug with indole-3-acetate was 69.1% lower than in its paired control ( $p=0.005$ ) (Figure 53d).

The Matrigel subcutaneous plug assay has a high inter-individual variability (Fridman, et al., 2012). Consistently, in this study, a two-way analysis of variance revealed that the individual effect accounted for 18.8% of the total variation ( $p=0.045$ ), whilst indole-3-acetate accounted for 31.8% of the total variance ( $p<0.0001$ ). The interaction between the indole-3-acetate and inter-individual effect was not significant ( $p=0.1521$ ).

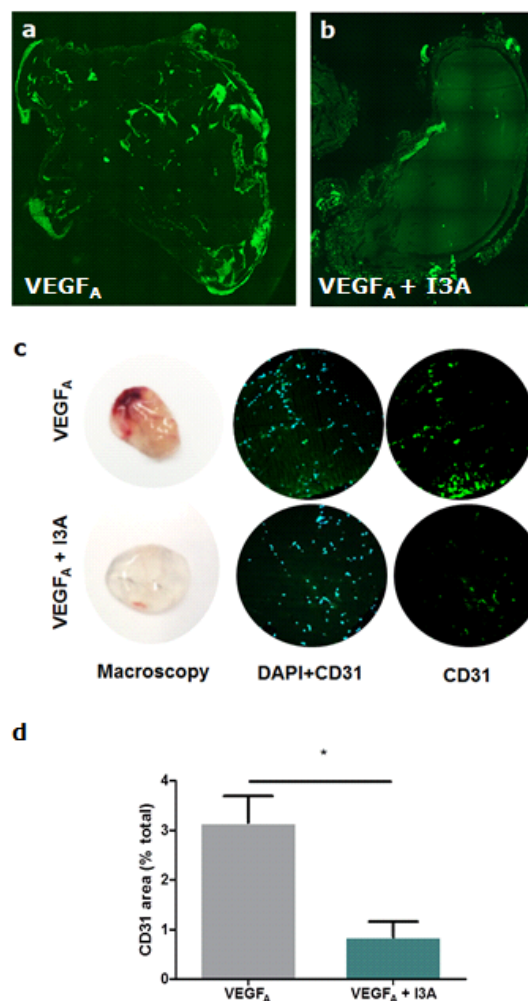


Figure 53. Indole-3-acetate inhibits angiogenesis in a subcutaneous Matrigel plug model.

a-b) Matrigel plugs were harvested after 7-day post- implantation, sectioned and stained with CD31 for immunohistochemistry. c) Sections were co-stained with DAPI (blue) for nucleus. Angiogenesis was quantified by measuring the CD31<sup>+</sup> -stained area for immunofluorescence with Image J software, in three random fields per plug. The average of the three measurements was calculated. d) Results are expressed as means  $\pm$  s.e.m. of three experiments). Two-tail paired t-test was used for comparisons between groups. \* $p<0.05$ ; I3A: indole-3-acetate; VEGF<sub>A</sub>: Vascular Endothelial Growth Factor A.

## 4. Discussion

It was previously demonstrated that indole-3-acetate is able to inhibit key steps of angiogenesis *in vitro*, using the HUVEC cell line (see Chapter 7). However, it is important to note that angiogenesis *in vivo* is a complex and highly regulated process that involves primarily endothelial cells but is further controlled by the cellular microenvironment and other cell types (*e.g.* macrophages, adipocytes) (Rahat, et al., 2014). In this chapter, the effect of indole-3-acetate on angiogenesis was confirmed, by employing a subcutaneous Matrigel plug assay. Specifically, we investigated the effect of the compound on the plug weight and the extension of the developed vascular area.

### 4.1. Indole-3-acetate reduces plug weight

In this study, we showed that indole-3-acetate significantly reduces the plug weight (-26.5%,  $p < 0.022$ ), when compared to control. These findings are in line with its predicted anti-angiogenic role based on VEGFR-2 inhibition and *in vitro* validation assays. Potent anti-angiogenic compounds (*e.g.* acacetin) are known to reduce the plug weight by up to 80%, when compared to the control (VEGF<sub>A</sub> only) (Bhat, et al., 2013). The decrease on plug weight was suggested to be a direct consequence of the inhibition of angiogenesis, with a resulting decrease on the number of vascular endothelial cells, vascular structures and intraluminal blood in the plug (Bhat, et al., 2013). However, these results must be interpreted with caution, as plug weight is highly affected by the size of the vessels and by stagnant pools of blood (AlMalki, et al., 2014). The quantification of angiogenesis by immunohistochemistry techniques has proven to produce more reliable and reproducible quantitative data (Schoell, et al., 1997; Simpson, et al., 1996).

### 4.2. Indole-3-acetate reduces the area of the vascular network

In order to investigate the impact of indole-3-acetate on the extension of the vascular network, CD31 immunohistochemistry was performed (Vermeulen, et al.,

1996). CD31, also named PECAM-1, is a member of the immunoglobulin superfamily and plays an important role in angiogenesis, leukocyte migration and integrin activation (Newman, et al., 1990). Various immunohistochemistry methods using anti-CD31 antibodies have been developed to quantify angiogenesis (Vermeulen et al. 1996). Moreover, the measurement of the CD31<sup>+</sup> area has been recognised as the most accurate index of vascularization (Schoell, et al., 1997; Simpson, et al., 1996) and was therefore the strategy adopted for quantification purposes in this experiment.

In this study, Matrigel plugs supplemented with indole-3-acetate resulted in a significant decrease in the vascularised area. The percentage of CD31<sup>+</sup> area was on average 69.1% lower in the plug treated with indole-3-acetate, when compared with its paired control ( $p=0.005$ ). Consistently with previous studies (Fridman, et al., 2012), a significant inter-individual variability was observed, accounting for 18.8% of the total variation ( $p=0.045$ ). The effect of individual variation on angiogenic response was minimised by implanting both a treated plug and a control in each mouse. This strategy allowed the use of paired statistical analyses, thus increasing statistical robustness. Regardless of inter-individual variability, all plugs treated with indole-3-acetate exhibited a smaller CD31<sup>+</sup> area, when compared with their respective controls. Altogether, these results suggest that indole-3-acetate is able to downregulate angiogenesis *in vivo*. These results are in line with the previous results *in vitro*, in which indole-3-acetate (at 100  $\mu$ M) was able to inhibit both cell migration and tube formation (see Chapter 7). This dual effect might account, at least partially, for lower *in vivo* vascularisation observed in the presence of indole-3-acetate.

## 5. Conclusion

Previous pharmacological inhibition assays and molecular docking analyses (Chapter 6) suggested that the indole-3-acetate might inhibit angiogenesis, through a direct competitive inhibition of VEGFR-2. This hypothesis was validated *in vitro* in Chapter 7, and finally *in vivo* in the present chapter.

Anti-angiogenic compounds have been proposed as a therapeutic approach in a range of pathologies characterised by increased proliferation of blood vessels (Taraboletti and Margosio, 2001; Vuorio, et al., 2012). This growing list of pathologies includes, amongst others, several cardiometabolic disorders (*e.g.* atherosclerosis, hypertension, obesity, diabetic retinopathy) and cancer (Tahergorabi and Khazaei, 2012). In order to validate a therapeutic role for indole-3-acetate in atherosclerosis, further work could include dosing *ApoE*<sup>-/-</sup> mice (a well-established mouse model for the study of atherosclerosis (Meir and Leitersdorf, 2004)) with indole-3-acetate, and assess its impact on the extension of the plaque, compared to non-treated controls. Xenograft studies could also clarify the role of indole-3-acetate in the development and progression of cancer, where sustained angiogenesis is a hallmark that allows the critical delivery of oxygen and energy to the neoplasia (Hanahan and Weinberg, 2011).

By identifying, for the first time, a role for indole-3-acetate as a gut microbial metabolite with anti-angiogenic properties, we identify a novel mechanism through which the gut microbiota can closely modulate a range of diseases with public health impact. Furthermore, this metabolite emerges both as a potential biomarker and therapeutic tool for such conditions, as will be further explored in the final chapter (*Chapter 9. General Discussion*).







## Chapter 9

# General Discussion

Over the last decades, various studies have highlighted the critical role of the gut microbiota in the onset and development of obesity and CVD (Delzenne and Cani, 2011; Jonsson and Backhed, 2017; Xu and Gordon, 2003).

Although protein kinases are one of the pharmaceutical industry's most important classes of drug targets - with almost 20 protein kinase inhibitors currently approved for clinical use and hundreds undergoing clinical trials (Cohen, 2002; Manning, et al., 2002) -, the human kinome remains an unexplored repertoire of microbial transduction triggers for human homeostasis and metabolism. Exploring the microbiome-kinome crosstalk opens, therefore, a novel research avenue to understand the impact of gut microbial metabolites in human health.

Amongst gut microbial metabolites, indoles were suggested to play a role as signalling molecules in obesity, as obese patients show higher modules for the degradation of tryptophan (Le Chatelier, et al., 2013). Most pro-inflammatory and pro-oxidative effects of indoles in vascular endothelial cells are thought to be mediated by AHR (Ito, et al., 2016; Jin, et al., 2014); however, as the expression of this receptor

is lost during adipocyte differentiation, this receptor is not likely to mediate its effects on energy metabolism in adipocytes (Shimba, et al., 2003).

The overall hypothesis of this work was that gut microbial metabolites - and in particular indoles - bind to human kinases, impact signalling pathways and interfere with cellular processes relevant to obesity and cardiometabolic disorders, thus acting as chemical messengers in the microbiome-kinome crosstalk.

This work describes two examples of indoles acting as chemical messengers in the microbiome-kinome crosstalk: while indoxyl sulfate shows a potential kinome-wide effect and pro-adipogenic properties, indole-3-acetate selectively inhibits VEGFR-2 and consequently angiogenesis. Both examples demonstrate how gut microbial metabolites can impact mechanisms relevant for the onset and development of obesity and cardiometabolic disorders. Although belonging to the same chemical family, these microbial metabolites demonstrate different patterns of impact on the human kinome. Hence, the impact of indoles in the microbial-mammalian crosstalk should be investigated at individual level, rather than assuming that their potential effects are class-specific.

In this Chapter, a summary of the key findings will be presented, followed by discussion of the strengths of the experiments, opportunities for future work and, finally, broader implications for human health.

## 1. Key findings

### 1.1. Gut microbial degradation of tryptophan converges to the production of indoxyl sulfate and indole-3-acetate

Using a systematic computational approach, it was demonstrated that the production of indoles is highly dependent on gut microbial metabolism, with 78% of the metabolic reactions being predictively performed by gut prokaryotes (Chapter 3). The metabolic reactions involved in the degradation of tryptophan into indoles converge to two main metabolites: indoxyl sulfate (*via* indole) and indole-3-acetate.

Regarding the production of indole (which is later sulfated in the liver in indoxyl sulfate), most predicted species belonged to the phyla Proteobacteria and Bacteroidetes. This systematic computational strategy predicted 23 species previously not associated with indole metabolism; two of these species were recently confirmed as indole-producing species, encoding tryptophanase in their genomes and/or producing detectable levels of indole (Alves, et al., 2006; Sasaki-Imamura, et al., 2011).

A different pattern was observed for indole-3-acetate, for which bacterial production was mostly associated with Proteobacteria and Firmicutes. *Pseudomonas putida*, *Pseudomonas fluorescens*, *Klebsiella oxytoca* and *Klebsiella pneumoniae* are amongst the species predicted to carry out several metabolic steps involved in the production of indole-3-acetate; interestingly, these species were previously demonstrated to be involved in the production of indole-3-acetate in plant studies (Karnwal, 2009; Patten and Glick, 2002). This data mining strategy also allowed the prediction of species previously not associated with indole-3-acetate production (*i.e.* *Pseudomonas nitroreducens* and *Raoultella terrigena*). Altogether, these results highlight the contribution of the gut microbiota to the production of indoles and demonstrate that each indolic compound has its own specific production path, to which a specific group of prokaryotes is likely to contribute.

## 1.2. Indoxyl sulfate exhibits pro-adipogenic properties through a putative multiple-kinase effect

Indoxyl sulfate, the only indolic compound unambiguously assigned by  $^1\text{H}$  NMR in the MetaHit study, is positively correlated with both adiposity (%Fat) and plasma levels of leptin (Chapter 4). No significant associations were found between indoxyl sulfate, inflammatory or metabolic parameters, or dietary tryptophan, thus excluding the effect of potential confounders. The pro-adipogenic effect of indoxyl sulfate was investigated *in vitro* (Chapter 5), where preliminary results suggest that it increases accumulation of intracellular lipids and production of leptin.

A high-throughput kinome screening demonstrated that indoxyl sulfate exhibits kinome-wide effects, binding to 46.3% of the kinases tested in a single-dose screening. Highest affinity bindings were found for PDGFRB, STK32C, c-KIT, MRCKA, MYLK3, MRCKB, GLK and CLK3, with PDGFRB emerging as the strongest candidate, with the lowest  $K_D$  and  $IC_{50}$  values. As PDGFRB was previously identified as a negative regulator of adipogenesis (Artemenko, et al., 2005; Fitter, et al., 2012), it was hypothesised that the pro-adipogenic effect of indoxyl sulfate could be mediated by inhibition of this receptor. However, this metabolite was not able to counteract the anti-adipogenic effect of PDGFRB activation, thus leading to the rejection of the initial hypothesis.

Given the kinome-wide impact of indoxyl sulfate, it is more plausible that its impact on adipocyte metabolism is mediated by a complex integration of effects of multiple kinases. Using a network-based analysis, the relevance of each kinase to the signalling pathways involved in adipogenesis was ranked. Kinases involved in the shortest betweenness-ranked paths leading to the production of triacylglycerol included PDGFRB, c-KIT, MRCKA, MYLK3, MRCKB and ZAK. The signalling of PDGFRB, ZAK, c-KIT and MRCKA converge to PLCB2, which therefore emerged as a potentially critical node for the pro-adipogenic effect of indoxyl sulfate. Altogether, these results identify indoxyl sulfate as a gut microbial metabolite with pro-adipogenic properties and provide seminal evidence on the putative targets and mechanisms mediating its effect.

### 1.3. Indole-3-acetate exhibits anti-angiogenic properties through a putative VEGFR-2 inhibition

Contrary to indoxyl sulfate, indole-3-acetate appears to have a very specific effect, with a single positive hit (VEGFR-2) in the kinome screening (Chapter 6). The single-dose kinome scan was complemented by the confirmation of the physical interaction ( $K_D = 890 \mu\text{M}$ ) and by the verification of inhibition of the phosphorylation activity ( $IC_{50} = 333.7 \mu\text{M}$ ). The physical interaction between indole-3-acetate and VEGFR-2 was further explored by molecular docking analysis, which suggested an ATP-competitive inhibitory mechanism. The inhibition of VEGFR-2 autophosphorylation by indole-3-acetate was subsequently confirmed in HUVEC; consistent with a putative role as VEGFR-2 inhibitor, indole-3-acetate significantly reduced the phosphorylation of AKT and the production of NO.

As VEGFR-2 is the main receptor involved in angiogenesis, it was hypothesised that indole-3-acetate would exhibit anti-angiogenic properties. Supporting this hypothesis, indole-3-acetate significantly inhibited both HUVEC cell migration and tube formation *in vitro* (Chapter 7). The anti-angiogenic effect of indole-3-acetate was further demonstrated *in vivo* using a Matrigel subcutaneous plug assay, where the compound inhibited the area of the plug's vascular network, expressed as the per cent of the CD31<sup>+</sup> area (Chapter 8).

Altogether, this combined strategy comprising a kinome screening and molecular docking analyses, combined with a panel of cell-based assays and, finally, *in vivo* validations provides a comprehensive body of evidence supporting an anti-angiogenic role for indole-3-acetate.

## 2. Experimental strengths and future work

The data-mining strategy proposed in Chapter 3 is the first attempt to use a systematic computational approach to clarify the gut microbial metabolism of indoles. This strategy provides a systematic overview of the metabolic pathways and predicted species involved in the metabolism of these compounds, identifying a list of novel species. However, as mentioned before, these predictions require experimental validation, such as the quantification of indole-3-acetate in pure cultures of bacteria grown in tryptophan-containing media (*e.g.* by LC-MS/MS). As a broader application, this methodology can be a useful tool to interrogate the metabolism of other gut microbial products.

In Chapter 4, a positive association between indoxyl sulfate and whole body fat (%Fat) was identified, in the context of non-genetic obesity.  $^1\text{H}$  NMR was used due to its high reproducibility and wide use in metabolic profiling in the context of obesity and cardiometabolic disorders, thus allowing comparisons with other studies (Calvani, et al., 2010; Dumas, et al., 2006; Elliott, et al., 2015). The assignment of indoxyl sulfate was confirmed by a spike-in experiment, as recommended when identifying a novel or putatively relevant biomarker (Dona, et al., 2016). The relationship between indoxyl sulfate and dietary intake was further explored, and potential metabolic and dietary confounders were excluded. In order to strengthen these findings, the statistical relationship between indoxyl sulfate and obesity should be further validated in other human cohorts. Of note, indoxyl sulfate was the only indolic compound unambiguously assigned by  $^1\text{H}$  NMR in this study. Indole-3-acetate was putatively assigned as a low-intensity signal in a standard ID experiment with water pre-saturation, and confirmed by 2D experiments (COSY and HSQC); however, spectral overlap with other compounds prevented an adequate quantification of this metabolite. Future work should address the unambiguous assignment and quantification of other indoles, using targeted mass spectrometry methods (Markley, et al., 2017), in order to explore their association with obesity and/or other cardiometabolic disorders.

One of the strengths of this work relies on capitalising on the clues provided by  $^1\text{H}$  NMR to investigate the role of indoles as pharmacological players, thus moving from correlation to causation. For the first time, it was demonstrated that indoxyl sulfate increases lipogenesis and leptin production in 3T3-L1, one of the most used cell lines to investigate the adipogenic potential of both endocrine and pharmacological compounds (Green and Kehinde, 1975; Green and Meuth, 1974) (Chapter 5). These findings can be strengthened by increasing the number of experiments and by validating the pro-adipogenic effect using other molecular biology techniques, such as quantification of the expression of adipocyte marker genes by qPCR. In particular, PPAR $\gamma$ , a nuclear receptor from the ligand-activated transcription factors family and vital for adipocyte differentiation (Gregoire, et al., 1998), and other markers of adipocyte differentiation (*i.e.* *C/EBP $\alpha$* , *FASN* and *ADIPOQ* (Gregoire, et al., 1998; Sul, et al., 2000)) should be further investigated.

By identifying key examples of how indoles act as chemical messengers in the microbiome-kinome crosstalk (Chapters 5 and 6), this work opens a new research avenue to understand the impact of gut microbial metabolites, and in particular indoles, in obesity and cardiometabolic disorders. Although being one of the pharmaceutical industry's most important classes of drug targets (Cohen, 2002; Manning, et al., 2002), the human kinome remained an unexplored repertoire of microbial transduction triggers for human homeostasis and metabolism. Importantly, this work highlights the within-class heterogeneity of indoles, with individual compounds exhibiting different targets and different kinome specificity. In the first example, indoxyl sulfate shows kinome-wide effects, representing an increased challenge regarding the clarification of targets and signalling pathways. Although a preliminary map of the strongest candidate targets is provided in this thesis, further work using phosphoproteomics coupled with network integration is needed to provide system-level information, to characterise the pathways involved and to identify potential inter-pathway cooperation.

Conversely, indole-3-acetate was identified as a specific inhibitor of VEGFR-2. VEGFR-2 is the main regulator of angiogenesis. Thus, the anti-angiogenic effect of indole-3-acetate was validated *in vitro*. As recommended, the effect of the compound

was evaluated in more than one mechanistic step (Auerbach, et al., 2003), using a panel of well-established cell-based assays and mimicking as close as possible physiological conditions (*e.g.* concentration, surface coating with Collagen-I and Matrigel), and finally validated *in vitro* using a Matrigel subcutaneous plug. Overall, one of the major strengths of this work is the consistency between *in silico*, *in vitro* and *in vivo* results that, combined in a comprehensive research strategy, support an anti-angiogenic effect of indole-3-acetate.

However, some experimental limitations should be noted. Although indole-3-acetate significantly impaired the phosphorylation of Tyr<sup>1175</sup>, one of the major autophosphorylation sites of VEGFR-2 in humans *in vitro* (Takahashi, et al., 2001), the magnitude of VEGFR-2 activation was not consistent between experiments performed under the same experimental conditions. As future work, these results could be potentially strengthened by testing higher concentrations of the compounds and the agonist, and by trying alternative normalisation techniques (*i.e.* normalisation to total protein content). Using molecular docking analysis, indole-3-acetate was predicted to inhibit VEGFR-2 through an ATP-competitive effect. Also, although indole-3-acetate was also shown to inhibit VEGFR-2 downstream signalling (AKT phosphorylation and NO production) (Chapter 6), it is important to note that one-tail testing was performed (*i.e.* testing for inhibition, as predicted by the *IC*<sub>50</sub>). These results can be strengthened in the future by increasing the number of experiments, optimising the concentrations used and investigating of the optimal time points to assess both outcomes. Additionally, in order to have a more comprehensive overview of the impact of indole-3-acetate on (Tyr<sup>1145</sup>)VEGFR-2 downstream signalling, it would be important to clarify its effect in other pathways (*i.e.* PKC/MEK/ERK, Src and FAK/paxillin). Finally, although *in silico* molecular docking is one of the most frequently used methods to predict the binding-conformation of small molecule ligands to a given target (Kitchen, et al., 2004), it requiring further *in vitro* validations (*i.e.* mutagenesis of the amino acids predicted to be involved in the binding).

Finally, it is important to consider that the effect of indole-3-acetate might be a multiparametric pharmacological response with a primary target (VEGFR-2) and multiple secondary targets. AHR, a nuclear receptor suggested to have a negative

impact in angiogenesis (Eckers, et al., 2016), was previously proposed as a target of indole-3-acetate; however, only a very weak AHR agonism was observed *in vitro* using high, non-physiological doses (250  $\mu$ M). These findings do not support that the effect of indole-3-acetate, at physiological concentrations, is mediated by AHR, however, small interfering RNA (siRNA) / short hairpin RNA (shRNA) experiments targeting other putative receptor involved should provide further insights in order to include, exclude or assess the magnitude of the contribution of the targets involved. In this case, a direct VEGFR-2 inhibition arises as a simpler, more logical and testable putative explanation, thus emerging as preferable according to the law of parsimony (*lex parsimoniae*, a problem-solving principle used as an heuristic guide in the development of theoretical models in science) (Gauch, 2003).

### **3. Consequences of kinome inhibition for obesity and cardiometabolic diseases**

#### **3.1. A novel contribution of the gut microbiota to human obesity**

Several microbial metabolites have been previously shown to play a role in obesity. As an example, SCFA (*i.e.* butyrate, propionate and acetate) exert a beneficial effect in diet-induced obesity (Chilloux, et al., 2016). While acetate alters the hypothalamic expression of neuropeptides involved in appetite suppression (Frost, et al., 2014), propionate upregulates the release of glucagon-like peptide-1 and peptide YY (Psichas, et al., 2015).

In this work, indoxyl sulfate is proposed as a novel gut microbial metabolite involved in human obesity. By demonstrating that the levels of indoxyl sulfate are positively correlated with adiposity, these results are in line with the findings of Zhang et al, who reported higher levels of this metabolite in a context of paediatric genetic obesity (Zhang, et al., 2015). Altogether, these results support that the association between indoxyl sulfate and obesity is not specific of a genetic context. The underlying conceptual model proposes that, in obese patients, gut dysbiosis and increased genomic

ability to degrade tryptophan (Le Chatelier, et al., 2013) result in an increased production of indoxyl sulfate, which exhibits pro-adipogenic properties and thus contributes to increased adiposity.

An exploratory target screening provided new knowledge on the impact of indoles on the human kinome, reinforcing the relevance of this repertoire of proteins as a critical object of study to better understand the microbial-mammalian crosstalk. By shortlisting putative human targets of indoxyl sulfate, this work provides a first attempt to clarify the targets and mechanisms by which indoxyl sulfate affects the onset and progression of obesity. *PLCB2* emerged as a potentially critical node for the pro-adipogenic effect of indoxyl sulfate; it codes for the enzyme 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase  $\beta 2$ , which hydrolyses PIP<sub>2</sub> to IP<sub>3</sub> and DAG. DAG is further esterified into TAG by the diglyceride acyltransferase (Saponaro, et al., 2015). Both IP<sub>3</sub> and DAG are involved in processes relevant to obesity. DAG stimulates PKC, which in turn phosphorylates myristoylated alanine-rich protein kinase C substrate (MARCKS); MARCKS then binds actin and calcium-calmodulin, finally inducing insulin exocytosis (Lustig, 2003). Insulin secretion is also promoted *via* IP<sub>3</sub>, which induces calcium release from intracellular stores (Lustig, 2003).

### **3.2. Impact of inhibition of angiogenesis by gut microbial metabolites in human health and disease**

The identification of indole-3-acetate as a gut microbial metabolite with anti-angiogenic properties has wide-ranging implications for human health. Angiogenesis is a tightly controlled process and, when dysregulated, contributes to a variety of human diseases. Insufficient angiogenesis has been associated with other cardiometabolic outcomes, such as hypertension (where impaired vasodilation leads to microvessel rarefaction) (Battegay, et al., 2007) or diabetes (characterised by impaired angiogenesis in ischemic limbs, but increased neovascularisation in the retina) (Tamarat, et al., 2004). Angiogenesis also seems to play a relevant role in atherosclerosis. Adventitial angiogenesis allows atherosclerotic plaques to develop beyond a critical thickness point (Barger and Beeuwkes, 1990). It was demonstrated that angiogenesis inhibitors (*i.e.* endostatin and TNP-470) are able to reduce the size of atherosclerotic plaques by 85%

and 70%, respectively (Moulton, et al., 1999). Furthermore, the neo-vascular network renders plaques more fragile and prone to rupture and is, therefore, a cause of luminal occlusion and acute coronary syndromes (Barger and Beeuwkes, 1990; Moreno, et al., 2004). As a gut microbial metabolite with anti-angiogenic properties, indole-3-acetate could exhibit a similar protective effect in patients with atherosclerosis.

The findings from this work provide new knowledge on the role of the gut microbiota in angiogenesis (a key process in the onset and development of atherosclerosis) and naturally complement the conceptual model proposed by Jonsson and Backhed that suggests the gut microbiota can impact atherosclerosis *via* infections/inflammatory responses that aggravate plaque development or rupture, or upon the action of gut microbial metabolites (Jonsson and Backhed, 2017).

Indole-3-acetate naturally emerges as a particularly interesting therapeutic tool in the context of angiogenesis-related health outcomes, as will be further explored in *Section 3.4. The gut microbial pharmacopeia as a potential therapeutic tool*.

### **3.3. The gut microbial pharmacopeia as a potential therapeutic tool**

Indoles are a heterogeneous class: each one of the studied metabolites exhibits a particular metabolic production pathway, different human targets and specificity pattern – and, therefore, different potential as therapeutic tools.

Indole-3-acetate is as a particularly interesting therapeutic tool in the context of angiogenesis-related health outcomes, particularly in atherosclerosis and cancer. As angiogenesis inhibitors are able to reduce the size of the atherosclerotic plaque (Moulton, et al., 1999), indole-3-acetate may have a similar beneficial effect, which remains to be validated *in vivo* (*e.g.* using an ApoE<sup>-/-</sup> mouse model). Sustained angiogenesis is also a hallmark of cancer (Hanahan and Weinberg, 2011), several anti-angiogenic drugs are currently used for its treatment (Oehler and Bicknell, 2000; Taraboletti and Margosio, 2001; Vasudev and Reynolds, 2014). VEGF<sub>A</sub> is overexpressed in most solid cancers, and its inhibition can suppress tumour growth in animal models (Oehler and Bicknell, 2000). Based on these findings, numerous anti-

cancer therapies targeting angiogenesis have been developed, many of which are approved by the Food and Drugs Administration (Carmeliet and Jain, 2011; Potente, et al., 2011). Contrary to most VEGFR-2 inhibitors that inhibit multiple kinases and, therefore, have a wide range of side effects, indole-3-acetate has shown a very specific inhibition for VEGFR-2 in our kinome screening – and, therefore, a lower potential for side effects. Therefore, indole-3-acetate could represent both a protective factor and a therapeutic tool in the context of cancer, especially in colon (where its absorption occurs and the area of contact is maximal) and in the liver and portal circulatory system (where its concentrations are particularly high).

Finally, putative microbial interventions aiming to modulate indoles should integrate the impact of indoles-producing species as part of a broader body of knowledge, as these species might also contribute to the production of other metabolites with either beneficial or deleterious effects for health outcomes. Due to the complexity of the gut microbial ecosystem, significant impact is unlikely to be achieved unless several species are simultaneously targeted.

### **3.4. Gut microbial metabolites in risk stratification and personalised medicine**

Currently, the clinical management of obesity and cardiometabolic disorders consists of a comprehensive, non-personalised lifestyle management, associated or not to pharmacological therapies. However, there is an obvious inter-individual heterogeneity. Rather than ignoring the existence of multiple phenotypes, the personalised medicine movement has embraced the belief that individual-level risk stratification can be used to inform treatment, but also to empower patients to adopt healthier choices and promote meaningful behaviour changes.

The potential of metabonomics in personalised health care has been extensively explored (Nicholson, et al., 2011); as an example, basal metabolite levels were able to predict future weight gain in patients undergoing chemotherapy for early stage breast cancer (Keun, et al., 2009). Similar to indoxyl sulfate, other gut microbial metabolites

(*e.g.* SCFA, bile acids, TMAO) have been suggested to play a role on the onset of obesity and other cardiometabolic disorders (Chilloux, et al., 2016; Dumas, et al., 2017). Altogether, the contributions of individual metabolites could be incorporated in the development of stratification tools in the context of obesity and cardiometabolic disorders, moving into a more personalised delivery of care. Similarly, if the protective effect of indole-3-acetate on atherosclerosis and cancer is confirmed, the endogenous levels of indole-3-acetate can be investigated as putative risk markers for the occurrence of acute coronary syndromes or cancer, respectively, in high-risk patients.

## 4. Conclusion

Although gut microbial metabolites have a well-recognised role in obesity and cardiometabolic disorders (Delzenne and Cani, 2011; Jonsson and Backhed, 2017; Xu and Gordon, 2003), their targets and mechanisms are not fully understood. In this context, the human kinome - one of the pharmaceutical industry's most important classes of drug targets - has remained mostly unexplored as a repertoire of transduction triggers for human metabolism.

This work highlights the role of indoles as chemical messengers in the microbiome-kinome crosstalk: whereas indoxyl sulfate shows a putative kinome-wide effect and pro-adipogenic properties, indole-3-acetate selectively inhibits VEGFR-2 and subsequently angiogenesis. These results demonstrate how, by binding to human kinases, indoles can critically impact pathways and cellular processes relevant for the onset and development of obesity and cardiometabolic diseases. Angiogenesis, in particular, has been proposed as a key mechanism in several cardiometabolic disorders (*e.g.* obesity, arterial hypertension, diabetic retinopathy) – and, importantly, in cancer.

By unravelling the targets, mechanisms and consequences of indoles as gut microbial pharmaco-metabolites, this work not only contributes to a better understanding of the link between the gut microbiota, obesity and cardiometabolic disorders – but also opens novel avenues for the use of the gut pharmacopeia in risk stratification and personalised health care delivery.







# List of published papers

Papers A and C are related to this work. Papers D to G were published during my time as a PhD candidate, however they are not directly related to work presented in this thesis. Copies of the first page of all papers can be found in Appendix A and respective reproduction permissions in Appendix B. Below, a\* symbol indicates equal contributions of two authors.

(A) Neves AL, Chilloux J, Sarafian MH, Rahim MB, Boulange CL, Dumas ME. The microbiome and its pharmacological targets: therapeutic avenues in cardiometabolic diseases. *Curr Opin Pharmacol*. 2015; 25:36-44. [review]

(B) Boulange CL, Neves AL, Chilloux J, Nicholson JK, Dumas ME. Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Genome Med*. 2016; 8(1):42.

(C) Chilloux J\*, Neves AL\*, Boulange CL, Dumas ME. The microbial-mammalian metabolic axis: a critical symbiotic relationship. *Curr Opin Clin Nutr Metab Care*. 2016; 19(4):250-256. [review]

(D) Neves AL, Couto L. Cardiovascular risk in overweight/obese and lean hypertensive patients. *Rev Port Cardiol*. 2014; 33(4):223-8.

(E) Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, Mendes GA, Lau AY. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. *J Am Med Inform Assoc*. 2015; 22(1):243-56.

(F) Pinho-Costa L, Yakubu K, Hoedebecke K, Laranjo L, Reichel CP, Colon-Gonzalez MD, Neves AL, Errami H. Healthcare hashtag index development: Identifying global impact in social media. *J Biomed Inform*. 2016; 63:390-399.

(G) Rodriguez-Martinez A, Ayala R, Posma JM, Neves AL, Gauguier D, Nicholson JK, Dumas ME. MetaboSignal: a network-based approach for topological analysis of metabotype regulation via metabolic and signaling pathways. *Bioinformatics* 2017; 33(5):773-775.



# References

- Adams, R.H. and Alitalo, K. (2007) Molecular regulation of angiogenesis and lymphangiogenesis, *Nat Rev Mol Cell Biol*, 8, 464-478.
- Adesso, S., *et al.* (2013) The uremic toxin indoxyl sulphate enhances macrophage response to LPS, *PLoS One*, 8, e76778.
- Aicher, A., Zeiher, A.M. and Dimmeler, S. (2005) Mobilizing endothelial progenitor cells, *Hypertension*, 45, 321-325.
- Ajay and Murcko, M.A. (1995) Computational methods to predict binding free energy in ligand-receptor complexes, *Journal Med Chem*, 38, 4953-4967.
- Al-Waiz, M., *et al.* (1987) Trimethylaminuria (fish-odour syndrome): an inborn error of oxidative metabolism, *Lancet*, 1, 634-635.
- Al-Waiz, M., *et al.* (1992) The exogenous origin of trimethylamine in the mouse, *Metabolism*, 41, 135-136.
- Alberti, K.G.M.M., *et al.* (2005) The metabolic syndrome - a new worldwide definition, *Lancet*, 366, 1059-1062.
- AlMalki, W.H., *et al.* (2014) Assessment methods for angiogenesis and current approaches for its quantification, *Ind J Pharmacol*, 46, 251-256.
- Altschul, S.F., *et al.* (1990) Basic local alignment search tool, *J Mol Biol*, 215, 403-410.
- Alves, M.S., *et al.* (2006) Identification of clinical isolates of indole-positive and indole-negative *Klebsiella* spp, *J Clin Microbiol*, 44, 3640-3646.
- Amandi, A., *et al.* (1982) Isolation and characterization of *Edwardsiella tarda* from fall chinook salmon (*Oncorhynchus tshawytscha*), *Appl Environ Microbiol*, 43, 1380-1384.
- Amar, J., *et al.* (2008) Energy intake is associated with endotoxemia in apparently healthy men, *Am J Clin Nutr*, 87, 1219-1223.
- Anandacoomarasamy, A., *et al.* (2008) The impact of obesity on the musculoskeletal system, *Int J Obesity (2005)*, 32, 211-222.
- Andersen, A.H., *et al.* (2012) Partial least squares for discrimination in fMRI data, *Magn Reson Imaging*, 30, 446-452.
- Arnaoutova, I., *et al.* (2009) The endothelial cell tube formation assay on basement membrane turns 20: state of the science and the art, *Angiogenesis*, 12, 267-274.
- Arnaoutova, I. and Kleinman, H.K. (2010) In vitro angiogenesis: endothelial cell tube formation on gelled basement membrane extract, *Nat Protoc*, 5, 628-635.
- Aronoff, S., *et al.* (2004) Glucose Metabolism and Regulation: Beyond Insulin and Glucagon, *Diabetes Spect*, 3, 183-190.
- Artemenko, Y., *et al.* (2005) Anti-adipogenic effect of PDGF is reversed by PKC inhibition, *J Cell Physiol*, 204, 646-653.
- Arumugam, M., *et al.* (2011) Enterotypes of the human gut microbiome, *Nature*, 473, 174-180.

- Atarashi, K., *et al.* (2013) Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota, *Nature*, 500, 232-236.
- Atarashi, K., *et al.* (2011) Induction of colonic regulatory T cells by indigenous Clostridium species, *Science*, 331, 337-341.
- Attwood, G., *et al.* (2006) Production of indolic compounds by rumen bacteria isolated from grazing ruminants, *J Appl Microbiol*, 100, 1261-1271.
- Auerbach, R., *et al.* (2003) Angiogenesis assays: a critical overview, *Clin Chem*, 49, 32-40.
- Backhed, F., *et al.* (2005) Host-bacterial mutualism in the human intestine, *Science*, 307, 1915-1920.
- Badenoch-Jones, J., *et al.* (1982) Mass spectrometric quantification of indole-3-acetic Acid in Rhizobium culture supernatants: relation to root hair curling and nodule initiation, *Appl Environ Microbiol*, 44, 275-280.
- Baker, J.H.E., *et al.* (2006) Vascular-specific quantification in an in vivo Matrigel chamber angiogenesis assay, *Microvasc Res*, 71, 69-75.
- Baker, J.L., Olsen, L.W. and Sorensen, T.I.A. (2007) Childhood body-mass index and the risk of coronary heart disease in adulthood, *NEJM*, 357, 2329-2337.
- Banasaz, M., *et al.* (2002) Increased enterocyte production in gnotobiotic rats mono-associated with Lactobacillus rhamnosus GG, *Appl Environ Microbiol*, 68, 3031-3034.
- Barger, A.C. and Beeuwkes, R., 3rd (1990) Rupture of coronary vasa vasorum as a trigger of acute myocardial infarction, *Am J Cardiol*, 66, 41G-43G.
- Barreto, F.C., *et al.* (2009) Serum indoxyl sulfate is associated with vascular disease and mortality in chronic kidney disease patients, *Clin J Am Soc Nephrol: CJASN*, 4, 1551-1558.
- Bashan, A., *et al.* (2012) Network physiology reveals relations between network topology and physiological function, *Nat Commun*, 3, 702.
- Battegay, E.J., *et al.* (2007) Effects of anti-hypertensive drugs on vessel rarefaction, *Curr Opin Pharmacol*, 7, 151-157.
- Bauer, H., *et al.* (1963) The response of the lymphatic tissue to the microbial flora. Studies on germfree mice, *Am J Pathol*, 42, 471-483.
- Benson, A.K. (2016) The gut microbiome - an emerging complex trait, *Nat Genet*, 48, 1301-1302.
- Benson, A.K., *et al.* (2010) Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors, *Proc Natl Acad Sci USA*, 107, 18933-18938.
- Benveniste, J., *et al.* (1971) Immunoglobulins in intact, immunized, and contaminated axenic mice: study of serum IgA, *J Immunol*, 107, 1647-1655.
- Berman, H.M., *et al.* (2000) The Protein Data Bank, *Nucleic Acids Res*, 28, 235-242.
- Bhat, T.A., *et al.* (2013) Acacetin inhibits in vitro and in vivo angiogenesis and downregulates Stat signaling and VEGF expression, *Cancer Prev Res*, 6, 1128-1139.
- Biagi, E., *et al.* (2010) Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians, *PLoS One*, 5, e10667.
- Biggs, N., Lloyd, E. and Wilson, R. (1986) *Graph Theory*. Oxford University Press.

- Bindels, L.B., Dewulf, E.M. and Delzenne, N.M. (2013) GPR43/FFA2: physiopathological relevance and therapeutic prospects, *Trends Pharmacol Sci*, 34, 226-232.
- Blaise, B.J., *et al.* (2009) Metabolic profiling strategy of *Caenorhabditis elegans* by whole-organism nuclear magnetic resonance, *J Prot Res*, 8, 2542-2550.
- Blasi, F. and Carmeliet, P. (2002) uPAR: a versatile signalling orchestrator, *Nat Rev Mol Cell Biol*, 3, 932-943.
- Blaum, C.S., *et al.* (2005) The association between obesity and the frailty syndrome in older women: the Women's Health and Aging Studies, *J Am Geriatr Soc*, 53, 927-934.
- Bleich, S., *et al.* (2008) Why is the developed world obese?, *Annu Rev Public Health*, 29, 273-295.
- Boerma, M., *et al.* (2006) Comparative expression profiling in primary and immortalized endothelial cells: changes in gene expression in response to hydroxy methylglutaryl-coenzyme A reductase inhibition, *Blood Coagul Fibrinolysis*, 17, 173-180.
- Bold, G., *et al.* (2016) A Novel Potent Oral Series of VEGFR2 Inhibitors Abrogate Tumor Growth by Inhibiting Angiogenesis, *J Med Chem*, 59, 132-146.
- Bollard, M.E., *et al.* (2005) NMR-based metabolomic approaches for evaluating physiological influences on biofluid composition, *NMR Biomed*, 18, 143-162.
- Bonacich, P. (1987) Power and Centrality: A Family of Measures, *Am J Sociol*, 92, 1170-1182.
- Bonder, M.J., *et al.* (2016) The effect of host genetics on the gut microbiome, *Nat Genet*, 48, 1407-1412.
- Borlak, J., *et al.* (2003) Verapamil: identification of novel metabolites in cultures of primary human hepatocytes and human urine by LC-MS(n) and LC-NMR, *Xenobiotica*, 33, 655-676.
- Borrel, G., *et al.* (2017) Genomics and metagenomics of trimethylamine-utilizing Archaea in the human gut microbiome, *ISMEJ*, 11, 2059-2074.
- Bouatra, S., *et al.* (2013) The human urine metabolome, *PLoS One*, 8, e73076.
- Brestoff, J.R. and Artis, D. (2013) Commensal bacteria at the interface of host metabolism and the immune system, *Nat Immunol*, 14, 676-684.
- Brill, A., *et al.* (2005) Platelet-derived microparticles induce angiogenesis and stimulate post-ischemic revascularization, *Cardiovasc Res*, 67, 30-38.
- Brito, J.S.d., *et al.* (2016) Is there a relationship between tryptophan dietary intake and plasma levels of indoxyl sulfate in chronic kidney disease patients on hemodialysis?, *J Bras Nefrol*, 38, 396-402.
- Browne, H.P., *et al.* (2016) Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation, *Nature*, 533, 543-546.
- Brun, P., *et al.* (2007) Increased intestinal permeability in obese mice: new evidence in the pathogenesis of nonalcoholic steatohepatitis, *Am J Physiol Gastrointest Liver Physiol*, 292, G518-525.
- Bruns, A.F., *et al.* (2010) Ligand-stimulated VEGFR2 signaling is regulated by co-ordinated trafficking and proteolysis, *Traffic*, 11, 161-174.
- Bylesjö, M., *et al.* (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification, *J Chemometr*, 20, 341-351.
- Bylesjö, M., *et al.* (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification, *J Chemom*, 20, 341-351.

- Calaf, R., *et al.* (2011) Determination of uremic solutes in biological fluids of chronic kidney disease patients by HPLC assay, *J Chromatogr B Analyt Technol Biomed Life Sci*, 879, 2281-2286.
- Caldwell, R.B., *et al.* (2005) Vascular endothelial growth factor and diabetic retinopathy: role of oxidative stress, *Curr Drug Targets*, 6, 511-524.
- Calvani, R., *et al.* (2010) Gut microbiome-derived metabolites characterize a peculiar obese urinary metabotype, *Int J Obesity*, 34, 1095-1098.
- Cani, P.D. (2014) Metabolism in 2013: The gut microbiota manages host metabolism, *Nat Rev Endocrinol*, 10, 74-76.
- Cani, P.D., *et al.* (2007) Metabolic endotoxemia initiates obesity and insulin resistance, *Diabetes*, 56, 1761-1772.
- Cani, P.D., *et al.* (2008) Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice, *Diabetes*, 57, 1470-1481.
- Cani, P.D., *et al.* (2007) Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced diabetes in mice through a mechanism associated with endotoxaemia, *Diabetologia*, 50, 2374-2383.
- Cani, P.D., *et al.* (2009) Changes in gut microbiota control inflammation in obese mice through a mechanism involving GLP-2-driven improvement of gut permeability, *Gut*, 58, 1091-1103.
- Cao, Y. (2007) Angiogenesis modulates adipogenesis and obesity, *J Clin Invest*, 117, 2362-2368.
- Cario, E., Gerken, G. and Podolsky, D.K. (2007) Toll-like receptor 2 controls mucosal inflammation by regulating epithelial barrier function, *Gastroenterology*, 132, 1359-1374.
- Carmeliet, P. (2005) Angiogenesis in life, disease and medicine, *Nature*, 438, 932-936.
- Carmeliet, P. and Jain, R.K. (2011) Molecular mechanisms and clinical applications of angiogenesis, *Nature*, 473, 298-307.
- Caro, J.F., *et al.* (1996) Leptin: the tale of an obesity gene, *Diabetes*, 45, 1455-1462.
- Carpentier, G. (2012) ImageJ contribution: Angiogenesis Analyzer., *ImageJ News*.
- Celletti, F.L., *et al.* (2001) Vascular endothelial growth factor enhances atherosclerotic plaque progression, *Nat Med*, 7, 425-429.
- Celloto, V.R., *et al.* (2012) Biosynthesis of indole-3-acetic acid by new *Klebsiella oxytoca* free and immobilized cells on inorganic matrices, *Scientific World Journal*, 2012, 495970.
- Chen, Y., *et al.* (2012) Excess body weight and the risk of primary liver cancer: an updated meta-analysis of prospective studies, *Eur J Cancer*, 48, 2137-2145.
- Cheng, Y., *et al.* (2015) Aryl Hydrocarbon Receptor Activity of Tryptophan Metabolites in Young Adult Mouse Colonocytes, *Drug Metab Dispos*, 43, 1536-1543.
- Cheung, B.M.Y., *et al.* (2006) Association of hypertension with single nucleotide polymorphisms in the quantitative trait locus for abdominal obesity-metabolic syndrome on chromosome 17, *J Hum Hypertens*, 20, 419-425.
- Cheung, S.Y., *et al.* (2015) Activation of transient receptor potential vanilloid 3 channel suppresses adipogenesis, *Endocrinology*, 156, 2074-2086.
- Chilloux, J., *et al.* (2016) The microbial-mammalian metabolic axis: a critical symbiotic relationship, *COCN*, 19, 250-256.

- Cho, I., *et al.* (2012) Antibiotics in early life alter the murine colonic microbiome and adiposity, *Nature*, 488, 621-626.
- Choy, J.S., *et al.* (2010) DNA methylation increases nucleosome compaction and rigidity, *J Am Chem Soc*, 132, 1782-1783.
- Claesson, M.J., *et al.* (2011) Composition, variability, and temporal stability of the intestinal microbiota of the elderly, *Proc Natl Acad Sci*, 108 Suppl 1, 4586-4591.
- Claesson, M.J., *et al.* (2012) Gut microbiota composition correlates with diet and health in the elderly, *Nature*, 488, 178-184.
- Cloarec, O., *et al.* (2005) Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets, *Anal Chem*, 77, 1282-1289.
- Cloarec, O., *et al.* (2005) Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies, *Anal Chem*, 77, 517-526.
- Cloarec, O., *et al.* (2005) Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies, *Anal Chem*, 77, 517-526.
- Coelho, M., Oliveira, T. and Fernandes, R. (2013) Biochemistry of adipose tissue: an endocrine organ, *Arch Med Sci*, 9, 191-200.
- Cohen, J.C., Vega, G.L. and Grundy, S.M. (1999) Hepatic lipase: new insights from genetic and metabolic studies, *Curr Opin Lipidol*, 10, 259-267.
- Cohen, P. (2002) Protein kinases--the major drug targets of the twenty-first century?, *Nat Rev Drug Discov*, 1, 309-315.
- Collis, T., *et al.* (2001) Relations of stroke volume and cardiac output to body composition: the strong heart study, *Circulation*, 103, 820-825.
- Coppola, A., *et al.* (2009) Effect of weight loss on coronary circulation and adiponectin levels in obese women, *Int J Cardiol*, 134, 414-416.
- Cotillard, A., *et al.* (2013) Dietary intervention impact on gut microbial gene richness, *Nature*, 500, 585-588.
- Cox, L.M., *et al.* (2014) Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences, *Cell*, 158, 705-721.
- Craciun, S. and Balskus, E.P. (2012) Microbial conversion of choline to trimethylamine requires a glycy radical enzyme, *Proc Natl Acad Sci USA*, 109, 21307-21312.
- Crampton, S.P., Davis, J. and Hughes, C.C.W. (2007) Isolation of human umbilical vein endothelial cells (HUVEC), *J Vis Exp*, 183.
- Cuschieri, S. and Mamo, J. (2016) Getting to grips with the obesity epidemic in Europe, *SAGE Open Med*, 4, 2050312116670406.
- Danaceau, J.P., *et al.* (2003) A liquid chromatographic-tandem mass spectrometric method for the analysis of serotonin and related indoles in human whole blood, *J Anal Toxicol*, 27, 440-444.
- David, L.A., *et al.* (2014) Diet rapidly and reproducibly alters the human gut microbiome, *Nature*, 505, 559-563.

- Davidovic, L., *et al.* (2011) A metabolomic and systems biology perspective on the brain of the fragile X syndrome mouse model, *Genome research*, 21, 2190-2202.
- de La Serre, C.B., *et al.* (2010) Propensity to high-fat diet-induced obesity in rats is associated with changes in the gut microbiota and gut inflammation, *Am J Physiol Gastrointest Liver Physiol*, 299, G440-448.
- Delzenne, N.M. and Cani, P.D. (2011) Interaction between obesity and the gut microbiota: relevance in nutrition, *Annu Rev Nutr*, 31, 15-31.
- DeMoss, R.D. and Moser, K. (1969) Tryptophanase in diverse bacterial species, *J Bacteriol*, 98, 167-171.
- Despres, J.-P., *et al.* (2008) Abdominal obesity: the cholesterol of the 21st century?, *Can J Cardiol*, 24 Suppl D, 7D-12D.
- Despres, J.-P., *et al.* (2008) Abdominal obesity and the metabolic syndrome: contribution to global cardiometabolic risk, *Arterioscler Thromb Vasc Biol*, 28, 1039-1049.
- Dharmashankar, K. and Widlansky, M.E. (2010) Vascular endothelial function and hypertension: insights and directions, *Curr Hypertens Rep*, 12, 448-455.
- Dieterle, F., *et al.* (2006) Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in <sup>1</sup>H NMR metabonomics, *Anal Chem*, 78, 4281-4290.
- Diez, J.J. and Iglesias, P. (2003) The role of the novel adipocyte-derived hormone adiponectin in human disease, *Eur J Endocrinol*, 148, 293-300.
- Digby, J.E., *et al.* (2012) Anti-inflammatory effects of nicotinic acid in human monocytes are mediated by GPR109A dependent mechanisms, *Arterioscler Thromb Vasc Biol*, 32, 669-676.
- DiMauro, E.F., *et al.* (2008) Structure-guided design of aminopyrimidine amides as potent, selective inhibitors of lymphocyte specific kinase: synthesis, structure-activity relationships, and inhibition of in vivo T cell activation, *J Med Chem*, 51, 1681-1694.
- Dolphin, C.T., *et al.* (1997) Missense mutation in flavin-containing mono-oxygenase 3 gene, FMO3, underlies fish-odour syndrome, *Nat Genet*, 17, 491-494.
- Dona, A.C., *et al.* (2014) Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping, *Anal Chem*, 86, 9887-9894.
- Dona, A.C., *et al.* (2016) A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments, *Comput Struct Biotechnol J*, 14, 135-153.
- Donini, L.M., *et al.* (2013) How to estimate fat mass in overweight and obese subjects, *Int J Endocrinol*, 2013, 285680.
- Dou, L., *et al.* (2004) The uremic solutes p-cresol and indoxyl sulfate inhibit endothelial proliferation and wound repair, *Kidney Int*, 65, 442-451.
- Dou, L., *et al.* (2007) The uremic solute indoxyl sulfate induces oxidative stress in endothelial cells, *J Thromb Haemostasis : JTH*, 5, 1302-1308.
- Dou, L., *et al.* (2015) The cardiovascular effect of the uremic solute indole-3 acetic acid, *J Am Soc Nephrol*, 26, 876-887.

- Dougan, M.M., *et al.* (2015) Prospective study of body size throughout the life-course and the incidence of endometrial cancer among premenopausal and postmenopausal women, *Int J Cancer*, 137, 625-637.
- Doyle, J.T., *et al.* (1962) Cigarette smoking and coronary heart disease. Combined experience of the Albany and Framingham studies, *NEJM*, 266, 796-801.
- Du, Y.-L., Alkhalaf, L.M. and Ryan, K.S. (2015) In vitro reconstitution of indolmycin biosynthesis reveals the molecular basis of oxazolinone assembly, *Proc Natl Acad Sci USA*, 112, 2717-2722.
- Dumas, M.-E. (2011) The microbial-mammalian metabolic axis: beyond simple metabolism, *Cell Metab*, 13, 489-490.
- Dumas, M.-E., *et al.* (2006) Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice, *Proc Natl Acad Sci USA*, 103, 12511-12516.
- Dumas, M.-E. and Davidovic, L. (2013) Metabolic phenotyping and systems biology approaches to understanding neurological disorders, *F1000Prime Rep*, 5, 18.
- Dumas, M.-E., *et al.* (2006) Assessment of analytical reproducibility of <sup>1</sup>H NMR spectroscopy based metabolomics for large-scale epidemiological research: the INTERMAP Study, *Anal Chem*, 78, 2199-2208.
- Dumas, M.-E., *et al.* (2017) Microbial-Host Co-metabolites Are Prodromal Markers Predicting Phenotypic Heterogeneity in Behavior, Obesity, and Impaired Glucose Tolerance, *Cell Rep*, 20, 136-148.
- Duncan, R.E., *et al.* (2007) Regulation of lipolysis in adipocytes, *Annu Rev Nutr*, 27, 79-101.
- Duranton, F., *et al.* (2012) Normal and pathologic concentrations of uremic toxins, *J Am Soc Nephrol*, 23, 1258-1270.
- Eble, J.A. and Niland, S. (2009) The extracellular matrix of blood vessels, *Curr Pharm Des*, 15, 1385-1400.
- Eckburg, P.B., *et al.* (2005) Diversity of the human intestinal microbial flora, *Science*, 308, 1635-1638.
- Eckers, A., *et al.* (2016) The aryl hydrocarbon receptor promotes aging phenotypes across species, *Sci Rep*, 6, 19618.
- Elliott, P., *et al.* (2015) Urinary metabolic signatures of human adiposity, *Science Transl Med*, 7, 285ra262.
- Ellis, L.M. and Hicklin, D.J. (2008) VEGF-targeted therapy: mechanisms of anti-tumour activity, *Nat Rev Cancer*, 8, 579-591.
- Emoto, T., *et al.* (2017) Characterization of gut microbiota profiles in coronary artery disease patients using data mining analysis of terminal restriction fragment length polymorphism: gut microbiota could be a diagnostic marker of coronary artery disease, *Heart Vessels*, 32, 39-46.
- Emoto, T., *et al.* (2016) Analysis of Gut Microbiota in Coronary Artery Disease Patients: a Possible Link between Gut Microbiota and Coronary Artery Disease, *J Atheroscler Thromb*, 23, 908-921.
- Engeland, A., *et al.* (2003) Body mass index in adolescence in relation to total mortality: 32-year follow-up of 227,000 Norwegian boys and girls, *Am J Epidemiol*, 157, 517-523.
- Engin, A. (2017) *Human Protein Kinases and Obesity: Obesity and Lipotoxicity*.

- Engvall, E. and Perlmann, P. (1971) Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G, *Immunochemistry*, 8, 871-874.
- Eriksson, E.J., *et al.* (1999) *Multi- and Megavariate Data Analysis*. Umetrics Academy, Sweden.
- Eriksson, L., *et al.* (2004) Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm), *Anal Bioanal Chem*, 380, 419-429.
- Erridge, C., *et al.* (2007) A high-fat meal induces low-grade endotoxemia: evidence of a novel mechanism of postprandial inflammation, *Am J Clin Nutr*, 86, 1286-1292.
- Ezzati M, L.A., Rodgers A (2004) A comparative quantification of health risks: global and regional burden of disease attribution to selected major risk factors. Geneva: World Health Organization.
- Fabian, M.A., *et al.* (2005) A small molecule-kinase interaction map for clinical kinase inhibitors, *Nature Biotech*, 23, 329-336.
- Faith, J.J., *et al.* (2013) The long-term stability of the human gut microbiota, *Science*, 341, 1237439.
- Fallani, M., *et al.* (2010) Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics, *J Pediatr Gastroenterol Nutr*, 51, 77-84.
- Farmer, J.J., 3rd, *et al.* (1985) *Escherichia fergusonii* and *Enterobacter taylorae*, two new species of Enterobacteriaceae isolated from clinical specimens, *J Clin Microbiol*, 21, 77-81.
- Faure, V., *et al.* (2006) Elevation of circulating endothelial microparticles in patients with chronic renal failure, *J Thromb Haemostasis : JTH*, 4, 566-573.
- Fedorov, O., *et al.* (2007) A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases, *Proc Natl Acad Sci USA*, 104, 20523-20528.
- Fenton, T.R., *et al.* (2014) Higher versus lower protein intake in formula-fed low birth weight infants, *Cochrane Database Syst Rev*, CD003959.
- Feoktistova, M., Geserick, P. and Leverkus, M. (2016) Crystal Violet Assay for Determining Viability of Cultured Cells, *Cold Spring Harb Protoc*, 2016, pdb.prot087379.
- Ferreira, I., *et al.* (2004) Central fat mass versus peripheral fat and lean mass: opposite (adverse versus favorable) associations with arterial stiffness? The Amsterdam Growth and Health Longitudinal Study, *J Clin Endocrinol Metab*, 89, 2632-2639.
- Fitter, S., *et al.* (2012) Suppression of PDGF-induced PI3 kinase activity by imatinib promotes adipogenesis and adiponectin secretion, *J Mol Endocrinol*, 48, 229-240.
- Flint, H.J., *et al.* (2012) The role of the gut microbiota in nutrition and health, *Nat Rev Gastroenterol Hepatol*, 9, 577-589.
- Fonville, J.M., *et al.* (2010) The evolution of partial least squares models and related chemometric approaches in metabonomics and metabolic phenotyping, *J Chemom*, 24, 636-649.
- Ford, E.S., Williamson, D.F. and Liu, S. (1997) Weight change and diabetes incidence: findings from a national cohort of US adults, *Am J Epidemiol*, 146, 214-222.
- Forslund, K., *et al.* (2015) Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota, *Nature*, 528, 262-266.
- Forstermann, U., Xia, N. and Li, H. (2017) Roles of Vascular Oxidative Stress and Nitric Oxide in the Pathogenesis of Atherosclerosis, *Circ Res*, 120, 713-735.

- Fowler, S.D. and Greenspan, P. (1985) Application of Nile red, a fluorescent hydrophobic probe, for the detection of neutral lipid deposits in tissue sections: comparison with oil red O, *J Histochem Cytochem*, 33, 833-836.
- Fox, C.S., *et al.* (2004) Trends in cardiovascular complications of diabetes, *JAMA*, 292, 2495-2499.
- Foxall, P.J., *et al.* (1993) Analysis of biological fluids using 600 MHz proton NMR spectroscopy: application of homonuclear two-dimensional J-resolved spectroscopy to urine and blood plasma for spectral simplification and assignment, *J Pharm Biomed Anal*, 11, 21-31.
- Franchi, L., *et al.* (2012) NLRC4-driven production of IL-1 $\beta$  discriminates between pathogenic and commensal bacteria and promotes host intestinal defense, *Nat Immunol*, 13, 449-456.
- Frayn, K.N. (2002) Adipose tissue as a buffer for daily lipid flux, *Diabetologia*, 45, 1201-1210.
- Freeman, L. (1977) A set of measures of centrality based upon betweenness, *Sociometry*, 40, 35-41.
- Fridman, R., *et al.* (2012) Increased initiation and growth of tumor cell lines, cancer stem cells and biopsy material in mice using basement membrane matrix protein (Cultrex or Matrigel) co-injection, *Nat Protoc*, 7, 1138-1144.
- Frost, G., *et al.* (2014) The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism, *Nat Commun*, 5, 3611.
- Fukumura, D., *et al.* (2003) Paracrine regulation of angiogenesis and adipocyte differentiation during in vivo adipogenesis, *Circ Res*, 93, e88-97.
- Gao, X., *et al.* (2014) Dietary trimethylamine N-oxide exacerbates impaired glucose tolerance in mice fed a high fat diet, *J Biosci Bioeng*, 118, 476-481.
- Garcia-Perez, I., *et al.* (2012) Urinary metabolic phenotyping the slc26a6 (chloride-oxalate exchanger) null mouse model, *J Prot Res*, 11, 4425-4435.
- Gartland, K.P., *et al.* (1990) Pattern recognition analysis of high resolution  $^1\text{H}$  NMR spectra of urine. A nonlinear mapping approach to the classification of toxicological data, *NMR in Biomedicine*, 3, 166-172.
- Gartland, K.P.R., *et al.* (1991) Application of Pattern-Recognition Methods to the Analysis and Classification of Toxicological Data Derived from Proton Nuclear-Magnetic-Resonance Spectroscopy of Urine, *Mol Pharmacol*, 39, 629-642.
- Gauch, H.G. (2003) *Scientific Method in Practice*. Cambridge University Press.
- Gavaghan, C.L., *et al.* (2000) An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse, *FEBS Lett*, 484, 169-174.
- Gelber, R.P., *et al.* (2007) A prospective study of body mass index and the risk of developing hypertension in men, *Am J Hypertens*, 20, 370-377.
- Genkinger, J.M., *et al.* (2011) A pooled analysis of 14 cohort studies of anthropometric factors and pancreatic cancer risk, *Int J Cancer*, 129, 1708-1717.
- Gentleman, R., *et al.* (2005) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer.
- Gerhard, K. (2008) *Biochemistry of Signal Transduction and Regulation*. Wiley-VCH.

- Gerhardt, H., *et al.* (2003) VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia, *J Cell Biol*, 161, 1163-1177.
- Gerhold, D.L., *et al.* (2002) Gene expression profile of adipocyte differentiation and its regulation by peroxisome proliferator-activated receptor-gamma agonists, *Endocrinology*, 143, 2106-2118.
- Gerken, T., *et al.* (2007) The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase, *Science*, 318, 1469-1472.
- Geuking, M.B., *et al.* (2011) Intestinal bacterial colonization induces mutualistic regulatory T cell responses, *Immunity*, 34, 794-806.
- Gill, S.R., *et al.* (2006) Metagenomic analysis of the human distal gut microbiome, *Science*, 312, 1355-1359.
- Giskeodegard, G.F., *et al.* (2015) Diurnal rhythms in the human urine metabolome during sleep and total sleep deprivation, *Sci Rep*, 5, 14843.
- Giugliano, D., Ceriello, A. and Esposito, K. (2008) Glucose metabolism and hyperglycemia, *Am J Clin Nutr*, 87, 217S-222S.
- Gondouin, B., *et al.* (2013) Indolic uremic solutes increase tissue factor production in endothelial cells by the aryl hydrocarbon receptor pathway, *Kidney Int*, 84, 733-744.
- Goodrich, J.K., *et al.* (2016) Genetic Determinants of the Gut Microbiome in UK Twins, *Cell Host Microbe*, 19, 731-743.
- Goodrich, J.K., *et al.* (2014) Human genetics shape the gut microbiome, *Cell*, 159, 789-799.
- Gordon, D.J., *et al.* (1989) High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies, *Circulation*, 79, 8-15.
- Gordon, H.A. and Bruckner-Kardoss, E. (1961) Effect of normal microbial flora on intestinal surface area, *Am J Physiol*, 201, 175-178.
- Grant, D.S., *et al.* (1991) Intracellular mechanisms involved in basement membrane induced blood vessel differentiation in vitro, *In Vitro Cell Dev Biol*, 27A, 327-336.
- Green, E.D., *et al.* (1995) The human obese (OB) gene: RNA expression pattern and mapping on the physical, cytogenetic, and genetic maps of chromosome 7, *Genome Res*, 5, 5-12.
- Green, H. and Kehinde, O. (1975) An established preadipose cell line and its differentiation in culture. II. Factors affecting the adipose conversion, *Cell*, 5, 19-27.
- Green, H. and Meuth, M. (1974) An established pre-adipose cell line and its differentiation in culture, *Cell*, 3, 127-133.
- Gregoire, F.M., Smas, C.M. and Sul, H.S. (1998) Understanding adipocyte differentiation, *Physiol Rev*, 78, 783-809.
- Groschwitz, K.R. and Hogan, S.P. (2009) Intestinal barrier function: molecular regulation and disease pathogenesis, *J Allergy Clin Immunol*, 124, 3-20; quiz 21-22.
- Guyton, A.C. and Hall, J.E. (2005) *Textbook of Medical Physiology*. Elsevier Saunders.
- Habibi, I., Emamian, E.S. and Abdi, A. (2014) Advanced fault diagnosis methods in molecular networks, *PLoS One*, 9, e108830.
- Hahn, E.L. (1950) "Spin echoes", *Physical Review*, 80, 580-594.

- Hajri, T. and Abumrad, N.A. (2002) Fatty acid transport across membranes: relevance to nutrition and metabolic pathology, *Annu Rev Nutr*, 22, 383-415.
- Han, H.-S., *et al.* (2016) Regulation of glucose metabolism from a liver-centric perspective, *Exp Mol Med*, 48, e218.
- Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation, *Cell*, 144, 646-674.
- Hansson, G.K. (2005) Inflammation, atherosclerosis, and coronary artery disease, *NEJM*, 352, 1685-1695.
- Haralabopoulos, G.C., *et al.* (1994) Inhibitors of basement membrane collagen synthesis prevent endothelial cell alignment in matrigel in vitro and angiogenesis in vivo, *Lab Invest*, 71, 575-582.
- Hasan, J., *et al.* (2004) Quantitative angiogenesis assays in vivo--a review, *Angiogenesis*, 7, 1-16.
- Hayashi, C., *et al.* (2011) Porphyromonas gingivalis accelerates inflammatory atherosclerosis in the innominate artery of ApoE deficient mice, *Atherosclerosis*, 215, 52-59.
- Heeschen, C., *et al.* (2001) Nicotine stimulates angiogenesis and promotes tumor growth and atherosclerosis, *Nat Med*, 7, 833-839.
- Hein, M. and Graver, S. (2013) Tumor cell response to bevacizumab single agent therapy in vitro, *Cancer Cell Int*, 13, 94.
- Heinzmann, S.S., *et al.* (2012) Stability and robustness of human metabolic phenotypes in response to sequential food challenges, *J Prot Res*, 11, 643-655.
- Heiss, M., *et al.* (2015) Endothelial cell spheroids as a versatile tool to study angiogenesis in vitro, *FASEB*, 29, 3076-3084.
- Herbert, S.P. and Stainier, D.Y.R. (2011) Molecular control of endothelial cell behaviour during blood vessel morphogenesis, *Nat Rev Mol Cell Biol*, 12, 551-564.
- Herrera, B.M., Keildson, S. and Lindgren, C.M. (2011) Genetics and epigenetics of obesity, *Maturitas*, 69, 41-49.
- Hewitt, M.J., *et al.* (1993) Hydration of the fat-free body mass in children and adults: implications for body composition assessment, *Am J Physiol*, 265, E88-95.
- Higashi, R.M., *et al.* (2014) Stable isotope-labeled tracers for metabolic pathway elucidation by GC-MS and FT-MS, *Methods Mol Biol*, 1198, 147-167.
- Hildebrand, F., *et al.* (2013) Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice, *Genome Biol*, 14, R4.
- Hildebrandt, M.A., *et al.* (2009) High-fat diet determines the composition of the murine gut microbiome independently of obesity, *Gastroenterology*, 137, 1716-1724.e1711-1712.
- Holm, C. (2003) Molecular mechanisms regulating hormone-sensitive lipase and lipolysis, *Biochem Soc Trans*, 31, 1120-1124.
- Holmes, E., *et al.* (1994) Automatic data reduction and pattern recognition methods for analysis of <sup>1</sup>H nuclear magnetic resonance spectra of human urine from normal and pathological states, *Anal Biochem*, 220, 284-296.

- Holmes, E., *et al.* (1997) 750 MHz  $^1\text{H}$  NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease, *J Pharm Biomed Anal*, 15, 1647-1659.
- Holmes, E., *et al.* (2011) Understanding the role of gut microbiome-host metabolic signal disruption in health and disease, *Trends Microbiol*, 19, 349-359.
- Holmes, E., Wilson, I.D. and Nicholson, J.K. (2008) Metabolic phenotyping in health and disease, *Cell*, 134, 714-717.
- Holmqvist, K., *et al.* (2004) The adaptor protein shb binds to tyrosine 1175 in vascular endothelial growth factor (VEGF) receptor-2 and regulates VEGF-dependent cellular migration, *J Biol Chem*, 279, 22267-22275.
- Hornemann, U., *et al.* (1971) The biosynthesis of indolmycin, *J Am Chem Soc*, 93, 3028-3035.
- Hornemann, U., *et al.* (1970) Demonstration of a C-methylating enzyme in cell free extracts of indolmycin-producing *Streptomyces griseus*, *Biochem Biophys Res Commun*, 39, 594-599.
- Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components, *J Educ Psychol*, 24, 417-441; 498-520.
- Hoyles, L., *et al.* (2017) Microbiome–host systems interactions: Protective effects of propionate upon the blood–brain barrier *bioRxiv* 170548, [Epub ahead of print].
- Huang, L., *et al.* (2012) Development and strategies of VEGFR-2/KDR inhibitors, *Future Med Chem*, 4, 1839-1852.
- Huang, Z., *et al.* (2014) The stem cell factor/Kit signalling pathway regulates mitochondrial function and energy expenditure, *Nat Commun*, 5, 4282.
- Hubbard, S.R. (2004) Juxtamembrane autoinhibition in receptor tyrosine kinases, *Nat Rev Mol Cell Biol*, 5, 464-471.
- Hubert, H.B., *et al.* (1983) Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study, *Circulation*, 67, 968-977.
- Hulkower, K.I. and Herber, R.L. (2011) Cell migration and invasion assays as tools for drug discovery, *Pharmaceutics*, 3, 107-124.
- IDF (2005) IDF Consensus Worldwide Definition of the Metabolic Syndrome.
- Ito, S., *et al.* (2016) Crucial Role of the Aryl Hydrocarbon Receptor (AhR) in Indoxyl Sulfate-Induced Vascular Inflammation, *J Atheroscler Thromb*, 23, 960-975.
- Jaffe, E.A., *et al.* (1973) Culture of human endothelial cells derived from umbilical veins. Identification by morphologic and immunologic criteria, *J Clin Invest*, 52, 2745-2756.
- Jakobsson, H.E., *et al.* (2014) Decreased gut microbiota diversity, delayed *Bacteroidetes* colonisation and reduced Th1 responses in infants delivered by caesarean section, *Gut*, 63, 559-566.
- Janderova, L., *et al.* (2003) Human mesenchymal stem cells as an in vitro model for human adipogenesis, *Obes Res*, 11, 65-74.
- Jardetzky, O. and Roberts, G.C.K. (1981) *NMR in molecular biology*. Academic Press.
- Jean, M. and DeMoss, R.D. (1968) Indolelactate dehydrogenase from *Clostridium sporogenes*, *Can J Microbiol*, 14, 429-435.

- Jeffery, I.B., *et al.* (2012) Categorization of the gut microbiota: enterotypes or gradients?, *Nat Rev Microbiol*, 10, 591-592.
- Jeffery, I.B., Lynch, D.B. and O' Toole, P.W. (2016) Composition and temporal stability of the gut microbiota in older persons, *ISMEJ*, 10, 170-182.
- Jensen, M.T., Cox, R.P. and Jensen, B.B. (1995) 3-Methylindole (skatole) and indole production by mixed populations of pig fecal bacteria, *Appl Environ Microbiol*, 61, 3180-3184.
- Jesmin, S., *et al.* (2005) Age-related changes in cardiac expression of VEGF and its angiogenic receptor KDR in stroke-prone spontaneously hypertensive rats, *Mol Cell Biochem*, 272, 63-73.
- Jimenez, E., *et al.* (2005) Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section, *Curr Microbiol*, 51, 270-274.
- Jimenez, E., *et al.* (2008) Is meconium from healthy newborns actually sterile?, *Res Microbiol*, 159, 187-193.
- Jin, U.-H., *et al.* (2014) Microbiome-derived tryptophan metabolites and their aryl hydrocarbon receptor-dependent agonist and antagonist activities, *Molecular Pharmacol*, 85, 777-788.
- Jo, J., *et al.* (2009) Hypertrophy and/or Hyperplasia: Dynamics of Adipose Tissue Growth, *PLoS Comput Biol*, 5, e1000324.
- Johnson, M., *et al.* (2008) NCBI BLAST: a better web interface, *Nucleic Acids Res*, 36, W5-9.
- Jolliffe, I.T. (2002) *Principal Component Analysis*. Springer Series in Statistics. Springer.
- Jonsson, A.L. and Backhed, F. (2017) Role of gut microbiota in atherosclerosis, *Nat Rev Cardiol*, 14, 79-87.
- Jung, Y.D., *et al.* (2002) The role of the microenvironment and intercellular cross-talk in tumor angiogenesis, *Semin Cancer Biol*, 12, 105-112.
- Jungas, R.L., Halperin, M.L. and Brosnan, J.T. (1992) Quantitative analysis of amino acid oxidation and related gluconeogenesis in humans, *Physiol Rev*, 72, 419-448.
- Kam, Y., *et al.* (2008) A novel circular invasion assay mimics in vivo invasive behavior of cancer cell lines and distinguishes single-cell motility in vitro, *BMC Cancer*, 8, 198.
- Kamat, B.R., *et al.* (1987) Neovascularization and coronary atherosclerotic plaque: cinematographic localization and quantitative histologic analysis, *Hum Pathol*, 18, 1036-1042.
- Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res*, 28, 27-30.
- Kang, Z., *et al.* (2013) Protocatechuic acid induces angiogenesis through PI3K-Akt-eNOS-VEGF signalling pathway, *Basic Clin Pharmacol Toxicol*, 113, 221-227.
- Kannel, W.B., Castelli, W.P. and Gordon, T. (1979) Cholesterol in the prediction of atherosclerotic disease. New perspectives based on the Framingham study, *Annals of internal medicine*, 90, 85-91.
- Karaman, M.W., *et al.* (2008) A quantitative analysis of kinase inhibitor selectivity, *Nat Biotech*, 26, 127-132.
- Karlsson, F.H., *et al.* (2012) Symptomatic atherosclerosis is associated with an altered gut metagenome, *Nat Commun*, 3, 1245.

- Karlsson, F.H., *et al.* (2013) Gut metagenome in European women with normal, impaired and diabetic glucose control, *Nature*, 498, 99-103.
- Karnwal, A. (2009) Production of indole acetic acid by Fluorescent *Pseudomonas* in the presence of L-tryptophan and rice root exudates, *J Plant Pathol*, 1, 61-63.
- Keeler, J. (2002) Understanding NMR spectroscopy. University of Cambridge.
- Kersten, S. (2001) Mechanisms of nutritional and hormonal regulation of lipogenesis, *EMBO Rep*, 2, 282-286.
- Keun, H.C., *et al.* (2002) Analytical reproducibility in (1)H NMR-based metabonomic urinalysis, *Chem Res Toxicol*, 15, 1380-1386.
- Keun, H.C., *et al.* (2009) Serum molecular signatures of weight change during early breast cancer chemotherapy, *Clin Cancer Res*, 15, 6716-6723.
- Keys, A., *et al.* (1972) Indices of relative weight and obesity, *J Chronic Dis*, 25, 329-343.
- Khurana, R., *et al.* (2005) Role of angiogenesis in cardiovascular disease: a critical appraisal, *Circulation*, 112, 1813-1824.
- Kim, B.-S., Jeon, Y.-S. and Chun, J. (2013) Current status and future promise of the human microbiome, *Pediatr Gastroenterol Hepatol Nutr*, 16, 71-79.
- Kim, S.H. and Reaven, G.M. (2004) The metabolic syndrome: one step forward, two steps back, *Diab Vasc Dis Res*, 1, 68-75.
- Kimura, I., *et al.* (2013) The gut microbiota suppresses insulin-mediated fat accumulation via the short-chain fatty acid receptor GPR43, *Nat Commun*, 4, 1829.
- Kinexus (2017) PhosphoNET - Human Phosphosite Knowledgebase.
- Kitchen, D.B., *et al.* (2004) Docking and scoring in virtual screening for drug discovery: methods and applications, *Nat Rev Drug Discov*, 3, 935-949.
- Knights, D., *et al.* (2014) Rethinking "enterotypes", *Cell host & microbe*, 16, 433-437.
- Koenig, J.E., *et al.* (2011) Succession of microbial consortia in the developing infant gut microbiome, *Proc Natl Acad Sci USA*, 108 Suppl 1, 4578-4585.
- Koeth, R.A., *et al.* (2014) gamma-Butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO, *Cell Metab*, 20, 799-812.
- Koo, H.-J., *et al.* (2015) Introduction of Methyl Groups at C2 and C6 Positions Enhances the Antiangiogenesis Activity of Curcumin, *Sci Rep*, 5, 14205.
- Korb, O., Stutzle, T. and Exner, T.E. (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS, *J Chem Inf Model*, 49, 84-96.
- Kornev, A.P., *et al.* (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism, *Proc Natl Acad Sci USA*, 103, 17783-17788.
- Ku, C.S., *et al.* (1994) Left ventricular filling in young normotensive obese adults, *Am J Cardiol*, 73, 613-615.
- Kubis, N., *et al.* (2002) Role of microvascular rarefaction in the increased arterial pressure in mice lacking for the endothelial nitric oxide synthase gene (eNOS3pt/-), *J Hypertens*, 20, 1581-1587.
- Kubota, Y., *et al.* (1988) Role of laminin and basement membrane in the morphological differentiation of human endothelial cells into capillary-like structures, *J Cell Biol*, 107, 1589-1598.

- Kuri-Harcuch, W., Wise, L.S. and Green, H. (1978) Interruption of the adipose conversion of 3T3 cells by biotin deficiency: differentiation without triglyceride accumulation, *Cell*, 14, 53-59.
- Lagier, J.-C., *et al.* (2016) Culture of previously uncultured members of the human gut microbiota by culturomics, *Nat Microbiol*, 1, 16203.
- Lamallice, L., Le Boeuf, F. and Huot, J. (2007) Endothelial cell migration during angiogenesis, *Circ Res*, 100, 782-794.
- Le Chatelier, E., *et al.* (2013) Richness of human gut microbiome correlates with metabolic markers, *Nature*, 500, 541-546.
- Le Chatelier, E., *et al.* (2013) Richness of human gut microbiome correlates with metabolic markers, *Nature*, 500, 541-546.
- Lees, H.J., *et al.* (2013) Hippurate: the natural history of a mammalian-microbial cometabolite, *J Prot Res*, 12, 1527-1546.
- Leon-Mimila, P., *et al.* (2013) Contribution of common genetic variants to obesity and obesity-related traits in mexican children and adults, *PLoS One*, 8, e70640.
- Lewington, S., *et al.* (2002) Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies, *Lancet*, 360, 1903-1913.
- Ley, R.E., *et al.* (2006) Microbial ecology: human gut microbes associated with obesity, *Nature*, 444, 1022-1023.
- Li, J., *et al.* (2014) An integrated catalog of reference genes in the human gut microbiome, *Nature Biotech*, 32, 834-841.
- Libby, P., Ridker, P.M. and Hansson, G.K. (2011) Progress and challenges in translating the biology of atherosclerosis, *Nature*, 473, 317-325.
- Libby, P., Ridker, P.M. and Maseri, A. (2002) Inflammation and atherosclerosis, *Circulation*, 105, 1135-1143.
- Lin, G.-H., Chang, C.-Y. and Lin, H.-R. (2015) Systematic profiling of indole-3-acetic acid biosynthesis in bacteria using LC-MS/MS, *J Chromatogr B Analyt Technol Biomed Life Sci*, 988, 53-58.
- Lindon, J.C., Holmes, E. and Nicholson, J.K. (2003) So what's the deal with metabonomics?, *Anal Chem*, 75, 384A-391A.
- Lindon, J.C., Holmes, E. and Nicholson, J.K. (2004) Metabonomics: systems biology in pharmaceutical research and development, *Curr Opin Mol Ther*, 6, 265-272.
- Lindon, J.C. and Nicholson, J.K. (2008) Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics, *Annu Rev Anal Chem (Palo Alto Calif)*, 1, 45-69.
- Liou, A.P., *et al.* (2013) Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity, *Science Transl Med*, 5, 178ra141.
- Liu, J., *et al.* (2002) Caveolin-1 expression enhances endothelial capillary tubule formation, *J Biol Chem*, 277, 10661-10668.
- Lloyd-Price, J., *et al.* (2017) Strains, functions and dynamics in the expanded Human Microbiome Project, *Nature*, 550, 61-66.

- Lombard, G.L. and Dowell, V.R., Jr. (1983) Comparison of three reagents for detecting indole production by anaerobic bacteria in microtest systems, *J Clin Microbiol*, 18, 609-613.
- Lozupone, C.A., *et al.* (2012) Diversity, stability and resilience of the human gut microbiota, *Nature*, 489, 220-230.
- Lusis, A.J. (2000) Atherosclerosis, *Nature*, 407, 233-241.
- Lustig, R.H. (2003) Autonomic dysfunction of the beta-cell and the pathogenesis of obesity, *Rev Endocr Metab Disord*, 4, 23-32.
- Lutgendorff, F., Akkermans, L.M.A. and Soderholm, J.D. (2008) The role of microbiota and probiotics in stress-induced gastro-intestinal damage, *Curr Mol Med*, 8, 282-298.
- Ma, Y., *et al.* (2013) Obesity and risk of colorectal cancer: a systematic review of prospective studies, *PLoS One*, 8, e53916.
- Mac Gabhann, F. and Popel, A.S. (2008) Systems biology of vascular endothelial growth factors, *Microcirculation*, 15, 715-738.
- MacDougald, O.A., *et al.* (1995) Regulated expression of the obese gene product (leptin) in white adipose tissue and 3T3-L1 adipocytes, *Proc Natl Acad Sci USA*, 92, 9034-9037.
- Maffei, M., *et al.* (1995) Increased expression in adipocytes of ob RNA in mice with lesions of the hypothalamus and with mutations at the db locus, *Proc Natl Acad Sci USA*, 92, 6957-6960.
- Malinda, K.M. (2001) In vivo matrigel migration and angiogenesis assays, *Methods Mol Med*, 46, 47-52.
- Manning, G., *et al.* (2002) The protein kinase complement of the human genome, *Science*, 298, 1912-1934.
- Manulis, S., *et al.* (1994) Biosynthesis of indole-3-acetic acid via the indole-3-acetamide pathway in *Streptomyces* spp, *Microbiology*, 140 ( Pt 5), 1045-1050.
- Margetic, S., *et al.* (2002) Leptin: a review of its peripheral actions and interactions, *Int J Obes Relat Metab Disord*, 26, 1407-1433.
- Marin, V., *et al.* (2001) Endothelial cell culture: protocol to obtain and cultivate human umbilical endothelial cells, *J Immunol Methods*, 254, 183-190.
- Markley, J.L., *et al.* (2017) The future of NMR-based metabolomics, *Curr Opin Biotechnol*, 43, 34-40.
- Marletta, M.A., *et al.* (1988) Macrophage oxidation of L-arginine to nitrite and nitrate: nitric oxide is an intermediate, *Biochemistry*, 27, 8706-8711.
- Mastyugin, V., *et al.* (2005) A quantitative high-throughput endothelial cell migration assay, *Drug Discov Today*, 1, 43-49.
- Medscape (2014) Medscape Lab Values, Normal Adult.
- Mehta, J.L., *et al.* (2007) Value of CRP in coronary risk determination, *Indian Heart J*, 59, 173-177.
- Meir, K.S. and Leitersdorf, E. (2004) Atherosclerosis in the apolipoprotein-E-deficient mouse: a decade of progress, *Arterioscler Thromb Vasc Biol*, 24, 1006-1014.
- Melo, L.G., *et al.* (2004) Endothelium-targeted gene and cell-based therapies for cardiovascular disease, *Arterioscler Thromb Vasc Biol*, 24, 1761-1774.
- Miles, A.M., *et al.* (1996) Determination of nitric oxide using fluorescence spectroscopy, *Methods Enzymol*, 268, 105-120.

- Miles, J.M. and Jensen, M.D. (1993) Does glucagon regulate adipose tissue lipolysis?, *J Clin Endocrinol Metab*, 77, 5A-5B.
- Mitchell, M. (1996) *An Introduction to Genetic Algorithms*. The MIT Press.
- Miyazawa-Hoshimoto, S., *et al.* (2003) Elevated serum vascular endothelial growth factor is associated with visceral fat accumulation in human obese subjects, *Diabetologia*, 46, 1483-1488.
- Mohammed, N., Onodera, R. and Or-Rashid, M.M. (2003) Degradation of tryptophan and related indolic compounds by ruminal bacteria, protozoa and their mixture in vitro, *Amino Acids*, 24, 73-80.
- Morales-Ruiz, M., *et al.* (2000) Vascular endothelial growth factor-stimulated actin reorganization and migration of endothelial cells is regulated via the serine/threonine kinase Akt, *Circ Res*, 86, 892-896.
- Moreno, P.R., *et al.* (2004) Plaque neovascularization is increased in ruptured atherosclerotic lesions of human aorta: implications for plaque vulnerability, *Circulation*, 110, 2032-2038.
- Moulton, K.S., *et al.* (1999) Angiogenesis inhibitors endostatin or TNP-470 reduce intimal neovascularization and plaque growth in apolipoprotein E-deficient mice, *Circulation*, 99, 1726-1732.
- MSD (2011) Phospho(Ser473)/Total Akt Assay Whole Cell Lysate Kit (Product Insert). Gaithersburg.
- Mueller, W.M., *et al.* (2000) Effects of metformin and vanadium on leptin secretion from cultured rat adipocytes, *Obes Res*, 8, 530-539.
- Muoio, D.M. and Newgard, C.B. (2008) Mechanisms of disease: Molecular and metabolic mechanisms of insulin resistance and beta-cell failure in type 2 diabetes, *Nat Rev Mol Cell Biol*, 9, 193-205.
- Murphy, E.F., *et al.* (2010) Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models, *Gut*, 59, 1635-1642.
- Muruganandham, M., *et al.* (2006) Preclinical evaluation of tumor microvascular response to a novel antiangiogenic/antitumor agent RO0281501 by dynamic contrast-enhanced MRI at 1.5 T, *Mol Cancer Ther*, 5, 1950-1957.
- Musri, M.M. and Parrizas, M. (2012) Epigenetic regulation of adipogenesis, *COCN*, 15, 342-349.
- Muteliefu, G., *et al.* (2009) Indoxyl sulphate induces oxidative stress and the expression of osteoblast-specific proteins in vascular smooth muscle cells, *Nephrol Dial Transplant*, 24, 2051-2058.
- Nagayama, K., *et al.* (1977) Two-dimensional J-resolved <sup>1</sup>H NMR spectroscopy for studies of biological macromolecules, *Biochem Biophys Res Commun*, 78, 99-105.
- Newman, P.J., *et al.* (1990) PECAM-1 (CD31) cloning and relation to adhesion molecules of the immunoglobulin gene superfamily, *Science*, 247, 1219-1222.
- Nicholson, J.K., *et al.* (2002) Metabonomics: a platform for studying drug toxicity and gene function, *Nat Rev Drug Discov*, 1, 153-161.
- Nicholson, J.K., *et al.* (1995) 750 MHz <sup>1</sup>H and <sup>1</sup>H-<sup>13</sup>C NMR spectroscopy of human blood plasma, *Anal Chem*, 67, 793-811.
- Nicholson, J.K., Holmes, E. and Elliott, P. (2008) The metabolome-wide association study: a new look at human disease risk factors, *J Proteome Res*, 7, 3637-3638.

- Nicholson, J.K., Holmes, E. and Lindon, J.C. (2007) *Metabonomics and Metabolomics Techniques and Their Applications in Mammalian Systems. In The Handbook of Metabonomics and Metabolomics*. Elsevier.
- Nicholson, J.K., Holmes, E. and Wilson, I.D. (2005) Gut microorganisms, mammalian metabolism and personalized health care, *Nat Rev Microbiol*, 3, 431-438.
- Nicholson, J.K., Lindon, J.C. and Holmes, E. (1999) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data, *Xenobiotica*, 29, 1181-1189.
- Nicholson, J.K., *et al.* (1984) Monitoring metabolic disease by proton NMR of urine, *Lancet*, 2, 751-752.
- Nicholson, J.K., Wilson, I.D. and Lindon, J.C. (2011) Pharmacometabonomics as an effector for personalized medicine, *Pharmacogenomics*, 12, 103-111.
- Noiri, E., *et al.* (1997) Permissive role of nitric oxide in endothelin-induced migration of endothelial cells, *J Biol Chem*, 272, 1747-1752.
- O' Toole, P.W. and Jeffery, I.B. (2015) Gut microbiota and aging, *Science*, 350, 1214-1215.
- OECD (2012) 'Overweight and obesity among adults,' in *Health at a Glance: Europe 2012*.
- Oehler, M.K. and Bicknell, R. (2000) The promise of anti-angiogenic cancer therapy, *Br J Cancer*, 82, 749-752.
- Olsson, A.-K., *et al.* (2006) VEGF receptor signalling - in control of vascular function, *Nat Rev Mol Cell Biol*, 7, 359-371.
- Orton, R.J., *et al.* (2005) Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway, *Biochem J*, 392, 249-261.
- Oscar, T.P. (1991) Glucagon-stimulated lipolysis of primary cultured broiler adipocytes, *Poult Sci*, 70, 326-332.
- Pan, H., Guo, J. and Su, Z. (2014) Advances in understanding the interrelations between leptin resistance and obesity, *Physiol Behav*, 130, 157-169.
- Papin, J.A., *et al.* (2005) Reconstruction of cellular signalling networks and analysis of their properties, *Nat Rev Mol Cell Biol*, 6, 99-111.
- Passaniti, A., *et al.* (1992) A simple, quantitative method for assessing angiogenesis and antiangiogenic agents using reconstituted basement membrane, heparin, and fibroblast growth factor, *Lab Invest*, 67, 519-528.
- Patten, C.L., Blakney, A.J.C. and Coulson, T.J.D. (2013) Activity, distribution and function of indole-3-acetic acid biosynthetic pathways in bacteria, *Crit Rev Microbiol*, 39, 395-415.
- Patten, C.L. and Glick, B.R. (2002) Role of *Pseudomonas putida* indoleacetic acid in development of the host plant root system, *Appl Environ Microbiol*, 68, 3795-3801.
- Pearson, K. (1901) On Lines and Planes of Closest Fit to Systems of Points in Space, *Philosoph Mag*, 2 559-572.
- Phillips, L.K. and Prins, J.B. (2008) The link between abdominal obesity and the metabolic syndrome, *Curr Hypertens Rep*, 10, 156-164.
- Pi-Sunyer, F.X., *et al.* (1999) Therapeutic controversy: Obesity - a modern-day epidemic, *J Clin Endocrinol Metab*, 84, 3-12.

- Pilkis, S.J., el-Maghrabi, M.R. and Claus, T.H. (1988) Hormonal regulation of hepatic gluconeogenesis and glycolysis, *Annu Rev Biochem*, 57, 755-783.
- Plagemann, A., *et al.* (2009) Hypothalamic proopiomelanocortin promoter methylation becomes altered by early overfeeding: an epigenetic model of obesity and the metabolic syndrome, *J Physiol*, 587, 4963-4976.
- Ponten, F., Jirstrom, K. and Uhlen, M. (2008) The Human Protein Atlas - a tool for pathology, *J Pathol*, 216, 387-393.
- Pories, W.J., Dohm, L.G. and Mansfield, C.J. (2010) Beyond the BMI: the search for better guidelines for bariatric surgery, *Obesity*, 18, 865-871.
- Posma, J.M., *et al.* (2012) Subset optimization by reference matching (STORM): an optimized statistical approach for recovery of metabolic biomarker structural information from <sup>1</sup>H NMR spectra of biofluids, *Anal Chem*, 84, 10694-10701.
- Posma, J.M., *et al.* (2017) Integrated Analytical and Statistical Two-Dimensional Spectroscopy Strategy for Metabolite Identification: Application to Dietary Biomarkers, *Anal Chem*, 89, 3300-3309.
- Posner, I., *et al.* (1994) Kinetics of inhibition by tyrphostins of the tyrosine kinase activity of the epidermal growth factor receptor and analysis by a new computer program, *Mol Pharmacol*, 45, 673-683.
- Potente, M., Gerhardt, H. and Carmeliet, P. (2011) Basic and therapeutic aspects of angiogenesis, *Cell*, 146, 873-887.
- Poujade, M., *et al.* (2007) Collective migration of an epithelial monolayer in response to a model wound, *Proc Natl Acad Sci USA*, 104, 15988-15993.
- Poulos, S.P., Dodson, M.V. and Hausman, G.J. (2010) Cell line models for differentiation: preadipocytes and adipocytes, *Exp Biol Med*, 235, 1185-1193.
- Premji, S.S., Fenton, T.R. and Sauve, R.S. (2006) Higher versus lower protein intake in formula-fed low birth weight infants, *Coch Datab Syst Rev*, CD003959.
- Psichas, A., *et al.* (2015) The short chain fatty acid propionate stimulates GLP-1 and PYY secretion via free fatty acid receptor 2 in rodents, *Int J Obesity (2005)*, 39, 424-429.
- Pussinen, P.J., *et al.* (2011) Endotoxemia is associated with an increased risk of incident diabetes, *Diabetes Care*, 34, 392-397.
- Qin, K., *et al.* (2011) Inactive-state preassembly of G(q)-coupled receptors and G(q) heterotrimers, *Nat Chem Biol*, 7, 740-747.
- R Core Team (2013) R: A language and environment for statistical computing. Vienna, Austria.
- Ragel, B.T., *et al.* (2007) Celecoxib inhibits meningioma tumor growth in a mouse xenograft model, *Cancer*, 109, 588-597.
- Rahat, M.A., Hemmerlein, B. and Iragavarapu-Charyulu, V. (2014) The regulation of angiogenesis by tissue cell-macrophage interactions, *Front Physiol*, 5, 262.
- Ramirez-Puebla, S.T., *et al.* (2013) Gut and root microbiota commonalities, *Appl Environ Microbiol*, 79, 2-9.
- Rehman, A., *et al.* (2011) Nod2 is essential for temporal development of intestinal microbial communities, *Gut*, 60, 1354-1362.

- Rezzi, S., *et al.* (2007) Human metabolic phenotypes link directly to specific dietary preferences in healthy individuals, *J Proteome Res*, 6, 4469-4477.
- Ridaura, V.K., *et al.* (2013) Gut microbiota from twins discordant for obesity modulate metabolism in mice, *Science*, 341, 1241214.
- Risau, W. and Flamme, I. (1995) Vasculogenesis, *Annu Rev Cell Dev Biol*, 11, 73-91.
- Rivard, A., *et al.* (1999) Rescue of diabetes-related impairment of angiogenesis by intramuscular gene therapy with adeno-VEGF, *Am J Pathol*, 154, 355-363.
- Rivollier, A., *et al.* (2012) Inflammation switches the differentiation program of Ly6Chi monocytes from antiinflammatory macrophages to inflammatory dendritic cells in the colon, *J Exp Med*, 209, 139-155.
- Roberts, J. and Rosenfeld, H.J. (1977) Isolation, crystallization, and properties of indolyl-3-alkane alpha-hydroxylase. A novel tryptophan-metabolizing enzyme, *J Biol Chem*, 252, 2640-2647.
- Rodriguez-Martinez A, A.R., Posma JM, Neves AL, Gauguier D, Nicholson JK, Dumas ME (2016) MetaboSignal: a network-based approach for topological analysis of metabotype regulation via metabolic and signaling pathways, *Bioinformatics*, pii: btw697.
- Rodriguez-Martinez, A., *et al.* (2017) MWAStools: an R/Bioconductor package for metabolome-wide association studies, *Bioinformatics*, btx477.
- Rodriguez-Martinez, A., *et al.* (2017) J-Resolved 1H NMR 1D-Projections for Large-Scale Metabolic Phenotyping Studies: Application to Blood Plasma Analysis., *Anal Chem*.
- Rodriguez, J.M., *et al.* (2015) The composition of the gut microbiota throughout life, with an emphasis on early life, *Microb Ecol Health Dis*, 26, 26050.
- Rodwell, V., *et al.* (2015) *Harper's Illustrated Biochemistry*. McGraw-Hill.
- Romero-Corral, A., *et al.* (2008) Accuracy of body mass index in diagnosing obesity in the adult general population, *Int J Obesity (2005)*, 32, 959-966.
- Rorsman, P. and Braun, M. (2013) Regulation of insulin secretion in human pancreatic islets, *Annu Rev Physiol*, 75, 155-179.
- Rosen, E.D. and MacDougald, O.A. (2006) Adipocyte differentiation from the inside out, *Nat Rev Mol Cell Biol*, 7, 885-896.
- Rossmeislova, L., *et al.* (2013) Weight loss improves the adipogenic capacity of human preadipocytes and modulates their secretory profile, *Diabetes*, 62, 1990-1995.
- Rubin, C.S., *et al.* (1978) Development of hormone receptors and hormonal responsiveness in vitro. Insulin receptors and insulin sensitivity in the preadipocyte and adipocyte forms of 3T3-L1 cells, *J Biol Chem*, 253, 7570-7578.
- Ruiz-Ojeda, F.J., *et al.* (2016) Cell Models and Their Application for Studying Adipogenic Differentiation in Relation to Obesity: A Review, *Int J Mol Sci*, 17.
- Rupnick, M.A., *et al.* (2002) Adipose tissue mass can be regulated through the vasculature, *Proc Natl Acad Sci USA*, 99, 10730-10735.
- Russell, T.R. and Ho, R. (1976) Conversion of 3T3 fibroblasts into adipose cells: triggering of differentiation by prostaglandin F2alpha and 1-methyl-3-isobutyl xanthine, *Proc Natl Acad Sci USA*, 73, 4516-4520.

- Russell, W.R., *et al.* (2013) Major phenylpropanoid-derived metabolites in the human gut can arise from microbial fermentation of protein, *Mol Nutr Food Res*, 57, 523-535.
- Russell, W.R., *et al.* (2013) Colonic bacterial metabolites and human health, *Curr Opin Microbiol*, 16, 246-254.
- Sachdev, D.P., *et al.* (2009) Isolation and characterization of indole acetic acid (IAA) producing *Klebsiella pneumoniae* strains from rhizosphere of wheat (*Triticum aestivum*) and their effect on plant growth, *Indian J Exp Biol*, 47, 993-1000.
- Salminen, S., *et al.* (2004) Influence of mode of delivery on gut microbiota composition in seven year old children, *Gut*, 53, 1388-1389.
- Samad, F., Pandey, M. and Loskutoff, D.J. (1998) Tissue factor gene expression in the adipose tissues of obese mice, *Proc Natl Acad Sci USA*, 95, 7591-7596.
- Sane, D.C., Anton, L. and Brosnihan, K.B. (2004) Angiogenic growth factors and hypertension, *Angiogenesis*, 7, 193-201.
- Santacruz, A., *et al.* (2010) Gut microbiota composition is associated with body weight, weight gain and biochemical parameters in pregnant women, *Br J Nutr*, 104, 83-92.
- Saponaro, C., *et al.* (2015) The Subtle Balance between Lipolysis and Lipogenesis: A Critical Point in Metabolic Homeostasis, *Nutrients*, 7, 9453-9474.
- Sasaki-Imamura, T., *et al.* (2011) Molecular basis of indole production catalyzed by tryptophanase in the genus *Prevotella*, *FEMS Microbiol Lett*, 322, 51-59.
- Schloss, P.D. and Handelsman, J. (2004) Status of the microbial census, *Microbiol Mol Biol Rev*, 68, 686-691.
- Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis, *Nat Methods*, 9, 671-675.
- Schoell, W.M., *et al.* (1997) Tumor angiogenesis as a prognostic factor in ovarian carcinoma: quantification of endothelial immunoreactivity by image analysis, *Cancer*, 80, 2257-2262.
- Seldin, M.M., *et al.* (2016) Trimethylamine N-Oxide Promotes Vascular Inflammation Through Signaling of Mitogen-Activated Protein Kinase and Nuclear Factor-kappaB, *J Am Heart Assoc*, 5.
- Sender, R., Fuchs, S. and Milo, R. (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body, *PLoS Biol*, 14, e1002533.
- Senthong, V., *et al.* (2016) Plasma Trimethylamine N-Oxide, a Gut Microbe-Generated Phosphatidylcholine Metabolite, Is Associated With Atherosclerotic Burden, *J Am Coll Cardiol*, 67, 2620-2628.
- Seymour, G. (1993) *Predictive Inference*. New York.
- Shah, N.R. and Braverman, E.R. (2012) Measuring adiposity in patients: the utility of body mass index (BMI), percent body fat, and leptin, *PLoS One*, 7, e33308.
- Shamai, L., *et al.* (2011) Association of body mass index and lipid profiles: evaluation of a broad spectrum of body mass index patients including the morbidly obese, *Obes Surg*, 21, 42-47.
- Shannon, P., *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res*, 13, 2498-2504.

- Shi, Y. and Burn, P. (2004) Lipid metabolic enzymes: emerging drug targets for the treatment of obesity, *Nat Rev Drug Discov*, 3, 695-710.
- Shibuya, M. (2011) Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies, *Genes Cancer*, 2, 1097-1105.
- Shimba, S., *et al.* (2003) Transcriptional regulation of the AhR gene during adipose differentiation, *Biol Pharm Bull*, 26, 1266-1271.
- Shimba, S., *et al.* (1998) Depletion of arylhydrocarbon receptor during adipose differentiation in 3T3-L1 cells, *Biochem Biophys Res Commun*, 249, 131-137.
- Shiojima, I., *et al.* (2005) Disruption of coordinated cardiac hypertrophy and angiogenesis contributes to the transition to heart failure, *J Clin Invest*, 115, 2108-2118.
- Simpson, J.F., *et al.* (1996) Endothelial area as a prognostic indicator for invasive breast carcinoma, *Cancer*, 77, 2077-2085.
- Smith, T. (1897) A modification of the method for determining the production of indol by bacteria, *J Exp Med*, 2, 543-547.
- Smolinska, A., *et al.* (2012) NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review, *Anal Chim Acta*, 750, 82-97.
- Speedie, M.K., Hornemann, U. and Floss, H.G. (1975) Isolation and characterization of tryptophan transaminase and indolepyruvate C-methyltransferase. Enzymes involved in indolmycin biosynthesis in *Streptomyces griseus*, *J Biol Chem*, 250, 7819-7825.
- Speliotes, E.K., *et al.* (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index, *Nat Genet*, 42, 937-948.
- Sridharan, G.V., *et al.* (2014) Prediction and quantification of bioactive microbiota metabolites in the mouse gut, *Nat Commun*, 5, 5492.
- Stamler, J., Stamler, R. and Neaton, J.D. (1993) Blood pressure, systolic and diastolic, and cardiovascular risks. US population data, *Arch Intern Med*, 153, 598-615.
- Staton, C.A., Reed, M.W.R. and Brown, N.J. (2009) A critical analysis of current in vitro and in vivo angiogenesis assays, *Int J Exp Pathol*, 90, 195-221.
- Staton, C.A., *et al.* (2004) Current methods for assaying angiogenesis in vitro and in vivo, *Int J Exp Pathol*, 85, 233-248.
- Steel, R.G.D. and Torrie, J.H. (1960) *Principles and Procedures of Statistics with Special Reference to the Biological Sciences*. McGraw-Hill.
- Stella, C., *et al.* (2006) Susceptibility of human metabolic phenotypes to dietary modulation, *J Proteome Res*, 5, 2780-2788.
- Stockler-Pinto, M.B., *et al.* (2016) The uremic toxin indoxyl sulfate exacerbates reactive oxygen species production and inflammation in 3T3-L1 adipose cells, *Free Radic Res*, 50, 337-344.
- Strilic, B., *et al.* (2009) The molecular basis of vascular lumen formation in the developing mouse aorta, *Dev Cell*, 17, 505-515.
- Student, A.K., Hsu, R.Y. and Lane, M.D. (1980) Induction of fatty acid synthetase synthesis in differentiating 3T3-L1 preadipocytes, *J Biol Chem*, 255, 4745-4750.

- Stull, T.L., *et al.* (1995) Production and oxidation of indole by *Haemophilus influenzae*, *J Biol Chem*, 270, 5-8.
- Suez, J., *et al.* (2014) Artificial sweeteners induce glucose intolerance by altering the gut microbiota, *Nature*, 514, 181-186.
- Sul, H.S., *et al.* (2000) Regulation of the fatty acid synthase promoter by insulin, *J Nutr*, 130, 315S-320S.
- Sun, J., *et al.* (2013) Evaluating effects of penicillin treatment on the metabolome of rats, *J Chromatogr B Analyt Technol Biomed Life Sci*, 932, 134-143.
- Suzuki, K., *et al.* (2010) The sensing of environmental stimuli by follicular dendritic cells promotes immunoglobulin A generation in the gut, *Immunity*, 33, 71-83.
- Swidsinski, A., *et al.* (2005) Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence in situ hybridization study in mice, *World J Gastroenterology*, 11, 1131-1140.
- Sylvester, J.J. (1878) Chemistry and Algebra, *Nature*, 17, 284.
- Tahergorabi, Z. and Khazaei, M. (2012) A review on angiogenesis and its assays, *Iran J Basic Med Sci*, 15, 1110-1126.
- Takahashi, H. and Shibuya, M. (2005) The vascular endothelial growth factor (VEGF)/VEGF receptor system and its role under physiological and pathological conditions, *Clin Sci*, 109, 227-241.
- Takahashi, T., *et al.* (2001) A single autophosphorylation site on KDR/Flk-1 is essential for VEGF-A-dependent activation of PLC-gamma and DNA synthesis in vascular endothelial cells, *EMBO J*, 20, 2768-2778.
- Takai, K., *et al.* (1977) Crystalline hemoprotein from *Pseudomonas* that catalyzes oxidation of side chain of tryptophan and other indole derivatives, *J Biol Chem*, 252, 2648-2656.
- Tamarat, R., *et al.* (2004) Impairment in ischemia-induced neovascularization in diabetes: bone marrow mononuclear cell dysfunction and therapeutic potential of placenta growth factor treatment, *Am J Pathol*, 164, 457-466.
- Tan, W.H., Popel, A.S. and Mac Gabhann, F. (2013) Computational model of VEGFR2 pathway to ERK activation and modulation through receptor trafficking, *Cell Signal*, 25, 2496-2510.
- Tannock, G.W. (1977) Characteristics of *Bacteroides* isolates from the cecum of conventional mice, *Appl Environ Microbiol*, 33, 745-750.
- Tansey, J.T., *et al.* (2003) Functional studies on native and mutated forms of perilipins. A role in protein kinase A-mediated lipolysis of triacylglycerols, *J Biol Chem*, 278, 8401-8406.
- Taraboletti, G. and Margosio, B. (2001) Antiangiogenic and antivascular therapy for cancer, *Curr Opin Pharmacol*, 1, 378-384.
- Thangaraju, M., *et al.* (2009) GPR109A is a G-protein-coupled receptor for the bacterial fermentation product butyrate and functions as a tumor suppressor in colon, *Cancer Res*, 69, 2826-2832.
- Thomas, C., *et al.* (2008) Targeting bile-acid signalling for metabolic diseases, *Nat Rev Drug Discov*, 7, 678-693.
- Tims, S., *et al.* (2013) Microbiota conservation and BMI signatures in adult monozygotic twins, *ISMEJ*, 7, 707-717.

- Tousoulis, D., *et al.* (2012) The role of nitric oxide on endothelial function, *Curr Vasc Pharmacol*, 10, 4-18.
- Towle, H.C., Kaytor, E.N. and Shih, H.M. (1997) Regulation of the expression of lipogenic enzyme genes by carbohydrate, *Annu Rev Nutr*, 17, 405-433.
- Trevisan, M. and Dorn, J. (2010) The relationship between periodontal disease (pd) and cardiovascular disease (cvd), *Mediterr J Hematol Infect Dis*, 2, e2010030.
- Trouillon, R., *et al.* (2010) Angiogenin induces nitric oxide synthesis in endothelial cells through PI-3 and Akt kinases, *Biochemistry*, 49, 3282-3288.
- Trygg, J., Holmes, E. and Lundstedt, T. (2007) Chemometrics in metabonomics, *J Proteome Res*, 6, 469-479.
- Trygg, J. and Wold, S. (2002) Orthogonal Projections to Latent Structures *Journal of Chemometrics*, 16, 119-128.
- Tsikas, D. (2005) Methods of quantitative analysis of the nitric oxide metabolites nitrite and nitrate in human biological fluids, *Free Radic Res*, 39, 797-815.
- Tuomisto, H. (2010) *A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity.* John Wiley & Sons.
- Turnbaugh, P.J., *et al.* (2008) Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome, *Cell host & microbe*, 3, 213-223.
- Turnbaugh, P.J., *et al.* (2009) A core gut microbiome in obese and lean twins, *Nature*, 457, 480-484.
- Turnbaugh, P.J., *et al.* (2006) An obesity-associated gut microbiome with increased capacity for energy harvest, *Nature*, 444, 1027-1031.
- Turner, J.R. (2009) Intestinal mucosal barrier function in health and disease, *Nat Rev Immunol*, 9, 799-809.
- Tuteja, N. (2009) Signaling through G protein coupled receptors, *Plant Signal Behav*, 4, 942-947.
- Tyers, M. and Mann, M. (2003) From genomics to proteomics, *Nature*, 422, 193-197.
- Tzoulaki, I., *et al.* (2014) Design and analysis of metabolomics studies in epidemiologic research: a primer on -omic technologies, *Am J Epidemiol*, 180, 129-139.
- Unger, R.H., *et al.* (2010) Lipid homeostasis, lipotoxicity and the metabolic syndrome, *Biochim Biophys Acta*, 1801, 209-214.
- Valbuena, S. and Lerma, J. (2016) Non-canonical Signaling, the Hidden Life of Ligand-Gated Ion Channels, *Neuron*, 92, 316-329.
- Van Belle, E., *et al.* (1997) Hypercholesterolemia attenuates angiogenesis but does not preclude augmentation by angiogenic cytokines, *Circulation*, 96, 2667-2674.
- Van, Q.N., *et al.* (2008) Comparison of 1D and 2D NMR spectroscopy for metabolic profiling, *J Proteome Res*, 7, 630-639.
- van Schaftingen, E. and Gerin, I. (2002) The glucose-6-phosphatase system, *Biochem J*, 362, 513-532.
- Vandeputte, D., *et al.* (2016) Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates, *Gut*, 65, 57-62.
- Vasudev, N.S. and Reynolds, A.R. (2014) Anti-angiogenic therapy for cancer: current progress, unresolved questions and future directions, *Angiogenesis*, 17, 471-494.

- Venkatesh, M., *et al.* (2014) Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and Toll-like receptor 4, *Immunity*, 41, 296-310.
- Vermeulen, P.B., *et al.* (1996) Quantification of angiogenesis in solid human tumours: an international consensus on the methodology and criteria of evaluation, *Eur J Cancer*, 32A, 2474-2484.
- Vuorio, T., Jauhiainen, S. and Yla-Herttuala, S. (2012) Pro- and anti-angiogenic therapy and atherosclerosis with special emphasis on vascular endothelial growth factors, *Expert Opin Biol Ther*, 12, 79-92.
- Wacklin, P., *et al.* (2011) Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine, *PLoS One*, 6, e20113.
- Wahlen, K., Sjolín, E. and Hoffstedt, J. (2008) The common rs9939609 gene variant of the fat mass- and obesity-associated gene FTO is related to fat cell lipolysis, *J Lipid Res*, 49, 607-611.
- Waltenberger, J. (2001) Impaired collateral vessel development in diabetes: potential cellular mechanisms and therapeutic implications, *Cardiovasc Res*, 49, 554-560.
- Wang, F. and Xu, Y. (2014) Body mass index and risk of renal cell cancer: a dose-response meta-analysis of published cohort studies, *Int J Cancer*, 135, 1673-1686.
- Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics, *Nat Rev Genet*, 10, 57-63.
- Warrier, M., *et al.* (2015) The TMAO-Generating Enzyme Flavin Monooxygenase 3 Is a Central Regulator of Cholesterol Balance, *Cell Rep*.
- Weiss, M.M., *et al.* (2008) Evaluation of a series of naphthamides as potent, orally active vascular endothelial growth factor receptor-2 tyrosine kinase inhibitors, *J Med Chem*, 51, 1668-1680.
- Weljie, A.M., *et al.* (2006) Targeted profiling: quantitative analysis of 1H NMR metabolomics data, *Anal Chem*, 78, 4430-4442.
- Wells, J.C.K. and Fewtrell, M.S. (2006) Measuring body composition, *Arch Dis Child*, 91, 612-617.
- Wesoly, R. and Weiler, U. (2012) Nutritional Influences on Skatole Formation and Skatole Metabolism in the Pig, *Animals*, 2, 221-242.
- Whitehead, T.R., *et al.* (2008) Catabolic pathway for the production of skatole and indoleacetic acid by the acetogen *Clostridium drakei*, *Clostridium scatologenes*, and swine manure, *Appl Environ Microbiol*, 74, 1950-1953.
- WHO (2016) Obesity and overweight. Geneva, Switzerland.
- Widiker, S., *et al.* (2010) High-fat diet leads to a decreased methylation of the Mc4r gene in the obese B6MI and the lean B6 mouse lines, *J Appl Genet*, 51, 193-197.
- Wikoff, W.R., *et al.* (2009) Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites, *Proc Natl Acad Sci USA*, 106, 3698-3703.
- Williams, J.K., Armstrong, M.L. and Heistad, D.D. (1988) Vasa vasorum in atherosclerotic coronary arteries: responses to vasoactive stimuli and regression of atherosclerosis, *Circ Res*, 62, 515-523.
- Wishart, D.S., *et al.* (2007) HMDB: the Human Metabolome Database, *Nucleic Acids Res*, 35, D521-526.
- Wu, G.D., *et al.* (2011) Linking long-term dietary patterns with gut microbial enterotypes, *Science*, 334, 105-108.

- Wu, H.-T., Lin, S.-H. and Chen, Y.-H. (2005) Inhibition of cell proliferation and in vitro markers of angiogenesis by indole-3-carbinol, a major indole metabolite present in cruciferous vegetables, *J Agric Food Chem*, 53, 5164-5169.
- Wu, I.W., *et al.* (2011) p-Cresyl sulphate and indoxyl sulphate predict progression of chronic kidney disease, *Nephrol Dial Transplant*, 26, 938-947.
- Xie, L., *et al.* (2008) Adiponectin and leptin are secreted through distinct trafficking pathways in adipocytes, *Biochim Biophys Acta*, 1782, 99-108.
- Xiong, Y., *et al.* (2004) Short-chain fatty acids stimulate leptin production in adipocytes through the G protein-coupled receptor GPR41, *Proc Natl Acad Sci USA*, 101, 1045-1050.
- Xu, J. and Gordon, J.I. (2003) Honor thy symbionts, *Proc Natl Acad Sci USA*, 100, 10452-10459.
- Xu, M., McCanna, D.J. and Sivak, J.G. (2015) Use of the viability reagent PrestoBlue in comparison with alamarBlue and MTT to assess the viability of human corneal epithelial cells, *J Pharmacol Toxicol Methods*, 71, 1-7.
- Xu, Z.-R., Hu, C.-H. and Wang, M.-Q. (2002) Effects of fructooligosaccharide on conversion of L-tryptophan to skatole and indole by mixed populations of pig fecal bacteria, *J Gen Appl Microbiol*, 48, 83-90.
- Yang, W. and Xia, S.-H. (2006) Mechanisms of regulation and function of G-protein-coupled receptor kinases, *World J Gastroenterol*, 12, 7753-7757.
- Yarrow, J.C., *et al.* (2004) A high-throughput cell migration assay using scratch wound healing, a comparison of image-based readout methods, *BMC Biotechnol*, 4, 21.
- Yassour, M., *et al.* (2016) Sub-clinical detection of gut microbial biomarkers of obesity and type 2 diabetes, *Genome Med*, 8, 17.
- Yisireyili, M., *et al.* (2013) Indoxyl sulfate promotes cardiac fibrosis with enhanced oxidative stress in hypertensive rats, *Life Sci*, 92, 1180-1185.
- Zeeb, M., Strilic, B. and Lammert, E. (2010) Resolving cell-cell junctions: lumen formation in blood vessels, *Curr Opin Cell Biol*, 22, 626-632.
- Zhang, Y., *et al.* (1993) Immunohistochemical study of intimal microvessels in coronary atherosclerosis, *Am J Pathol*, 143, 164-172.
- Zhang, C., *et al.* (2015) Dietary modulation of gut microbiota contributes to alleviation of both genetic and simple obesity in children. *EBioMedicine*, 2(8), 968-84.
- Zhang, Y., *et al.* (1994) Positional cloning of the mouse obese gene and its human homologue, *Nature*, 372, 425-432.
- Zhao, Y. (2012) Auxin biosynthesis: a simple two-step pathway converts tryptophan to indole-3-acetic acid in plants, *Mol Plant*, 5, 334-338.
- Zhu, W., *et al.* (2016) Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk, *Cell*, 165, 111-124.
- Ziche, M., *et al.* (1994) Nitric oxide mediates angiogenesis in vivo and endothelial cell growth and migration in vitro promoted by substance P, *J Clin Invest*, 94, 2036-2044.







# Appendix A

## Published papers





Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

Current Opinion in  
Pharmacology



## The microbiome and its pharmacological targets: therapeutic avenues in cardiometabolic diseases

Ana Luisa Neves<sup>1</sup>, Julien Chilloux<sup>1</sup>, Magali H Sarafian,  
Mohd Badrin Abdul Rahim, Claire L Boulange<sup>2</sup> and  
Marc-Emmanuel Dumas

Consisting of trillions of non-pathogenic bacteria living in a symbiotic relationship with their mammalian host, the gut microbiota has emerged in the past decades as one of the key drivers for cardiometabolic diseases (CMD). By degrading dietary substrates, the gut microbiota produces several metabolites that bind human pharmacological targets, impact subsequent signalling networks and *in fine* modulate host's metabolism. In this review, we revisit the pharmacological relevance of four classes of gut microbial metabolites in CMD: short-chain fatty acids (SCFA), bile acids, methylamines and indoles. Unravelling the signalling mechanisms of the microbial-mammalian metabolic axis adds one more layer of complexity to the physiopathology of CMD and opens new avenues for the development of microbiota-based pharmacological therapies.

### Address

Division of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, Exhibition Road, London SW7 2AZ, UK

Corresponding author: Dumas, Marc-Emmanuel  
([m.dumas@imperial.ac.uk](mailto:m.dumas@imperial.ac.uk))

<sup>1</sup> These authors contributed equally to this review.

<sup>2</sup> Current address: Metabomatrix Ltd, Bio-Incubator, Prince Consort Road, South Kensington, London SW7 2BP, UK.

### Chemical compounds studied in this article

Acetate (PubChem CID: 176) Butyrate (PubChem CID: 264) Cholic acid (PubChem CID: 221493) Chenodeoxycholic acid (PubChem CID: 10133) Deoxycholic acid (PubChem CID: 222528) Indole-3-propionate (PubChem CID: 3744) 3-Indoxylsulfate (PubChem CID: 10258) Propionate (PubChem CID: 1032) Trimethylamine (PubChem CID: 1146) Trimethylamine-N-oxide (PubChem CID: 1145)

Current Opinion in Pharmacology 2015, 25:36-44

This review comes from a themed issue on **Endocrine and metabolic diseases**

Edited by Kevin G Murphy

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 28th October 2015

<http://dx.doi.org/10.1016/j.coph.2015.09.013>

1471-4892/© 2015 Elsevier Ltd. All rights reserved.

### Introduction

Cardiometabolic diseases (CMD) present a complex array of interrelated risk factors affecting more than one billion

people with a dramatic impact on mortality, morbidity and quality of life [1]. These factors (including impaired glucose tolerance, dyslipidemia, arterial hypertension, insulin resistance and central obesity) are epidemiologically clustered — the presence of at least three of five of these symptoms corresponding to the 'metabolic syndrome' clinical diagnosis [1]. Although many pharmacological mechanisms have been suggested, the underlying causes of CMD and its potential therapeutic avenues remain to be fully explored. With the advent of high-throughput methodologies (metagenomics, metabolomics), the gut microbiome emerged as one of the key drivers for CMD [2]. The gut ecosystem, as well as its individual members, was shown to contribute to the host metabolism [3<sup>\*</sup>]. A lower bacterial gene count (LGC) is associated to increased adiposity, insulin resistance and dyslipidemia [4<sup>\*\*</sup>] and dietary intervention can improve both bacterial gene richness and clinical metabolic outcomes [5<sup>\*\*</sup>]. Patients with type 2 diabetes (T2D) also show specific compositional and functional changes in their metagenomes [6<sup>\*\*</sup>].

With the increasing number of clinical studies reporting associations between the composition of the gut microbiota and CMD outcomes, one question arises — how are these changes in microbial ecology translated into pharmacological messages to the mammalian host? Consisting of trillions of non-pathogenic bacteria living in a symbiotic relationship with their host, gut microbiota produces several signalling molecules (e.g., LPS, peptidoglycans, but also metabolites) that bind host proteins and impact signalling networks, therefore playing a central role as chemical messengers in the microbial-mammalian crosstalk [7]. The identification of the pharmacological targets and signalling pathways of these metabolites is key to a better understanding the molecular crosstalk supporting the microbial-mammalian metabolic axis — and provides a suitable framework for the discovery of the mechanistic basis of these associations. In this context, fine mapping of the microbial signalling metabolome and its host molecular targets opens up novel pharmacological avenues for microbiome interventions.

### The microbiome interacts with its host through microbial metabolites

In this review, we shall present four classes of gut microbial metabolites impacting host molecular mechanisms

Boulangé et al. *Genome Medicine* (2016) 8:42  
DOI 10.1186/s13073-016-0303-2

Genome Medicine

REVIEW

Open Access



## Impact of the gut microbiota on inflammation, obesity, and metabolic disease

Claire L. Boulangé<sup>1</sup>, Ana Luisa Neves<sup>2</sup>, Julien Chilloux<sup>2</sup>, Jeremy K. Nicholson<sup>1,2\*</sup> and Marc-Emmanuel Dumas<sup>2\*</sup>

### Abstract

The human gut harbors more than 100 trillion microbial cells, which have an essential role in human metabolic regulation via their symbiotic interactions with the host. Altered gut microbial ecosystems have been associated with increased metabolic and immune disorders in animals and humans. Molecular interactions linking the gut microbiota with host energy metabolism, lipid accumulation, and immunity have also been identified. However, the exact mechanisms that link specific variations in the composition of the gut microbiota with the development of obesity and metabolic diseases in humans remain obscure owing to the complex etiology of these pathologies. In this review, we discuss current knowledge about the mechanistic interactions between the gut microbiota, host energy metabolism, and the host immune system in the context of obesity and metabolic disease, with a focus on the importance of the axis that links gut microbes and host metabolic inflammation. Finally, we discuss therapeutic approaches aimed at reshaping the gut microbial ecosystem to regulate obesity and related pathologies, as well as the challenges that remain in this area.

**The essential role of the gut microbiota in human health**

Trillions of microbes live in our guts, which are collectively termed "gut microbiota" [1]. The process of colonization with these microbes starts prenatally, through microbial transmission from mother to fetus [2]. Colonization of the human gut continues after birth and is modulated by factors including gestational age, mode of delivery (natural or by Caesarean section), diet (breastfeeding or infant formula), hygiene, and antibiotic exposure. The environment and diet during the first 3 years of life are crucial to the acquisition of an adult-like microbiota and to the establishment of bacterial-host symbiosis that influences the development of the immune and neurologic systems. The human gut microbiota reaches the characteristics of an adult microbiota between the ages of 2 and 5 years [2].

Gene sequencing data have shown that although a great diversity of bacterial species is found among healthy individuals, the gut metagenome (that is, all the genes in the community of gut microorganisms) is involved in core functions, such as the digestion and degradation of otherwise indigestible nutrients, and the development and stimulation of the immune system and digestive tract of the host [3–7]. The gut microbiota also produces pharmacologically active signaling molecules that interact with the metabolism of the host [8–10]. For example, short-chain fatty acids (SCFAs) are produced by fermentation of dietary fibers by gut bacteria. Their interaction with G protein-coupled receptors (GPCRs) affects insulin sensitivity in adipocytes and peripheral organs, thus regulating energy metabolism [11]. Transient changes in the intestinal ecosystem occur throughout life and in some cases can result in the disruption of microbial-host symbiosis [12]. Owing to the essential role of the gut ecosystem in maintaining host physiology, its alteration can trigger a wide range of physiological disorders, including low-grade inflammation, metabolic disorders, excess lipid accumulation, and loss of insulin sensitivity, which increase the risk of developing metabolic diseases.

\* Correspondence: j.nicholson@imperial.ac.uk, m.dumas@imperial.ac.uk  
<sup>1</sup>Metabomatrix Ltd, Bio-incubator, Prince Consort Road, South Kensington, London SW7 2BP, UK  
<sup>2</sup>Division of Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, Exhibition Road, South Kensington, London SW7 2PH, UK



© 2016 Boulangé et al. Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Europe PMC Funders Group

Author Manuscript

*Curr Opin Clin Nutr Metab Care*. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

*Curr Opin Clin Nutr Metab Care*. 2016 July ; 19(4): 250–256. doi:10.1097/MCO.0000000000000284.

## The microbial-mammalian metabolic axis, a critical symbiotic relationship

Julien Chilloux<sup>#</sup>, Ana Luisa Neves<sup>#</sup>, Claire L. Boulangé<sup>2</sup>, and Marc-Emmanuel Dumas<sup>#</sup>

Division of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom

<sup>#</sup> These authors contributed equally to this work.

### Abstract

**Purpose of review—**The microbial-mammalian symbiosis plays a critical role in metabolic health. Microbial metabolites emerge as key messengers in the complex communication between the gut microbiota and their host. These chemical signals are mainly derived from nutritional precursors, which also are in turn able to modify gut microbiota population. Recent advances in the characterization of the gut microbiome and the mechanisms involved in this symbiosis allow the development of nutritional interventions. This review covers the latest findings on the microbial-mammalian metabolic axis as a critical symbiotic relationship particularly relevant to clinical nutrition.

**Recent findings—**The modulation of host metabolism by metabolites derived from the gut microbiota highlights the importance of gut microbiota in disease prevention and causation. The composition of microbial populations in our gut ecosystem is a critical pathophysiological factor, mainly regulated by diet, but also by the host's characteristics (e.g. genetics, circadian clock, immune system, age). Tailored interventions, including dietary changes, the use of antibiotics, prebiotic and probiotic supplementation and faecal transplantation are promising strategies to manipulate microbial ecology.

**Summary—**The microbiota is now considered as an easily reachable target to prevent and treat related diseases. Recent findings in both mechanisms of its interactions with host metabolism and in strategies to modify gut microbiota will allow us to develop more effective treatments especially in metabolic diseases.

### Keywords

Microbiota; signalling metabolites; host metabolism; dietary intervention

### Introduction

Humans have evolved as part of a critical symbiotic relationship with their gut microbes. The gut ecosystem harbours thousands of microbial species and millions of genes,

Corresponding author Dr Marc-Emmanuel Dumas +44 20 7594 1820 m.dumas@imperial.ac.uk.

<sup>2</sup>Current address: Metabometrix Ltd, Bio-incubator, Prince Consort Road, South Kensington, London SW7 2BP, United Kingdom.

Conflicts of Interest: none.

Rev Port Cardiol. 2014;33(4):223–228



Revista Portuguesa de  
**Cardiologia**  
Portuguese Journal of Cardiology  
[www.revportcardiol.org](http://www.revportcardiol.org)



ORIGINAL ARTICLE

## Cardiovascular risk in overweight/obese and lean hypertensive patients



Ana Luísa Neves<sup>a,b</sup>, Luciana Couto<sup>b</sup>

<sup>a</sup> GABBA Program, Abel Salazar Biomedical Sciences Institute, University of Porto, Portugal

<sup>b</sup> Department of Social Sciences and Health, Family Medicine Unit, University of Porto, Portugal

Received 5 September 2011; accepted 19 October 2013

Available online 13 May 2014

### KEYWORDS

Cardiovascular risk;  
Obesity;  
Overweight;  
Arterial hypertension

### Abstract

**Introduction:** Obesity and hypertension have been identified as independent risk factors for cardiovascular disease. Nevertheless, the role of obesity in the development and progression of target-organ disease in hypertensive patients is controversial. The objective of this study was to assess the impact of body weight on cardiovascular risk factors, target-organ disease and global cardiovascular risk in hypertensive patients in a primary care setting.

**Methods:** A cross-sectional observational study was carried in Vila Nova de Gaia, Portugal (n=150). A detailed medical and personal history was obtained and a physical examination was performed. Venous blood and 24-hour urine samples were collected, and an electrocardiogram was performed. Cardiovascular risk was assessed using the Framingham score. The statistical analysis was performed using SPSS®. A p-value <0.05 was considered statistically significant.

**Results:** The sample was 71.8% female, with a mean age of 74.3±10.8 years. The prevalence of obesity was 29.5%. Overweight/obese subjects presented lower mean HDL cholesterol (51.2±13.9 mg/dl vs. 65.4±35.2, p<0.005), higher triglycerides (137.8±70.4 mg/dl vs. 111.5±68.8 mg/dl, p<0.001), higher fasting glucose (111.9±32.8 mg/dl vs. 98.4±13.1 mg/dl, p<0.011) and more frequent mild valve disease (57.9% vs. 29.6%, p=0.021). Global cardiovascular risk was also significantly higher (10.9±7.7 vs. 6.5±5.7, p<0.001).

**Conclusion:** Overweight and obesity appear to be related to a less favorable lipid and blood glucose profile and higher cardiovascular risk in hypertensive patients. On the basis of our findings we suggest strict metabolic monitoring and improved education on weight reduction and control at primary health care clinics.

© 2011 Sociedade Portuguesa de Cardiologia. Published by Elsevier España, S.L. All rights reserved.

Corresponding author.

E-mail address: [ana.luiza.neves@gmail.com](mailto:ana.luiza.neves@gmail.com) (A.L. Neves).

<http://dx.doi.org/10.1016/j.repo.2013.10.016>

0870-2551/© 2011 Sociedade Portuguesa de Cardiologia. Published by Elsevier España, S.L. All rights reserved.

Laranjo L, et al. J Am Med Inform Assoc 2015;22:243–256. doi:10.1136/ami-2014-002841, Review

## The influence of social networking sites on health behavior change: a systematic review and meta-analysis

RECEIVED 27 March 2014  
REVISED 4 June 2014  
ACCEPTED 24 June 2014  
PUBLISHED ONLINE FIRST 8 July 2014



Liliana Laranjo<sup>1</sup>, Amañi Arguel<sup>2</sup>, Ana L Neves<sup>3</sup>, Aldeen M Gallagher<sup>4</sup>, Ruth Kaplan<sup>5</sup>, Nathan Mortimer<sup>4</sup>, Guilherme A Mendes<sup>6</sup>, Annie Y S Lau<sup>2</sup>

### ABSTRACT

**Objective** Our aim was to evaluate the use and effectiveness of interventions using social networking sites (SNSs) to change health behaviors.

**Materials and methods** Five databases were scanned using a predefined search strategy. Studies were included if they focused on patients/consumers, involved an SNS intervention, had an outcome related to health behavior change, and were prospective. Studies were screened by independent investigators, and assessed using Cochrane's 'risk of bias' tool. Randomized controlled trials were pooled in a meta-analysis.

**Results** The database search retrieved 4656 citations; 12 studies (7411 participants) met the inclusion criteria. Facebook was the most utilized SNS, followed by health-specific SNSs, and Twitter. Eight randomized controlled trials were combined in a meta-analysis. A positive effect of SNS interventions on health behavior outcomes was found (Hedges'  $g$  0.24; 95% CI 0.04 to 0.43). There was considerable heterogeneity ( $I^2 = 84.0\%$ ;  $\tau^2 = 0.058$ ) and no evidence of publication bias.

**Discussion** To the best of our knowledge, this is the first meta-analysis evaluating the effectiveness of SNS interventions in changing health-related behaviors. Most studies evaluated multi-component interventions, posing problems in isolating the specific effect of the SNS. Health behavior change theories were seldom mentioned in the included articles, but two particularly innovative studies used 'network alteration', showing a positive effect. Overall, SNS interventions appeared to be effective in promoting changes in health-related behaviors, and further research regarding the application of these promising tools is warranted.

**Conclusions** Our study showed a positive effect of SNS interventions on health behavior-related outcomes, but there was considerable heterogeneity.

**Protocol registration** The protocol for this systematic review is registered at <http://www.crd.york.ac.uk/PROSPERO> with the number CRD42013004140.

**Key words:** Social Networking Site, Social Network, Social Media, Consumer Health, Behavior Change

### BACKGROUND AND SIGNIFICANCE

Social networking sites (SNSs) have become a global phenomenon. They are generally defined as web-based platforms that allow individuals to create their own personal profile and build a network of connections with other users.<sup>1</sup> As of September 2013, 73% of online adults were using an SNS of some kind and 42% were using more than one.<sup>2,3</sup> Facebook is the most popular platform (with more than 1.19 billion monthly active users<sup>4</sup>), followed by Twitter (500 million users worldwide<sup>5</sup>).

In parallel to general purpose SNSs like Facebook and Twitter, health-specific SNSs are also emerging.<sup>6</sup> Some are oriented towards patients with a specific chronic condition

(eg, TuDiabetes), others are more general and open to patients with any chronic condition (eg, PatientsLikeMe), and a few others target people wanting to change a particular health-risk behavior (eg, smoking cessation<sup>7</sup>) or other health-related lifestyle factors.

The application of SNSs in the health domain shows tremendous potential.<sup>8</sup> At the population level, they are currently being used for public health surveillance,<sup>9</sup> both for communicable<sup>9,10</sup> and non-communicable diseases.<sup>11,12</sup> At the individual level, they are able to facilitate access to health-related information<sup>13–16</sup> and social support,<sup>7,17</sup> promoting better-informed treatment decisions.<sup>18,19</sup> Given that lifestyle behaviors are

Correspondence to Dr Liliana Laranjo, Portuguese School of Public Health, Universidade Nova de Lisboa, Avenida Padre Cruz, Lisboa 1600-560, Portugal; [lc051@mail.harvard.edu](mailto:lc051@mail.harvard.edu)

© The Author 2014. Published by Oxford University Press on behalf of the American Medical Informatics Association.

All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

For numbered affiliations see end of article.



## Healthcare hashtag index development: Identifying global impact in social media



Luís Pinho-Costa MD <sup>a, \*</sup>, Kenneth Yakubu FWACP FMCfM MB.BS <sup>b</sup>, Kyle Hoedebecke MD CKTP RMT <sup>c</sup>,  
Liliana Laranjo MD MPH PhD <sup>d</sup>, Christofer Patrick Reichel MD <sup>e</sup>, Maria del C. Colon-Gonzalez MD <sup>f</sup>,  
Ana Luísa Neves MD MSc <sup>g</sup>, Hassna Errami MD <sup>h</sup>

<sup>a</sup> Fânzeres Family Health Unit, Praceta da Barrosa, s/n, 4510-513 Fânzeres, Gondomar, Portugal

<sup>b</sup> Department of Family Medicine, University of Jos & Jos University Teaching Hospital, P.M.B. 2076, Jos, Plateau State, Nigeria

<sup>c</sup> Robinson Health Clinic, Team 11722 Tagaytay Drive Ft. Bragg, NC 28310, USA

<sup>d</sup> Australian Institute of Health Innovation, Centre for Health Informatics, Level 6, 75 Taberna Road, Macquarie University, NSW 2109, Australia

<sup>e</sup> Austrian Association of General Practice and Family Medicine, Grenzasse 11, 3100 St. Pölten, Austria

<sup>f</sup> Department of Family & Preventive Medicine, University of Texas, Rio Grande Valley 205 E. Toronto Ave, McAllen, TX 78503, USA

<sup>g</sup> Unit of Family Medicine, Department of Social Sciences and Health, Faculty of Medicine, University of Porto, Portugal

<sup>h</sup> Cabinet médical des Aygalades, 57 chemin de Saint Antoine à Saint Joseph, 13015 Marseille, France

### article info

#### Article history:

Received 15 February 2016

Revised 12 September 2016

Accepted 14 September 2016

Available online 16 September 2016

#### Keywords:

Biomedinformatics

Internet

Social media

Family practice

Primary health care

### abstract

**Purpose:** Create an index of global reach for healthcare hashtags and tweeters therein, filterable by topic of interest.

**Materials and methods:** For this proof-of-concept study we focused on the field of Primary Care and Family Medicine. Six hashtags were selected based on their importance, from the ones included in the 'Healthcare Hashtag Project'. Hashtag Global Reach (HGR) was calculated using the additive aggregation of five weighted, normalized indicator variables: number of impressions, tweets, tweeters, user locations, and user languages. Data were obtained for the last quarter of 2014 and first quarter of 2015 using Symplur Signals. Topic-specific HGR were calculated for the top 10 terms and for sets of quotes mapped after a thematic analysis. Individual Global Reach, IGR, was calculated across hashtags as additive indexes of three indicators: replies, retweets and mentions.

**Results:** Using the HGR score we were able to rank six selected hashtags and observe their performance throughout the study period. We found that #PrimaryCare and #FMRrevolution had the highest HGR score in both quarters; Interestingly, #FMChangeMakers experienced a marked increase in its global visibility during the study period. 'Health Policy' was the commonest theme, while 'Care', 'Family' and 'Health' were the most common terms.

**Discussion:** This is the first study describing an altmetric hashtag index. Assuming analytical soundness, the index might prove generalizable to other healthcare hashtags. If released as a real-time business intelligence tool with customizable settings, it could aid publishing and strategic decisions by nations, organizations, and analysts. IGR could also serve to augment academic evaluation and professional development.

**Conclusion:** Our study demonstrates the feasibility of using an index on the global reach of healthcare hashtags and tweeters.

© 2016 Elsevier Inc. All rights reserved.

### 1. Introduction

Social media use has risen exponentially each year on a global scale [1–4]. As a group of internet-based applications, they allow

for the exchange of user-generated content, building on the concept of Web 2.0 [1,5]. These applications include blogs, discussion boards, wikis, and social networking sites [1]. In the medical domain, social media are increasingly used by clinicians and researchers as an efficient way of sharing information, keeping up-to-date with scientific knowledge and collaborating with both peers and patients [1,6–14].

Corresponding author at: Unidade de Saúde Familiar de Fânzeres, Praceta da Barrosa, s/n, 4510-513 Fânzeres, Gondomar, Portugal.  
E-mail address: [luisdepinhocosta@gmail.com](mailto:luisdepinhocosta@gmail.com) (L. Pinho-Costa).

<http://dx.doi.org/10.1016/j.jbi.2016.09.010>

1532-0464/ © 2016 Elsevier Inc. All rights reserved.



## Systems biology

# MetaboSignal: a network-based approach for topological analysis of metabotype regulation via metabolic and signaling pathways

Andrea Rodriguez-Martinez<sup>1</sup>, Rafael Ayala<sup>1</sup>, Joram M. Posma<sup>1</sup>,  
Ana L. Neves<sup>1</sup>, Dominique Gauguier<sup>1,2</sup>, Jeremy K. Nicholson<sup>1</sup> and  
Marc-Emmanuel Dumas<sup>1,\*</sup>

<sup>1</sup>Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ London, UK and Sorbonne Universities, University Pierre & Marie Curie, University Paris Descartes, Sorbonne Paris Cité INSERMUMR\_S 1138, Cordeliers Research Centre, 75006 Paris, France

\*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on August 1, 2016; revised on October 4, 2016; editorial decision on October 20, 2016; accepted on November 3, 2016

## Abstract

**Summary:** MetaboSignal is an R package that allows merging metabolic and signaling pathways reported in the Kyoto Encyclopaedia of Genes and Genomes (KEGG). It is a network-based approach designed to navigate through topological relationships between genes (signaling- or metabolic-genes) and metabolites, representing a powerful tool to investigate the genetic landscape of metabolic phenotypes.

**Availability and Implementation:** MetaboSignal is available from Bioconductor: <https://bioconductor.org/packages/MetaboSignal/>

**Contact:** m.dumas@imperial.ac.uk

**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics* online.

## 1 Introduction

'Omics' approaches such as genomics, metabolomics and metatranscriptomics are used to improve our understanding of integrated functioning of living organisms, with each 'omics' generating high-density datasets (Dumas, 2012; Nicholson et al., 2002). Genes and metabolites provide complementary information about biological processes. Metabolic reactions are catalyzed by enzymes encoded by genes. The activity of these enzymes is regulated by transcriptional, translational and post-translational mechanisms impacting metabolism. Metabolic processes are also regulated by signaling transduction pathways, allowing the organism to adapt to environmental changes and maintain homeostasis.

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) is a popular reference database storing biological pathways and cellular processes from a wide range of reference organisms as graphical diagrams, where the nodes are biological entities (e.g. genes, metabolites) and the edges represent the

relationship between them. In the recent years, several tools have been developed for analyzing, editing and customizing KEGG metabolic (Cottret et al., 2010; Posma et al., 2014) or signaling (Zhang and Wiemann, 2009) pathways independently. However, these tools ignore the interconnection between the genome and metabolome, and therefore they do not provide an integrated overview of regulatory events leading to the observed metabolic patterns.

We introduced an integrated metabolome and interactome mapping (iMIM) strategy to analyze the topology of protein interaction, signaling and metabolic networks, and implemented it as a standalone database and workflow (Davidovic et al., 2011). To make this network-modeling strategy accessible to a wider community, we have now developed an R-based approach, called MetaboSignal, which allows building organism-specific KEGG networks that account for the interaction between metabolites and genes (both metabolic and signaling). Advantages of MetaboSignal over other network-based approaches include: (i) network directionality; (ii) network



# Appendix B

## Reproduction permissions



**Elsevier Science and Technology Journals LICENSE  
TERMS AND CONDITIONS**

Aug 01, 2017

---

This is a License Agreement between Ana Soares ("You") and Elsevier Science and Technology Journals ("Elsevier Science and Technology Journals") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier Science and Technology Journals, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

|                                                     |                                                                                                                                                                                                     |
|-----------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| License Number                                      | 4160271381490                                                                                                                                                                                       |
| License date                                        | Aug 01, 2017                                                                                                                                                                                        |
| Licensed content publisher                          | Elsevier Science and Technology Journals                                                                                                                                                            |
| Licensed content title                              | Current opinion in pharmacology                                                                                                                                                                     |
| Licensed content date                               | Jan 1, 2001                                                                                                                                                                                         |
| Type of Use                                         | Thesis/Dissertation                                                                                                                                                                                 |
| Requestor type                                      | Author of requested content                                                                                                                                                                         |
| Format                                              | Print, Electronic                                                                                                                                                                                   |
| Portion                                             | chapter/article                                                                                                                                                                                     |
| Number of pages in chapter/article                  | 9                                                                                                                                                                                                   |
| Title or numeric reference of the portion(s)        | Entire paper                                                                                                                                                                                        |
| Title of the article or chapter the portion is from | The microbiome and its pharmacological targets: therapeutic avenues in cardiometabolic diseases. Curr Opin Pharmacol. 2015 Dec;25:36-44. doi: 10.1016/j.coph.2015.09.013. Epub 2015 Oct 28. Review. |
| Editor of portion(s)                                | Curr Opin Pharmacology                                                                                                                                                                              |
| Author of portion(s)                                | Neves AL, Chilloux J, Sarafian MH, Rahim MB, Boulangé CL, Dumas ME.                                                                                                                                 |
| Volume of serial or monograph.                      | 25                                                                                                                                                                                                  |
| Page range of the portion                           | 36-44                                                                                                                                                                                               |
| Publication date of portion                         | 2015                                                                                                                                                                                                |
| Rights for                                          | Main product                                                                                                                                                                                        |
| Duration of use                                     | Life of current edition                                                                                                                                                                             |
| Creation of copies for the disabled                 | no                                                                                                                                                                                                  |
| With minor editing privileges                       | no                                                                                                                                                                                                  |
| For distribution to                                 | Worldwide                                                                                                                                                                                           |
| In the following language(s)                        | Original language of publication                                                                                                                                                                    |
| With incidental promotional use                     | no                                                                                                                                                                                                  |

|                                           |                                                                                                                          |
|-------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| The lifetime unit quantity of new product | Up to 499                                                                                                                |
| Made available in the following markets   | education                                                                                                                |
| The requesting person/organization is:    | Ana Luisa Neves                                                                                                          |
| Order reference number                    |                                                                                                                          |
| Author/Editor                             | Ana Luisa Neves                                                                                                          |
| The standard identifier of New Work       | n/a                                                                                                                      |
| Title of New Work                         | Indoles as microbial pharmaco-metabolites: targets, mechanisms and consequences for obesity and cardiometabolic diseases |
| Publisher of New Work                     | Imperial College London                                                                                                  |
| Expected publication date                 | Dec 2017                                                                                                                 |
| Estimated size (pages)                    | 150                                                                                                                      |
| Total (may include CCC user fee)          | 0.00 USD                                                                                                                 |

Dear Dr. Neves,

Thank you for contacting BioMed Central.

The open access articles published in BioMed Central's journals are made available under the Creative Commons Attribution (CC-BY) license, which means they are accessible online without any restrictions and can be re-used in any way, subject only to proper attribution (which, in an academic context, usually means citation).

The re-use rights enshrined in our license agreement (<http://www.biomedcentral.com/about/policies/license-agreement>) include the right for anyone to produce printed copies themselves, without formal permission or payment of permission fees. As a courtesy, however, anyone wishing to reproduce large quantities of an open access article (250+) should inform the copyright holder and we suggest a contribution in support of open access publication (see suggested contributions at <http://www.biomedcentral.com/about/policies/reprints-and-permissions/suggested-contributions>).

Please note that the following journals have published a small number of articles that, while freely accessible, are not open access as outlined above: Alzheimer's Research & Therapy, Arthritis Research & Therapy, Breast Cancer Research, Critical Care, Genome Biology, Genome Medicine, Stem Cell Research & Therapy.

You will be able to find details about these articles at <http://www.biomedcentral.com/about/policies/reprints-and-permissions>

If you have any questions, please do not hesitate to contact me.

With kind regards,

---

**Ricardo Sison Jr.**

Global Open Research Support Executive  
Global Open Research Support

**Springer Nature**

T +44 (0)203 192 2009

**Wolters Kluwer Health LICENSE  
TERMS AND CONDITIONS**

Nov 20, 2017

This is a License Agreement between Ana Soares ("You") and Wolters Kluwer Health ("Wolters Kluwer Health") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Wolters Kluwer Health, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

|                                                     |                                                                            |
|-----------------------------------------------------|----------------------------------------------------------------------------|
| License Number                                      | 4160890178878                                                              |
| License date                                        | Aug 01, 2017                                                               |
| Licensed content publisher                          | Wolters Kluwer Health                                                      |
| Licensed content title                              | Current opinion in clinical nutrition                                      |
| Licensed content date                               | Jan 1, 2000                                                                |
| Type of Use                                         | Thesis/Dissertation                                                        |
| Requestor type                                      | Author of requested content                                                |
| Format                                              | Print, Electronic                                                          |
| Portion                                             | page                                                                       |
| Number of pages requested                           | 1                                                                          |
| The requesting person/organization is:              | Ana Luisa Neves                                                            |
| Title or numeric reference of the portion (s)       | First page                                                                 |
| Title of the article or chapter the portion is from | The microbial-mammalian metabolic axis: a critical symbiotic relationship. |
| Editor of portion(s)                                | Curr Opin Clin Nutr Metab Care.                                            |
| Author of portion(s)                                | Chilloux J, Neves AL, Boulangé CL, Dumas ME.                               |
| Volume of serial or monograph.                      | 19                                                                         |
| Issue, if republishing an article from a serial     | 4                                                                          |
| Page range of the portion                           | 250                                                                        |
| Publication date of portion                         | 2016                                                                       |
| Rights for                                          | Main product                                                               |
| Duration of use                                     | Life of current edition                                                    |
| Creation of copies for the disabled                 | no                                                                         |
| With minor editing privileges                       | no                                                                         |
| For distribution to                                 | Worldwide                                                                  |
| In the following language(s)                        | Original language of publication                                           |
| With incidental promotional use                     | no                                                                         |
| The lifetime unit quantity of new product           | Up to 499                                                                  |
| Billing Type                                        | Invoice                                                                    |
| Billing Address                                     | Ana Soares<br>SW7 2AZ<br><br>London, United Kingdom<br>Attn: Ana Soares    |

**From:** Revista Portuguesa de Cardiologia <repc@elsevier.com>  
**Sent:** Thursday, August 10, 2017 12:56 PM  
**To:** Neves, Ana Luísa  
**Subject:** Re: Request to reproduce own published paper [170801-014410]

Cara Dra. Ana Luísa,

Ao tratar-se da sua tese de doutoramento, não há qualquer inconveniente em utilizar o seu artigo já publicado na Revista Portuguesa de Cardiologia.

Com os melhores cumprimentos,  
**Ana Ramos Ribeiro**  
Journal Manager  
ELSEVIER Global Journals Production EMEA/LA  
Tel +34 610 54 51 09  
ELSEVIER

**Oxford University Press - Journals LICENSE  
TERMS AND CONDITIONS**

Aug 01, 2017

---

This is a License Agreement between Ana Soares ("You") and Oxford University Press - Journals ("Oxford University Press - Journals") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press - Journals, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

|                                                     |                                                                                                            |
|-----------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| License Number                                      | 4160280298153                                                                                              |
| License date                                        | Aug 01, 2017                                                                                               |
| Licensed content publisher                          | Oxford University Press - Journals                                                                         |
| Licensed content title                              | JOURNAL- AMERICAN MEDICAL INFORMATICS ASSOCIATION                                                          |
| Licensed content date                               | Jan 1, 1994                                                                                                |
| Type of Use                                         | Thesis/Dissertation                                                                                        |
| Requestor type                                      | Author of requested content                                                                                |
| Format                                              | Print, Electronic                                                                                          |
| Portion                                             | page                                                                                                       |
| Number of pages requested                           | 1                                                                                                          |
| Title or numeric reference of the portion(s)        | First page                                                                                                 |
| Title of the article or chapter the portion is from | The influence of social networking sites on health behavior change: a systematic review and meta-analysis. |
| Editor of portion(s)                                | J Am Med Inform Assoc                                                                                      |
| Author of portion(s)                                | Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, Mendes GA, Lau AY.                      |
| Volume of serial or monograph.                      | 22                                                                                                         |
| Issue, if republishing an article from a serial     | 1                                                                                                          |
| Page range of the portion                           | 243                                                                                                        |
| Publication date of portion                         | 2015                                                                                                       |
| Rights for                                          | Main product                                                                                               |
| Duration of use                                     | Life of current edition                                                                                    |
| Creation of copies for the disabled                 | no                                                                                                         |
| With minor editing privileges                       | no                                                                                                         |
| For distribution to                                 | Worldwide                                                                                                  |
| In the following language(s)                        | Original language of publication                                                                           |
| With incidental promotional use                     | no                                                                                                         |
| The lifetime unit quantity of new product           | Up to 499                                                                                                  |

## Elsevier Science and Technology Journals LICENSE TERMS AND CONDITIONS

Aug 01, 2017

---

This is a License Agreement between Ana Soares ("You") and Elsevier Science and Technology Journals ("Elsevier Science and Technology Journals") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier Science and Technology Journals, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

|                                                     |                                                                                                      |
|-----------------------------------------------------|------------------------------------------------------------------------------------------------------|
| License Number                                      | 4160270326766                                                                                        |
| License date                                        | Aug 01, 2017                                                                                         |
| Licensed content publisher                          | Elsevier Science and Technology Journals                                                             |
| Licensed content title                              | Journal of biomedical informatics                                                                    |
| Licensed content date                               | Sep 1, 2001                                                                                          |
| Type of Use                                         | Thesis/Dissertation                                                                                  |
| Requestor type                                      | Author of requested content                                                                          |
| Format                                              | Print, Electronic                                                                                    |
| Portion                                             | page                                                                                                 |
| Number of pages requested                           | 1                                                                                                    |
| Title or numeric reference of the portion(s)        | First page                                                                                           |
| Title of the article or chapter the portion is from | Healthcare hashtag index development: Identifying global impact in social media.                     |
| Editor of portion(s)                                | Journal of Biomedical Informatics                                                                    |
| Author of portion(s)                                | Pinho-Costa L, Yakubu K, Hoedebecke K, Laranjo L, Reichel CP, Colon-Gonzalez MD, Neves AL, Errami H. |
| Volume of serial or monograph.                      | 63                                                                                                   |
| Page range of the portion                           | 390                                                                                                  |
| Publication date of portion                         | 2016                                                                                                 |
| Rights for                                          | Main product                                                                                         |
| Duration of use                                     | Life of current edition                                                                              |
| Creation of copies for the disabled                 | no                                                                                                   |
| With minor editing privileges                       | no                                                                                                   |
| For distribution to                                 | Worldwide                                                                                            |
| In the following language(s)                        | Original language of publication                                                                     |
| With incidental promotional use                     | no                                                                                                   |
| The lifetime unit quantity of new product           | Up to 499                                                                                            |

|                                         |                                           |
|-----------------------------------------|-------------------------------------------|
| Made available in the following markets | education                                 |
| The requesting person/organization is:  | Ana Luisa Neves                           |
| Order reference number                  |                                           |
| Author/Editor                           | Ana Luisa Neves / Imperial College London |
| The standard identifier of New Work     | n/a                                       |
| Title of New Work                       | n/a                                       |
| Publisher of New Work                   | n/a                                       |
| Expected publication date               | Dec 2017                                  |
| Estimated size (pages)                  | 150                                       |
| Total (may include CCC user fee)        | 0.00 USD                                  |

Dear Ana Luisa Neves,

**RE.** Andrea Rodriguez-Martinez et al. MetaboSignal: a network-based approach for topological analysis of metabolite regulation via metabolic and signaling pathways. *Bioinformatics* (2017) 33 (5): 773-775

My apologies for the delay in responding to you and thank you for your email requesting permission to reuse all or part of your article in a thesis/dissertation.

As part of your copyright agreement with Oxford University Press you have retained the right, after publication, to use all or part of the article and abstract, in the preparation of derivative works, extension of the article into a booklength work, in a thesis/dissertation, or in another works collection, provided that a full acknowledgement is made to the original publication in the journal. As a result, you should not require direct permission from Oxford University Press to reuse your article.

#### **CC BY**

In addition to the reuses set out above, authors who have published their article under a [CC BY](#) licence may also share and distribute their article anywhere including commercial repositories immediately on publication.

When posting, distributing or reusing Open Access articles, the journal and OUP should be clearly attributed as the original place of publication and correct citation details should be given. Authors should also deposit the URL of their published article in any repository, in addition to the Version of Record.

When making their article available according to the terms of their Open Access licence, we strongly encourage authors of Open Access papers to deposit the version of record. This will guarantee that the definitive version is readily available to those accessing your article from such repositories, and means that your article is more likely to be cited correctly.

For full details of our publication and rights policy please see the attached link to our website:  
<http://www.oxfordjournals.org/en/access-purchase/rights-and-permissions/self-archiving-policyb.html>

If you have any other queries, please feel free to contact us.

Kind regards,  
Louise

**Miss Louise Eyre** | Permissions Executive | Rights Department  
Academic and Journals Divisions | Global Business Development  
Oxford University Press | Great Clarendon Street | Oxford | OX2 6DP  
Tel: +44(0)1865 354454 | Email: [louise.eyre@oup.com](mailto:louise.eyre@oup.com) | [www.oxfordjournals.org](http://www.oxfordjournals.org)



