Imperial College London

THE FRANCIS CRICK INSTITUTE

Investigation into the function and structure of Peptidyl Arginine Deiminase 6 in early embryonic development

Jack Peter Christopher Williams
MSci, ARCS

Under the supervision of:
Dr. Louise J. Walport & Prof. Edward W. Tate

In partial fulfilment of the requirements for the degree of:
Doctor of Philosophy
September 2019 to December 2023
Department of Chemistry, Imperial College London & The Francis Crick Institute

I dedicate this thesis to my parents.

“Deep in the human unconscious is a pervasive need for a logical universe that makes sense, but the real universe is always one step beyond logic.” – Frank Herbert, Dune

Statement of Originality

This dissertation is the result of my own work, conducted under the supervision of Dr Louise J. Walport[†‡], and Prof. Edward W. Tate[†] across both the Department of Chemistry at Imperial College London and the Francis Crick Institute. Except where indicated in the text, all aspects of this work are my own. Where work was done in collaboration this is indicated in the main text, and additionally highlighted in the Materials and Methods, referencing who performed the experiments. Due references have been provided on all supporting literatures and resources. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

This thesis has been completed in accordance with Imperial College London guidelines.

Thursday 22nd Feb 2024
Jack Peter Christopher Williams MSci, ARCS
London

† Department of Chemistry
Molecular Sciences Research Hub (MSRH),
Imperial College London,
White City Campus,
Wood Lane, London
W12 0BZ, UK

‡ The Francis Crick Institute
4SW - 1 Midland Road, London,
NW1 1AT, UK

Copyright

‘The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.’

Abstract

Investigation into the function and structure of Peptidyl Arginine Deiminase 6 in early embryonic development

Jack P. C. Williams, Prof. Edward W. Tate and Dr Louise J. Walport

Peptidyl arginine deiminase 6 (PADI6) is a maternal-effect gene essential for early embryo development and female fertility. Despite its clear biological and clinical significance, the function of PADI6 remains elusive. PADI6 belongs to a family of enzymes, that catalyse the post-translational conversion of arginine to citrulline, however, there is currently no reported catalytic activity of PADI6.

The work described in this thesis aimed to expand our understanding of the function of PADI6 in early embryo development, in particular, the question of whether PADI6 has a catalytic function. Using biophysical experiments, I have shown that human PADI6 is dimeric in solution and unlike other PADIs does not bind calcium. We have solved a 2.44 Å X-ray crystal structure of human PADI6, the first high resolution structure of PADI6 to date. Analysis of this structure showed the active site cleft of PADI6 was closed and provided further insight on the structural significance of reported pathogenic *PADI6* variants. In search of interacting proteins, I have confirmed interactions with five key early embryonic proteins and shown that these proteins do not have an activating effect on PADI6. Using two mouse models, I have shown that embryonic genome activation (EGA) is defective in embryos from *Padi6* knock-out mothers, and that there may be an additional non-essential catalytic function of PADI6 during EGA. Finally, I have shown that knocking out *Padi6* in mice has little to no effect on the establishment of the maternal imprinted methylation signature. Together this thesis answers outstanding questions regarding the function of PADI6 and offers new perspectives on its function in early embryonic development.

Acknowledgements

I would first like to thank my supervisors Louise Walport and Ed Tate, for giving me the opportunity to work on this exciting project and learn so much. I would especially like to thank Louise for her continuing support and mentoring over the past 4 years.

I would especially like to thank Tessa Bertran of the Walport Group for teaching me, a chemistry undergraduate, so many different biological techniques and her continuing advice throughout my project, and in particular for keeping me positive when things weren't going well.

Having studied chemistry at university, I could not have done what I have without the teaching and advice of different members of the Niakan and Turner groups at the Crick. Most of all I would like to thank Afshan McCarthy for teaching me so much about embryology, how to harvest and stain embryos, advising me on the best experiments, and continuing to support me in my project in any way she can. I would also like to thank Afshan for all the zygotes she harvested for the embryoscope work. I would like to thank Sugako Ogushi, formerly of the Turner group for teaching me how to dissect mice, and for her help harvesting oocytes for my bisulphite sequencing work. Agata Zielinska of the Turner group for teaching me how to extract, classify and handle oocytes, and for always letting me bounce ideas off her during long sessions in the embryo room. Andrew Ojarikre of the Turner lab, who along with GeMS and the BRF, have developed my mouse lines, looked after them, set up matings, and checked for plugs on so many mornings, most of all Daiva, Bobbi, and Trinity. I would like to thank Bryony Leeke and Daniel Snell, formerly of the Turner lab, for their advice on my RNA and bisulphite sequencing work, and especially Bryony for taking time away from her new postdoc to read over the methylation chapter of my thesis. Finally, I would like to thank James Turner and Kathy Niakan for being so willing to collaborate, and all the feedback and advice they have given me along the way.

This work wouldn't have been the same without the science technology platforms here at the Crick. Of the structural biology STP, I would most of all like to thank Stephane Mouilleron, for his continued advice, harvesting crystals and for solving the structure of human PADI6 for me. I would also like to thank Simone, Laura and

Roger for teaching me all the biophysical techniques I used, Chloë for teaching me how to express proteins in mammalian cells and for donating cells to use, Donald for acquiring a Cryo-EM dataset of mouse PADI6 for me, and Phil for his constant positivity and help. Of the proteomics STP I would like to thank Tania for all her hard work and advice on our collaborative single embryo proteomics work, and for her constant cheer and humour. I would also like to thank Sarah Maslen, for running so many Intact-MS samples for me. Of the Crick EM facility, I would like to thank Anne Weston and Lucy Collinson for helping us image the cytoplasmic lattices in GV oocytes. And of the Crick Advanced Sequencing Facility I would like to thank Daniel Leonce for preparing my scRNA-seq libraries for me, and for the rest of the facility for all of the QC and sequencing they perform.

I would like to thank my thesis committee Neil McDonald, Anna Barnard and Luiz Pedro Carvalho for all their advice and suggestions throughout my project.

I would also like to thank everyone in the Walport group. I would like to thank Cat, Jemima and Tessa for all taking the time to read my thesis chapters. Ewen for synthesising the chemical probe used in this work. Tom for his cheerfulness and positivity throughout. And everyone else for making the lab such an amazing environment to work in. Not in the Walport lab, but I would also like to thank Matthew Cockman for always being willing to discuss experimental ideas and provide suggestions, in particular regarding proteomics.

Finally, I would like to thank my parents for their unwavering support throughout my life, in whatever I choose to do. No matter how I feel or where I am, they have always done everything they possibly can to help. I would also like to thank my brother Tom for his continued life and career advice. And lastly, I would like to thank Amy Stokes for being my rock and bringing so much happiness and stability to my life.

Table of Contents

Copyright….. 5
Abstract…… 6
Acknowledgements 7
Table of Contents 9
List of Figures 13
List of Tables 16
Abbreviations and Acronyms 17
Chapter 1.Introduction 22
1.1 Peptidyl arginine deiminases 22
1.1.1 Citrullination 22
1.1.2 Peptidyl arginine deiminases 22
1.1.3 PADI conservation and overall structure 23
1.1.4 The PADI active site 25
1.1.5 Calcium binding 28
1.2 PADI6 in pre-implantation embryo development 31
1.2.1 Oogenesis 33
1.2.2 Fertilisation and zygotic genome activation 42
1.2.3 Gastrulation 46
1.2.4 PADIs in the early embryo 49
1.2.5 Citrullination in the early embryo 50
1.2.6 Is PADI6 regulated by 14-3-3 proteins in the early embryo? 51
1.3 PADI6, DNA methylation and maternal imprinting 52
1.3.1 DNA methylation 54
1.3.2 Genomic imprinting 57
1.3.3 DNA methylation through oogenesis and early embryo development 58
1.3.4 Imprinting disorders 61
1.4 Project Aims 64
Chapter 2.Materials and Methods 65
2.1 General materials 65
2.1.1 Antibodies 65
2.1.2 Bacterial strains 65
2.1.3 Antibiotics 66
2.1.4 Buffers 66
2.1.5 Primers 67
2.1.6 Plasmids 67
2.2 General molecular biology methods 69
2.2.1 Plasmid transformation 69
2.2.2 DNA sequencing 69
2.2.3 In-Fusion cloning 69
2.2.4 Site-directed mutagenesis 71
2.2.5 DNA plasmid purification 71
2.2.6 Reverse transcription-quantitative PCR 72
2.2.7 SDS-PAGE 73
2.2.8 Western blot 73
2.2.9 Co-immunoprecipitation 74
2.3 Recombinant protein expression and purification 76
2.3.1 *E. coli* large scale expression 76
2.3.2 *E. coli* cell lysis and clearing 76
2.3.3 *E. coli* small scale trial expression 76
2.3.4 Expi293 expression 77

2.3.5 Expi293 cell lysis and clearing 77
2.3.6 Nickel immobilised metal affinity chromatography 77
2.3.7 In solution tag cleavage by TEV or 3C protease 78
2.3.8 Strep-tag affinity purification and cleavage off resin 78
2.3.9 Size-exclusion chromatography 78
2.3.10 Ion-exchange chromatography 79
2.3.11 Band-ID 79
2.3.12 Intact-MS of purified proteins 79
2.4 Biophysical methods 80
2.4.1 Circular dichroism 80
2.4.2 Nano differential scanning fluorimetry (Nano-DSF) 80
2.4.3 NanoDSF calcium titration 80
2.4.4 Size Exclusion Chromatography multi-angle laser light scattering (SEC-MALLS) 81
2.4.5 Mass photometry 81
2.4.6 Buffer screening 81
2.5 X-Ray crystallography 82
2.5.1 Reagents 82
2.5.2 Protein crystallisation 82
2.5.3 Preparation of multi matrix seed stock 82
2.5.4 Seeded protein crystallisation 83
2.5.5 Crystal harvesting, cryoprotection and structure determination 83
2.6 Cryo-EM 84
2.6.1 Data acquisition 84
2.6.2 Electron density map determination 84
2.7 Cell culture methods 85
2.7.1 General cell culture methods 85
2.7.2 Transient transfection 85
2.7.3 Cell harvesting 86
2.8 Citrullination characterisation methods 86
2.8.1 COLDER assays 86
2.8.2 Phenyl glyoxal probe 86
2.9 Mouse work 88
2.9.1 General husbandry 88
2.9.2 Humane culling procedure 88
2.9.3 Timed natural matings 89
2.9.4 Collection of GV oocytes 89
2.9.5 Collection of MII oocytes 89
2.9.6 Collection of zygotes 90
2.9.7 Collection of 2-cell embryos 90
2.9.8 Dissociation of 2-cell embryos into blastomeres 91
2.9.9 Time-lapse imaging of zygote development to blastocyst 91
2.9.10 mRNA extraction from mouse ovaries 91
2.9.11 Ovary mRNA amplification and sequencing 92
2.10 Immunofluorescent imaging 93
2.10.1 Zygote fixation and staining 93
2.10.2 Image acquisition and analysis 93
2.11 Transmission electron microscopy of GV oocytes 94
2.11.1 GV oocyte harvesting and initial fixation 94
2.11.2 Full fixation, sectioning and imaging 94
2.12 Single cell RNA sequencing 95
2.12.1 Sample preparation 95
2.12.2 Library preparation and sequencing 95

2.12.3 Data processing 96
2.12.4 PCA 96
2.12.5 Gene counts per million box plots 96
2.12.6 MA plots 96
2.12.7 Minor and major EGA gene identification 97
2.12.8 EGA gene heatmap 97
2.13 Embryo proteomics 97
2.13.1 Sample preparation and processing 97
2.13.2 Lysis buffers 98
2.13.3 Initial method 98
2.13.4 Adapted One-Tip method 98
2.13.5 Data analysis 98
2.14 Low input bisulphite sequencing 99
2.14.1 GV oocyte sample harvesting and preparation 99
2.14.2 HeLa gDNA harvesting 99
2.14.3 In-house sequencing primers 99
2.14.4 Library preparation 99
2.14.5 SPRIselect bead clean-up 100
2.14.6 Data processing 101
2.14.7 CpG coverage analysis 101
2.14.8 PCA of datasets 102
2.14.9 Analysis of methylation at different genomic features 102
2.15 Statistical analysis 103
Chapter 3.Structural characterisation of PADI6 105
3.1 Introduction 105
3.2 Human PADI6 106
3.2.1 Initial expression and purification of hPADI6 in *E. coli* 106
3.2.2 Optimisation of hPADI6 expression and purification in *E. coli* 108
3.2.3 Biophysical characterisation of recombinant hPADI6 from *E. coli* 109
3.2.4 COLDER assay of recombinant hPADI6 from *E. coli* 112
3.2.5 Initial crystallisation of recombinant hPADI6 from *E. coli* 113
3.2.6 Optimisation of hPADI6 expression in *E. coli* using pET28a vector 114
3.2.7 Final *E. coli* expressed hPADI6 crystallisation 117
3.2.8 Expression and purification of hPADI6 in mammalian Expi293 cells 119
3.2.9 Biophysical characterisation of recombinant hPADI6 from Expi293 cells 120
3.2.10 Calcium binding of recombinant hPADI6 122
3.2.11 Crystallisation of recombinant hPADI6 from Expi293 cells 122
3.2.12 Structural analysis of hPADI6 125
3.3 Mouse PADI6 139
3.3.1 Initial expression and purification of mPADI6 in *E. coli* 139
3.3.2 Expression and purification of mPADI6 in Expi293 cells 140
3.3.3 Biophysical characterisation of recombinant mPADI6 141
3.3.4 Cryo-EM structure determination of recombinant mPADI6 143
3.4 Summary 145
Chapter 4.hPADI6 and the subcortical maternal complex 148
4.1 Introduction 148
4.2 Investigation into the interaction between the SCMC and PADI6 149
4.2.1 hPADI6 interaction with the SCMC 149
4.2.2 Individual SCMC protein interactions 150
4.2.3 Core SCMC protein interactions with hPADI6 variants 152
4.3 SCMC recombinant expression 153
4.3.1 Expression and purification of SCMC proteins from *E. coli* 154
4.3.2 Expression and purification of SCMC proteins from Expi293 cells 156

4.4 Investigation into possible citrullination by, or activation of hPADI6 by the SCMC 160
4.4.1 Probe validation 161
4.4.2 Investigation into possible citrullination of hPADI6 in HEK-293T cells 162
4.5 Summary 164
Chapter 5.PADI6 in the early embryo 168
5.1 Introduction 168
5.2 Catalytic mutant mouse model 169
5.2.1 Rationale 169
5.2.2 Characterisation of recombinant mPADI6^{C663A} 171
5.2.3 Generation of *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ female mice 173
5.3 Fertility assessment of *Padi6* transgenic female mice 174
5.4 Cytoplasmic lattices in *Padi6* transgenic oocytes 179
5.5 Transcriptome characterisation of *Padi6* transgenic oocytes and embryos 182
5.5.1 Experimental design 182
5.5.2 Sample harvesting and library preparation 183
5.5.3 Data quality analysis 183
5.5.4 Transcriptome differences compared to WT oocytes and embryos 189
5.5.5 Investigation of EGA impairment 195
5.6 Establishment of single embryo proteomics platform 199
5.6.1 Single embryo proteomics workflow 200
5.6.2 GV oocyte differential protein analysis 202
5.6.3 One-Tip adapted workflow 205
5.7 Summary 207
Chapter 6.PADI6 and maternal imprinted signature establishment in mice 210
6.1 Introduction 210
6.2 Low-input bisulphite sequencing protocol establishment 212
6.2.1 Experimental design 212
6.2.2 Sample harvesting 215
6.2.3 Initial library preparation trial 215
6.2.4 Further optimisation and Pilot Study 1 218
6.2.5 Pilot study 2 222
6.3 Bisulphite sequencing of *Padi6* transgenic GV oocytes 226
6.4 Summary 236
Chapter 7.Discussion 239
7.1 Overview 239
7.2 Does PADI6 have a catalytic function? 244
Chapter 8.References 248
Appendices 274

List of Figures

Figure 1.1 PADI-catalysed citrullination scheme
Figure 1.2 PADI structure
Figure 1.3 Proposed scheme for PADI-catalysed citrullination
Figure 1.4 Human PADI active site conservation
Figure 1.5 PADI calcium binding and activation
Figure 1.6 Human PADI6 calcium binding site conservation with human PADI4
Figure 1.7 Arrest timings of mouse and human embryos lacking PADI6
Figure 1.8 Oogenesis overview
Figure 1.9 Ovulation
Figure 1.10 Cytoplasmic lattices
Figure 1.11 Electron tomography of CPL structure
Figure 1.12 EGA timings in mouse and human pre-implantation embryos
Figure 1.13 Mammalian pre-implantation embryo development
Figure 1.14 Cl-Amidine structure and PADI inhibition
Figure 1.15 PADI6 variants associated with MLID
Figure 1.16 DNA methylation overview
Figure 1.17 DNMT domains and recognition
Figure 1.18 Imprinting gDMR overview
Figure 1.19 Maternal genome methylation dynamics in development
Figure 2.1 KingFisher IP Protocol
Figure 3.1 Expression and purification of pMCSG_His-TEV-hPADI6 from *E. coli*
Figure 3.2 Expression and purification of pMCSG_His-3C-hPADI6 from *E. coli*
Figure 3.3 Biophysical characterisation of recombinant hPADI6 from *E. coli*
Figure 3.4 COLDER assays of recombinant hPADI6 and hPADI4
Figure 3.5. Initial crystal hit of recombinant hPADI6 purified from *E. coli*
Figure 3.6. hPADI6 test expression comparison of pMCSG_His-3C and pET28a_His constructs
Figure 3.7. Expression and purification of pET28a_His-hPADI6 from *E. coli*
Figure 3.8. Test expressions to optimise hPADI6 solubility
Figure 3.9. Expression and purification of pcDNA3.1_SS-TEV-hPADI6 from Expi293 cells
Figure 3.10. Biophysical characterisation of hPADI6 from Expi293 cells
Figure 3.11 hPAD6 is not stabilised by $CaCl_2$
Figure 3.12. hPADI6 crystal optimisation

Figure 3.13 Overall hPADI6 X-Ray crystal structure
Figure 3.14 hPADI6 structural conservation with hPADI4
Figure 3.15 hPADI6 active site pocket structural conservation with hPADI4
Figure 3.16 hPADI6 active site structure
Figure 3.17 Cysteines around the hPADI6 active site
Figure 3.18 Local structure of selected hPADI6 dimerisation and interaction variants
Figure 3.19 hPADI6 variant expression in HEK-293T cells
Figure 3.20 Expression and purification of pMCSG_His-3C-mPADI6 from *E. coli*
Figure 3.21 Expression and purification of pcDNA3.1_SS-TEV-mPADI6 from Expi293 cells
Figure 3.22 Biophysical characterisation of mPADI6
Figure 3.23 mPADI6 Cryo-EM electron density map calculation workflow
Figure 4.1 Co-IP of full SCMC complex with hPADI6
Figure 4.2 Co-IP of individual SCMC proteins with hPADI6
Figure 4.3 Co-IP of individual core SCMC proteins with hPADI6 variants
Figure 4.4 Trial expressions of SCMC proteins in *E. coli*
Figure 4.5 Trial expressions of SCMC proteins in Expi293 cells
Figure 4.6 Large scale expressions of TLE6 and NALP5 in Expi293 cells
Figure 4.7 Optimisation of SCMC expression in Expi293 cells
Figure 4.8 PG-TAMRA citrulline probe
Figure 4.9 PG-TAMRA probe validation with hPADI4 and histone citrullination
Figure 4.10 PG-TAMRA probe investigation into possible citrullination by hPADI6 in transfected HEK-293T cells
Figure 4.11 PADI6 and SCMC suggested interactions
Figure 5.1 *Padi6* transgenic mouse lines
Figure 5.2 Expression and purification of mPADI6^{C663A} from Expi293 cells
Figure 5.3 Biophysical characterisation of mPADI6^{C663A}
Figure 5.4 Transgenic mouse line validation
Figure 5.5 Mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ embryo developmental potential
Figure 5.6 Mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ embryo developmental timings
Figure 5.7 CPLs in GV oocytes from WT, *Padi6*^{C663A} and *Padi6*$^{-/-}$ females
Figure 5.8 scRNA-seq sample QC and clustering
Figure 5.9 CPM of *Padi6* in MII oocytes, zygotes and 2-cell blastomeres
Figure 5.10 CPM of SCMC genes in MII oocytes, zygotes and 2-cell blastomeres
Figure 5.11 Up and downregulation of genes in the MII oocyte

Figure 5.12 Up and downregulation of genes in the zygote
Figure 5.13 Up and downregulation of genes in the 2-cell blastomere
Figure 5.14 GO analysis of up and downregulated genes in 2-cell blastomeres
Figure 5.15 Identification of minor and major EGA genes
Figure 5.16 Clustering of minor and major EGA gene levels transgenic embryos
Figure 5.17 Omics organisation
Figure 5.18 Initial single oocyte proteomics workflow
Figure 5.19 Initial WT GV oocyte proteomics trial
Figure 5.20 Initial workflow comparative proteomics of *Padi6*^{mutant} GV oocytes
Figure 5.21 Adapted One-Tip method WT GV oocyte trial
Figure 6.1 Bisulphite sequencing overview
Figure 6.2 Initial library preparation trial fragment distributions
Figure 6.3 Bead clean-up optimisation fragment distributions
Figure 6.4 Pilot Study 1 library fragment distributions
Figure 6.5 Pilot Study 1 CpG coverage
Figure 6.6 Pilot Study 2 library fragment distributions
Figure 6.7 Pilot Study 2 CpG coverage
Figure 6.8 Full experiment library fragment distributions
Figure 6.9 Data processing summary of full experiment compared with Pilot Study 2
Figure 6.10 PCA of CpG methylation in GV oocytes
Figure 6.11 Full experiment CpG coverage
Figure 6.12 Methylation percentage at different genomic loci in GV oocytes
Figure 6.13 Scatter plot of methylation percentage at different genomic loci in GV oocytes
Figure 6.14 Mouse maternal gDMR methylation heatmap in GV oocytes
Figure 7.1 Questions answered in this work
Figure 7.2 New questions and perspectives regarding the function of PADI6

List of Tables

Table 1.1 PADI tissue expression and disease relevance (8)
Table 1.2 Human and mouse PADI BLAST sequence conservation
Table 1.3 PADI6 variants that result in early embryonic arrest and female infertility
Table 1.4 SCMC genes and mouse knock-out phenotypes
Table 1.5 PADI6 variants resulting in late-stage pregnancy complications
Table 1.6 PADI6 variants that cause MLID
Table 2.1 Antibodies
Table 2.2 Bacterial strains
Table 2.3 Antibiotics
Table 2.4 Buffers
Table 2.5 Purchased, donated or InFusion cloned plasmids
Table 2.6 X-Ray crystallography reagents
Table 2.7 Transient transfection DNA quantities
Table 2.8 Cell lysate transient transfection DNA quantities
Table 2.9 Proteomic lysis buffers
Table 3.1 hPADI6 variants selected for further characterisation
Table 5.1 WT and mat-*Padi6*$^{-/-}$ embryo transition times
Table 6.1 Initial library preparation trial library DNA concentrations
Table 6.2 Pilot Study 1 library DNA concentrations
Table 6.3 Pilot Study 1 sequencing and data processing statistics
Table 6.4 Pilot Study 2 library DNA concentrations
Table 6.5 Pilot Study 2 sequencing and data processing statistics after 20 million reads
Table 6.6 Pilot Study 2 sequencing and data processing statistics after 200 million reads
Table 6.7 Full experiment library DNA concentrations

Abbreviations and Acronyms

5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
A (Ala)	Alanine
ADD	ATRX-DNMT3L-DNMT3A domain
AMOT	Angiomotin
aPKC	Atypical protein kinase C
Apo-hPADI4	Calcium free hPADI4 X-ray crystal structure (1WD8)
ArnA	Bifunctional polymyxin resistance protein ArnA
BAEE	N_{α}-Benzoyl-L-arginine ethyl ester
BS-seq	Bisulphite sequencing
BWS	Beckwith-Wiedemann syndrome
B2m	Beta-2-microglobulin
C (Cys)	Cysteine
Ca	Calcium ion
Ca-hPADI4^{C645A}-BAEE	Calcium and BAEE bound hPADI4 X-ray crystal structure (1WDA)
Can	Carbonic anhydrase 2
CD	Circular dichroism
CDX2	Homeobox protein CDX-2
CGI	CpG island
ChIP-seq	Chromatin immunoprecipitation sequencing
COLDER	Colour developing reagent
Co-IP	Co-immunoprecipitation
CpG	CG dinucleotide
CPL	Cytoplasmic lattice
CPM	Counts per million
Cryo-EM	Cryo electron microscopy
C57BL/6N	Black 6 mouse strain
D (Asp)	Aspartic acid
DDM	N-Dodecyl-β-D-maltoside
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3B	DNA methyltransferase 3B
DNMT3L	DNA methyltransferase 3L
DTT	Dithiothreitol
DUX	Double homeobox protein

E (Glu)	Glutamic acid
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EGA	Embryonic genome activation
EM	Electron microscopy
ER	Endoplasmic reticulum
F (Phe)	Phenylalanine
G (Gly)	Glycine
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GATA3	GATA-binding factor 3
gDMR	Germline differentially methylated region
GeMS	Crick genetic modification service
GFP	Green fluorescent protein
GlmS	Glutamine-fructose-6-phosphate aminotransferase
GO	Gene ontology
GRCh38	Homo sapiens genome assembly hg38
GRCm38	Mus musculus genome assembly mm10
GRCm39	Mus musculus genome assembly mm39
GV	Germinal vesicle
H (His)	Histidine
HEK-293T	Human embryonic kidney 293 cells
HM	Hydatidiform mole
hPADI4	Human peptidyl arginine deiminase 4
hPADI6	Human peptidyl arginine deiminase 6
HRP	Horseradish peroxidase
H1	Linker histone H1
H3R2	Histone 3 arginine 2
H3K4	Histone 3 lysine 4
H3K4me3	Histone 3 lysine 4 trimethylation
H3K9	Histone 3 lysine 9
H3K27	Histone 3 lysine 27
H3K27ac	Histone 3 lysine 27 acetylation
H3K36	Histone 3 lysine 36
H3K36me3	Histone 3 lysine 36 trimethylation
H4Cit3	Histone 4 citrullinated arginine at position 3
I (Ile)	Isoleucine
ICR	Imprinting control region
IEX	Ion exchange chromatography

IMAC	Immobilised metal affinity chromatography
IP	Immunoprecipitation
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
K (Lys)	Lysine
KHDC3L	KH domain-containing protein 3
KSCN	Potassium thiocyanate
L (Leu)	Leucine
LATS2	Large tumour suppressor kinase 2
LB	Lysogeny broth
LiRibo-seq	Low input ribosome sequencing
LC	Liquid chromatography
M (Met)	Methionine
MLID	Multi-locus imprinting disorder
mRNA	Messenger ribonucleic acid
mPADI4	Mouse peptidyl arginine deiminase 4
mPADI6	Mouse peptidyl arginine deiminase 6
MS	Mass spectrometry
MSY2	Y-box-binding protein 2
MS/MS	Tandem mass spectrometry
M_w	Molecular weight
MZT	Maternal-to-zygotic transition
MII	Metaphase 2 arrested
N (Asn)	Asparagine
NanoDSF	Nano differential scanning fluorimetry
NALP2	NACHT, LRR and PYD domains-containing protein 2
NALP5	NACHT, LRR and PYD domains-containing protein 5
NALP7	NACHT, LRR and PYD domains-containing protein 7
NLS	Nuclear localisation signal
OCT4	POU domain, class 5, transcription factor 1
OOEP	Oocyte-expressed protein
P (Pro)	Proline
PADI	Peptidyl arginine deiminase
PBS	Phosphate-buffered saline
PCA	Principle component analysis
PCNA	Proliferating cell nuclear antigen domain
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PG	Phenylglyoxal

PGC	Primordial germ cell
PPI	Protein-protein interaction
PTM	Post-translational modification
PWWP	Proline-Tryptophan-Tryptophan-Proline domain
Q (Gln)	Glutamine
R (Arg)	Arginine
RFTS	Replication foci-targeting sequence
RNA Pol II	RNA Polymerase II
RT	Room temperature
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
S (Ser)	Serine
S.O.C.	Catabolite repression media
scBS-seq	Single cell bisulphite sequencing
SCMC	Subcortical maternal complex
scRNA-seq	Single cell RNA sequencing
SDS	Sodium dodecyl-sulphate
SDS-PAGE	Sodium dodecyl-sulphate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SEC-MALLS	Size exclusion chromatography
SETD2	Histone-lysine N-methyltransferase SETD2
SlyD	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD
SRS	Silver-Russell syndrome
T (Thr)	Threonine
TAE	Tris base, acetic acid, EDTA
TAMRA	5-Carboxytetramethylrhodamine
TCA	Trichloroacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TE	Trophectoderm
TEM	Transmission electron microscopy
TET1	Methylcytosine dioxygenase TET1
TET2	Methylcytosine dioxygenase TET2
TET3	Methylcytosine dioxygenase TET3
TLE6	Transducin-like enhancer protein 6
T_m	Melting temperature
TRC	Transcription requiring complex
TS	Temple syndrome
TSS	Transcription start site
UHRF1	E3 ubiquitin-protein ligase UHRF1

V (Val)	Valine
W (Trp)	Tryptophan
WT	Wild-type
Y (Tyr)	Tyrosine
YAP1	Yes-associated protein 1
ZBED3	Zinc finger BED domain-containing protein 3
ZFP57	Zinc finger protein 57
ZFP445	Zinc finger protein 445
ZP	Zona pellucida

Chapter 1. Introduction

1.1 Peptidyl arginine deiminases

1.1.1 Citrullination

Citrullination is the post-translation conversion of arginine to citrulline (**Figure 1.1**) (1–3). The conversion of the coded arginine to the non-coded citrulline can lead to significant changes in a substrate protein's structure and function. At physiological pH, arginine is positively charged with five H-bond donors, whereas citrulline is neutral with three H-bond donors and two acceptors. As such, citrullination can result in changes to a substrate protein's folded structure, or protein-protein interactions (PPI) by changing the residue's capacity to interact with other amino acids (4–6). This can ultimately modulate protein function by stabilising or destabilising a protein, or by blocking or facilitating PPIs.

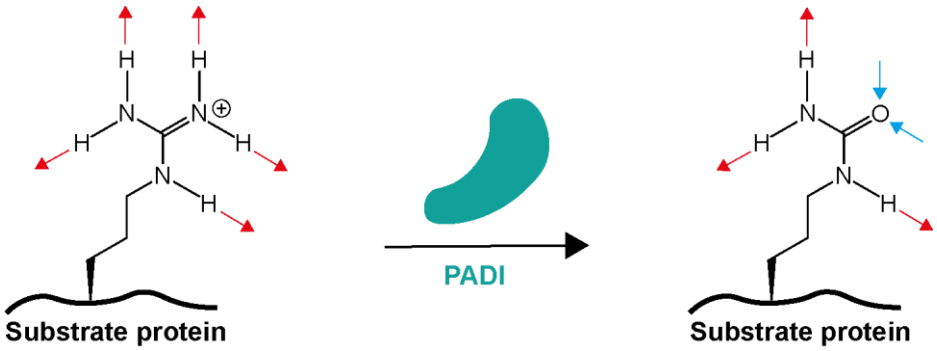


Figure 1.1 PADI-catalysed citrullination scheme
Scheme of the PADI-catalysed post-translational conversion of the positively charged arginine to the neutral citrulline. Red arrows = H-bond donors, Blue arrows = H-bond acceptors.

1.1.2 Peptidyl arginine deiminases

Citrullination is catalysed by peptidyl arginine deiminases (PADIs). In humans and mice there are five PADI paralogs, PADIs 1 to 4, and PADI6 (4,5,7–10). There is currently no reported catalytic activity for PADI6 however, and its classification as a PADI is based only on sequence conservation and genomic co-localisation with the other four (7,11). PADI enzymes are ubiquitous across vertebrate animals, but absent from yeast, worms and flies, and it has been proposed that *PADIs* were introduced into mammals by horizontal gene transfer from cyanobacteria (12).

The human PADI enzymes display varied and specific tissue expression patterns, and their dysregulation has been associated with several diseases (**Table 1.1**) (2,4,5,13,14,14–20). Out of the five human PADIs, only PADI4 possesses a canonical nuclear localisation signal, although there is considerable evidence that PADI2 and PADI4 are both also present in the nucleus (21,22). Whilst the substrate profile of each PADI is incomplete, various substrates have been confirmed for PADIs 1-4, with some overlap between PADIs; PADIs 1 and 3 are both reported to citrullinate keratins, with PADI2 and PADI4 both reported to citrullinate histones, despite PADI2 not possessing a canonical NLS (8,23,24). Other reported substrates of PADI2 include vimentin, tubulin, and RNA Polymerase II (RNA Pol II), where citrullination in the C-terminal domain of RNA Pol II has been shown to be crucial for regulating transcriptional pausing on DNA (25,26). As of now, no PADI6 substrates have been identified.

Table 1.1 PADI tissue expression and disease relevance (8)

PADI	Tissue expression	Associated diseases
1	Epidermis, uterus	-
2	Skeletal muscle, brain, inflammatory cells, secretory glands	Multiple sclerosis, Alzheimers disease, Rheumatoid arthritis
3	Hair follicles, keratinocytes	-
4	Granulocytes, cancerous cells	Cancer, Rheumatoid arthritis
6	Oocyte, pre-implantation embryo	Infertility, imprinting disorders

1.1.3 PADI conservation and overall structure

Overall, the sequences of the human PADIs are fairly conserved, with an average sequence identity of 54%. Human PADI6 is somewhat of an outlier, with an average identity of 44% with the other human PADIs. Full sequence alignments of the human PADIs can be found in **Appendix 1**. Mice, also have five PADI enzymes, and the human and mouse PADI sequences are highly conserved (**Table 1.2**) as such, mice are an ideal model system for studying PADI function.

Table 1.2 Human and mouse PADI BLAST sequence conservation

PADI	Human/mouse identity (%)	Human/mouse homology (%)
1	77	85
2	93	97
3	87	93
4	73	85
6	67	82

High resolution crystal structures have been reported for human PADI2 and human PADI4 (27,28). In mammals, it appears that the general structure of PADI enzymes remains conserved. They are composed of two domains: an N-terminal domain, containing two IgG like subdomains, and a final C-terminal catalytic 'PADI' domain (**Figure 1.3a**). It is within this C-terminal PADI domain that the active site and catalytic tetrad lie. The domain is composed of five ββαβ modules in a pseudo-five-fold symmetric structure known as an α/β-propeller. Interestingly, both crystallise as head-to-tail dimers. The function of this dimerization is unclear; however, it has been proposed that dimerization enhances citrullination activity (29). Although it must be noted that blocking dimerisation does not ablate PADI4 citrullination activity.

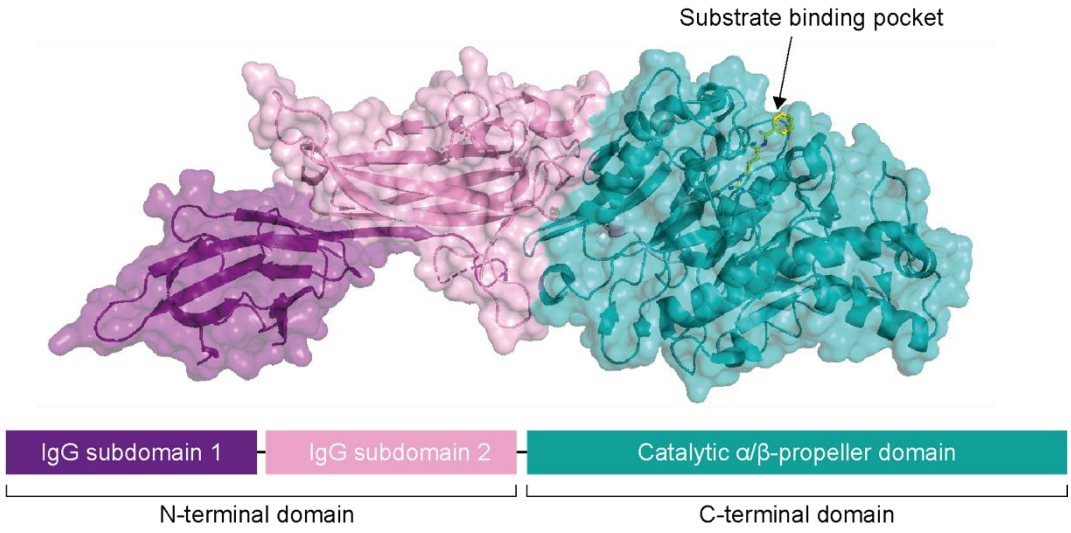


Figure 1.2 PADI structure
PADI domain architecture. Structure displayed is that of calcium and substrate bound human Ca-hPADI4^{C645A}-BAEE (PDB: 1WDA) published by Sato and co-workers (28).

1.1.4 The PADI active site

PADIs are cysteine hydrolase enzymes (30). Within the PADI catalytic pocket resides a nucleophilic cysteine residue, a histidine and two aspartic residues making up the catalytic tetrad. The chemical mechanism of citrullination was initially proposed in 2004 by Sato and co-workers based on biochemical and crystallographic information (**Figure 1.2**) (28). The conversion proceeds through four intermediate stages, initiated by the nucleophilic attack of the active site cysteine on the central guanidinium carbon of the arginine. Protonation of the newly formed amino group by the active site histidine results in the collapse of this intermediate and expulsion of ammonia, before hydrolysis of the imine to form citrulline. Substitution of any of the four key catalytic residues ablates any catalytic activity (28).

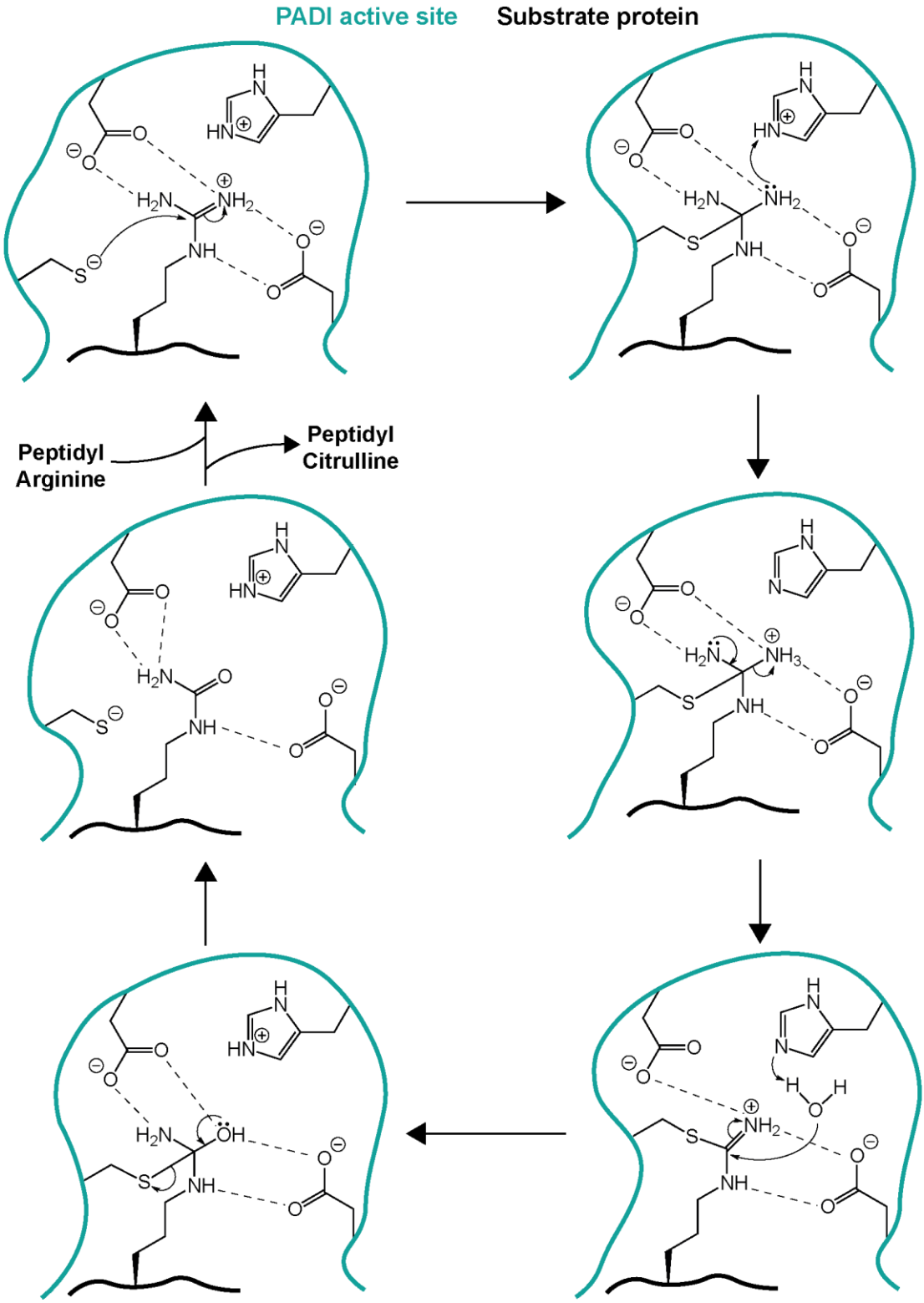


Figure 1.3 Proposed scheme for PADI-catalysed citrullination
Schematic of the proposed PADI-catalysed citrullination mechanism (28). The catalytic tetrad residues (aspartic acid, aspartic acid, histidine, cysteine) are shown along with the arginine residue of a substrate protein. After citrulline is formed, the cycle restarts with a new substrate arginine.

In the human PADI enzymes, the histidine and both aspartic acid residues are conserved (**Figure 1.4a-b**). However, the nucleophilic cysteine is only conserved in human PADIs 1 to 4, where human PADI6 has an alanine flanked by two cysteines (**Figure 1.4c**). The sequence proximity of the two cysteines to the position of the key nucleophilic cysteine in PADIs 1 to 4 raises the intriguing question of whether PADI6 does have a catalytic function, and if so, whether it is through one of these two cysteines. Another question that this raises is whether these two cysteines participate in disulphide bond formation and the possibility that catalytic activity could be modulated by redox chemistry. Interestingly, only the first of these two cysteines is conserved in the PADI6 sequence of multiple species (**Figure 1.4d**, **Appendix 2**), suggesting that if PADI6 is catalytically active, then this cysteine would be involved. The evolutionary reason for its positional displacement compared to the other PADIs is unknown.

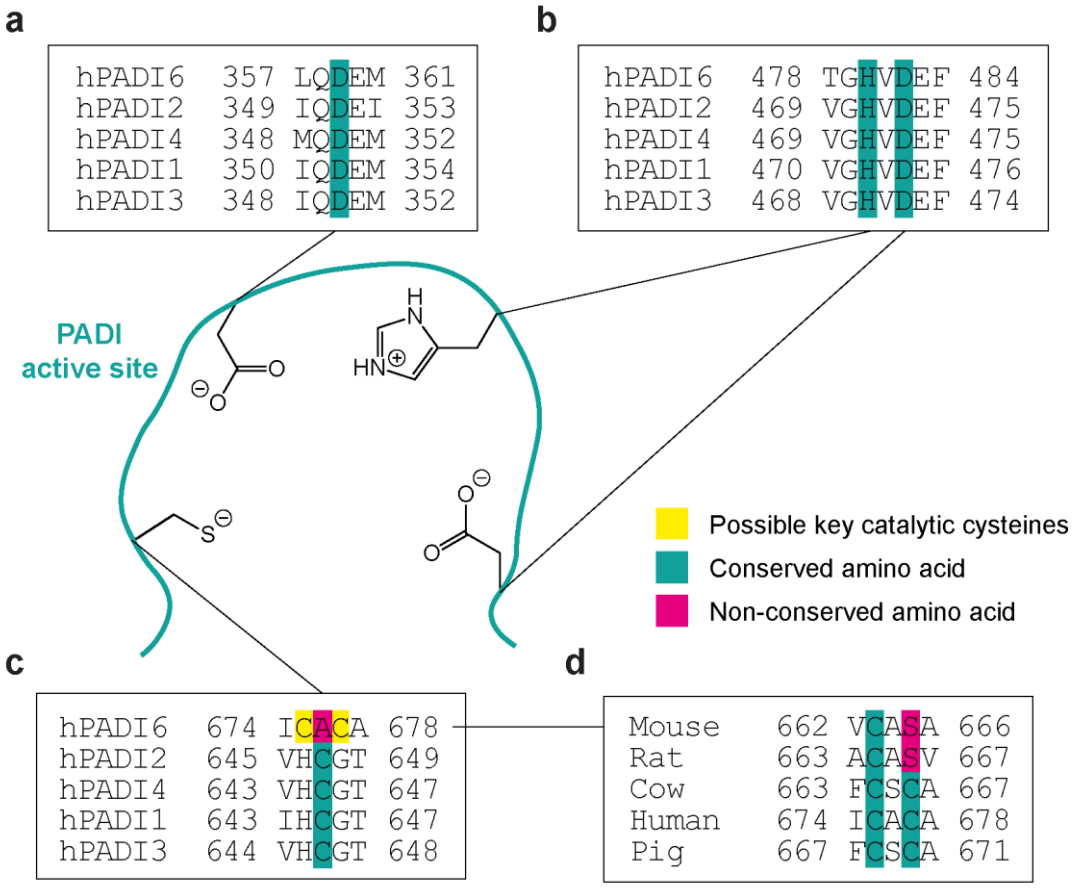


Figure 1.4 Human PADI active site conservation
a)-b) Sequence alignments of the human PADI active site, highlighting the conserved histidine and two aspartic acid residues. **c)** Sequence alignments of human PADI key catalytic cysteine residue, highlighted is the key catalytic cysteine conserved in human PADIs 1 to 4, but replaced by an alanine flanked by two possible key catalytic cysteines in PADI6. **d)** Sequence alignments of the mouse, rat, cow, human, and pig PADI6 active site. Sequence alignments produced using Clustal Omega (31).

1.1.5 Calcium binding

PADIs 1 to 4 bind between four and six Ca^{2+} ions in a stepwise manner at defined pockets in their structure (**Figure 1.5a**) (27,28). In the absence of Ca^{2+} ions, PADI enzymes are inactive *in vitro*. The mechanism of human PADI2 and PADI4 Ca^{2+}activation has been studied in depth through crystallographic, biophysical and mutagenesis experiments (27). Calcium binding induces structural changes in the protein, opening the active site pocket and moving the key catalytic cysteine residue into place for citrullination, activating the protein. This mechanism has been described in the most detail for human PADI2 by Thompson and co-workers, where high resolution crystal structures have been obtained with increasing calcium concentrations (27). This work identified that calcium ion (Ca) 1 and Ca6 first bind PADI2 with high affinity. This is then followed by binding of Ca3 to Ca5 which induce conformational changes that induce Ca2 binding. Ca3 to Ca5 binding is thus termed the calcium switch of the protein, as their affinities are much lower than the other calcium binding sites, and much closer to the concentration required to activate the protein *in vitro*. In the calcium free form an arginine residue shields the active site from substrate binding, R347 (**Figure 1.5b2**). Furthermore, the key active site cysteine residue (C647) is displaced away from the active site pocket. Upon binding of Ca2, R347 moves out of the active site, with tryptophan 348 (W348) swinging in to form a wall in the active site pocket, and C647 moving into the active site ready for catalysis. This calcium switching mechanism has been validated in both human PADI2 and PADI4 (27,28). Interestingly, the calcium concentrations required for maximal in vitro activity (1-10 mM) are far higher than even the highest physiological calcium concentrations (10 μM) (16). This has left the open question of how PADIs are activated *in vivo*.

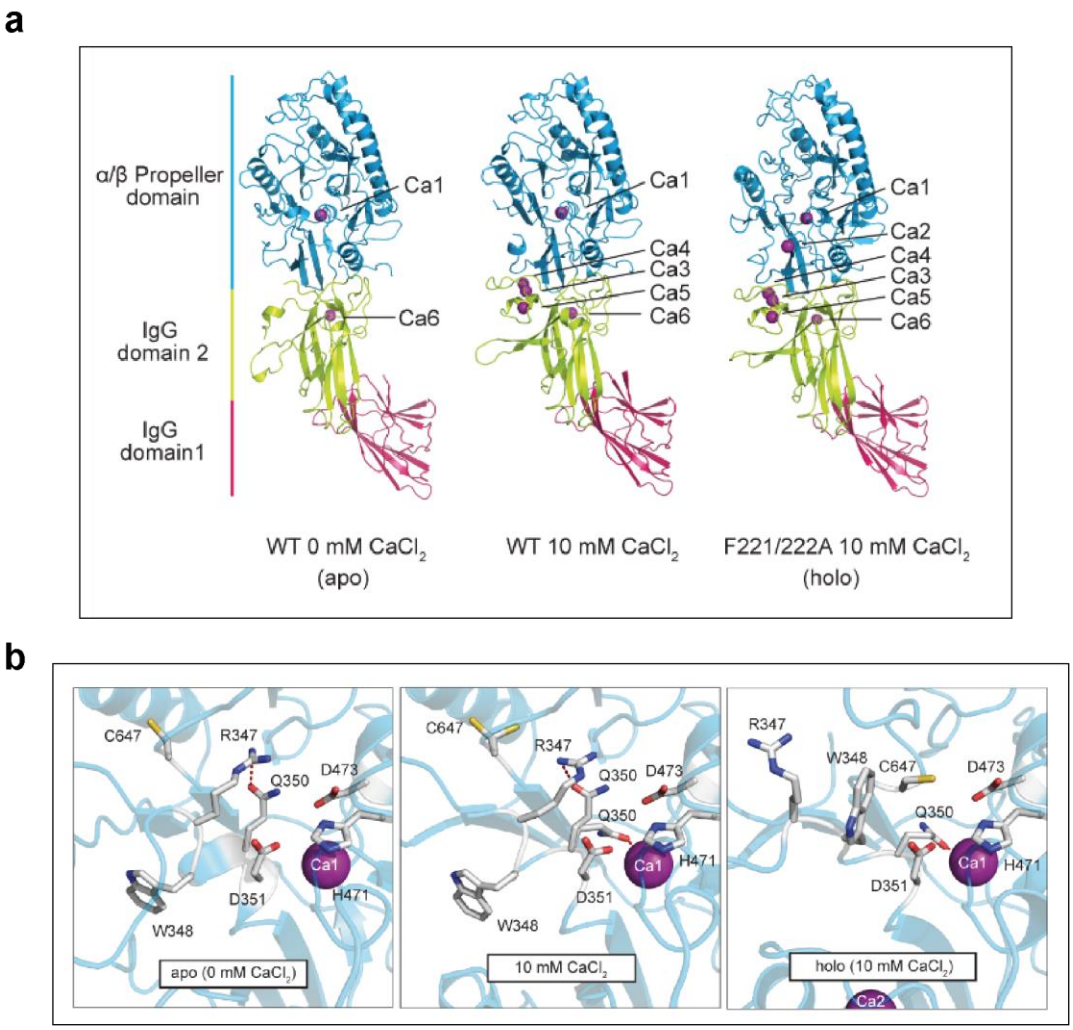
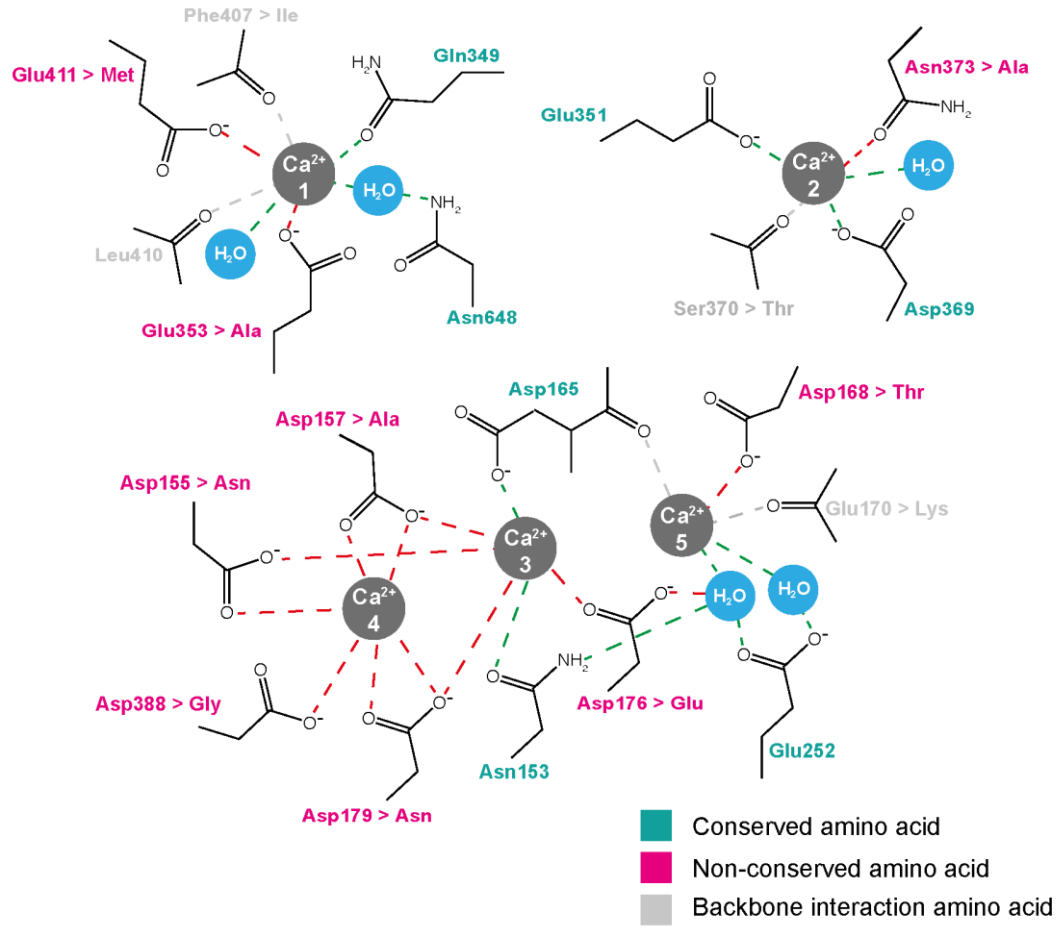


Figure 1.5 PADI calcium binding and activation
a) PADI2 calcium free and calcium bound structures, figure taken from reference (27). **b)** Comparison of the PAD2 active site in the apo (left), 10 mM calcium (centre), and holo (right), figure reproduced from reference (27) with permission.

Calcium binding and the subsequent activation switch described for PADI2 is conserved in PADIs 1 to 4, with the main difference being each PADI binds a different number of calcium ions (8). Human PADI6 however, is an outlier. Many of the residues involved in calcium binding in the other PADIs are not conserved in PADI6; only seven out of the 16 amino acids directly involved in binding in PADI4 are retained in human PADI6 (**Figure 1.6a-b**). For human PADI sequence alignments with calcium binding residues annotated see **Appendix 1**. Substitution of just one calcium binding residue has been reported to dramatically reduce the catalytic activity of PADI4 *in vitro* (28). As such, with none of the PADI4 calcium binding sites being fully conserved in PADI6, and nine of the 16 amino acids directly involved in calcium binding in PADI4 not conserved,

it is uncertain whether PADI6 has the capacity to bind calcium. The capacity of PADI6 to bind calcium, has not been investigated.

a



b

```
hPADI6  161 VNCNPADVGQQLEDKKTKKVIFSEEITNL 189    259 VEA 261
hPADI2  153 VNCDRETPWLPKEDCRDEKVYSKEDLKDM 189    252 VEG 254
hPADI4  152 VNCDRDNLESSAMDCEDDEVLDSEDLQDM 180    251 VEA 253
hPADI1  152 VNCDRDNHRSAEPDLTHSWLMSLADLQDM 180    251 VEG 253
hPADI3  152 VNCDRDDPSCDVQDNCDQHVHCLQDLEDM 180    251 VEG 253

hPADI6  357 LQDEMAFCYTQAPHKTTSLILDTPQAADLDEFPMKYSL 410
hPADI2  348 IQDEIEFGYIEAPHKGFPVVLDSPRDGNLKDFPVKELL 401
hPADI4  348 MQDEMEIGYIQAPHKTLPVVFDSPRNRGLKEFPIKRVM 401
hPADI1  349 IQDEMEFGYIEAPHKSFPVVFDSPRNRGLKDFPYKRIL 402
hPADI3  347 IQDEMELGYVQAPHKTLPVVFDSPRNGELQDFPYKRIL 400

hPADI6  410 SPGIMDSIGNLMV 422    678 ANI 680
hPADI2  401 GPDFLDSFGNLEV 413    649 TNV 651
hPADI4  401 GPDFLDSFGNLEV 413    647 TNV 649
hPADI1  402 GPDFLDSFGNLDV 414    647 TNV 649
hPADI3  400 GPDFLDSFGNLEV 412    648 TNV 650
```

Figure 1.6 Human PADI6 calcium binding site conservation with human PADI4
a) Schematic of the calcium binding sites for human PADI4, adapted from (28) to show the corresponding amino acids and general lack of conservation in calcium binding residues in human PADI6. **b)** Human PADI sequence conservation of PADI4 calcium binding residues.

1.2 PADI6 in pre-implantation embryo development

PADI6 is indispensable in early embryonic development in humans and mice (**Figure 1.7a**) (20,32). Numerous *PADI6* variants have been reported in infertile women with the resulting loss of PADI6 function being attributed as the cause of infertility (33–41). *Padi6* knock-out female mice are also infertile, with embryos from these females arresting at the 2-cell stage (32).

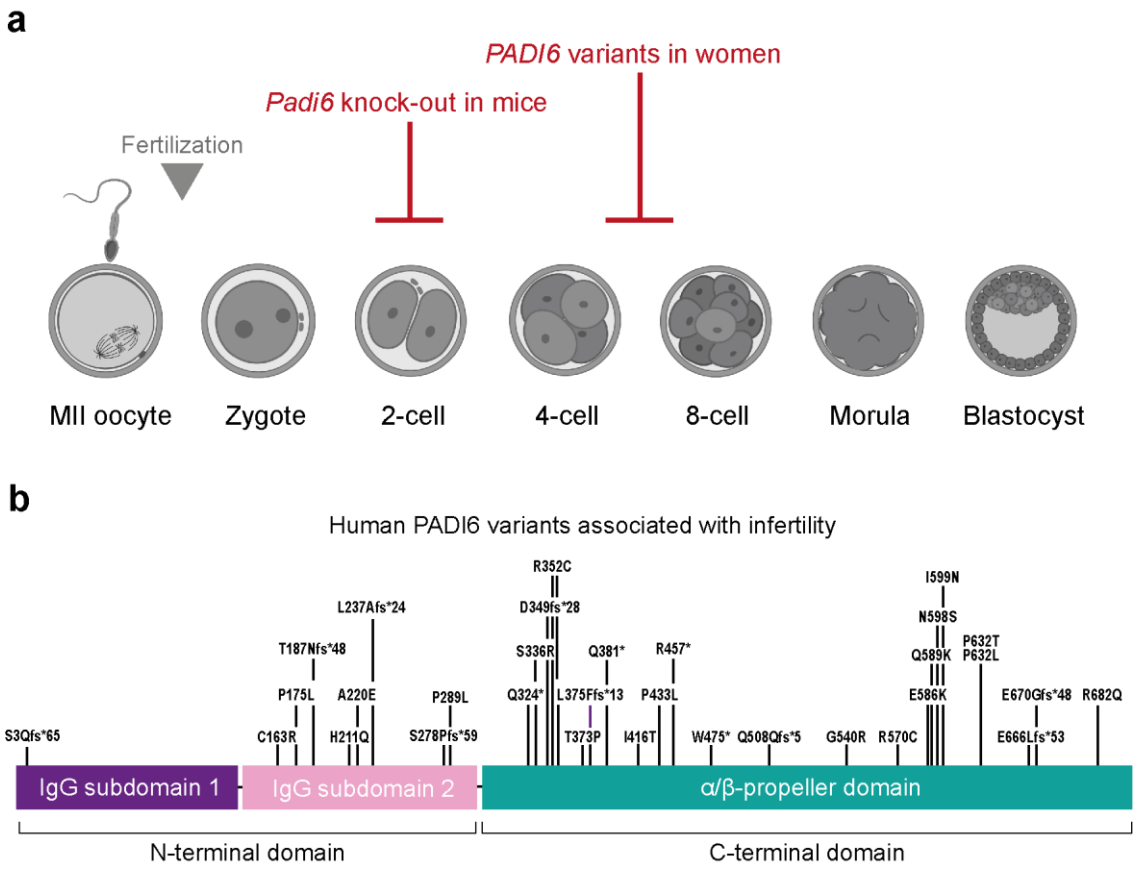


Figure 1.7 Arrest timings of mouse and human embryos lacking PADI6
a) Early embryo development. Embryos from *Padi6* knock-out female mice arrest at the 2-cell stage, embryos from women with loss-of-function *PADI6* variants arrest at the 4- to 8-cell stage. **b)** PADI6 variants that result in human female infertility displayed along the PADI protein domains. fs = frameshift.

Currently, 32 pathogenic *PADI6* variants have been identified in 26 infertile women by whole-exome-sequencing, with the variants being attributed as the cause of infertility (**Figure 1.7b**, **Table 1.3**).

Table 1.3 PADI6 variants that result in early embryonic arrest and female infertility

PADI6 variants	Zygosity	Ref
S3Qfs*65/C666Lfs*53	Compound heterozygous	(34)
C163R	Homozygous	(38)
T187Nfs*48/T187Nfs*48	Homozygous	(34)
H211Q/L237Afs*24	Compound heterozygous	(34)
H211Q/E670Gfs*48	Compound heterozygous	(39)
A220E/A220E	Homozygous	(34)
L237Afs*24/S336R	Compound heterozygous	(34)
S278Pfs*59/S278Pfs*59	Homozygous	(33,36)
P289L/P632L	Compound heterozygous	(42)
Q324*/G540R	Compound heterozygous	(39)
D349fs*28/P433L	Compound heterozygous	(34)
R352C	Homozygous	(34)
T373P/R570C	Compound heterozygous	(36)
L375Ffs*13	Homozygous	(42)
Q381*	Homozygous	(39)
I416T	Homozygous	(34)
R457*	Homozygous	(34)
W475*	Homozygous	(38)
S508Qfs*5	Homozygous	(36)
G540R	Homozygous	(34)
G540R/P632T	Compound heterozygous	(34)
E586K	Homozygous	(34)
Q589K/E670Gfs*48	Compound heterozygous	(34)
I416T / E670Gfs*48	Compound heterozygous	(37)

Embryos from 24 of the infertile women have been reported to predominantly arrest at the 4- to 8-cell stage (34,36,38,39). Of the 32 variants that result in infertility, 13 are predicted to produce a truncated or frameshifted protein and 19 are missense variants. Given the severe phenotypes attributed to these variants, it is likely that they are involved in crucial folding interactions or in key PPIs. Immunofluorescence analysis of oocytes from three of the infertile women possessing homozygous Q381*, compound heterozygous H211Q/E670Gfs*48 or homozygous L375Ffs*13 confirmed the deleterious effect of these variants compared with healthy oocytes (39). Validation of the variant's possible deleterious effects on PADI6 in oocytes or embryos from other women possessing pathogenic *PADI6* variants is not available. In the case of frameshifted and nonsense variants, the result is likely the absence of PADI6 protein in the oocyte and early embryo.

While PADI6 has been implicated in various oocyte and embryonic processes, the function and significance of PADI6 in early embryo development remains unknown. In this section, key stages of oocyte and early embryo development are discussed, and research linking PADI6 to these processes is highlighted.

1.2.1 Oogenesis

The general dogma of developmental biology is that females are born with a store of oocytes (the precursor to the mature egg cell) in their ovaries that are not replenished during their lifetime (43). At birth, these oocytes have already replicated their DNA, completed meiotic DNA recombination, and are arrested in early meiotic prophase (43). The plasma membranes of these immature oocytes are surrounded by a protective cross-linked glycoprotein matrix called the zona pellucida (ZP) and reside in distinct units termed primordial follicles (44–47). Each follicle is comprised of a ZP bound oocyte encased in cumulus cells. The cumulus cells are somatic cells that provide the oocyte with the necessary macromolecule precursors required to grow (**Figure 1.8a**) (47,48). Oocytes are present within the ovary at different stages of growth, and in humans it typically takes oocytes 4-6 months to grow fully to the point where it can resume meiosis. A menstrual cycle occurs approximately once per month, induced by a surge of luteinizing hormone, where typically one fully grown oocyte enters a maturation phase, resumes meiosis, and is released into the oviduct. The periodicity of the reproductive cycle varies between species, along with the number of oocytes that mature and are released for fertilization. For example, the oestrous cycle of mice lasts for 4 to 5 days and results in the release of upwards of 8 eggs depending on the strain (49,50).

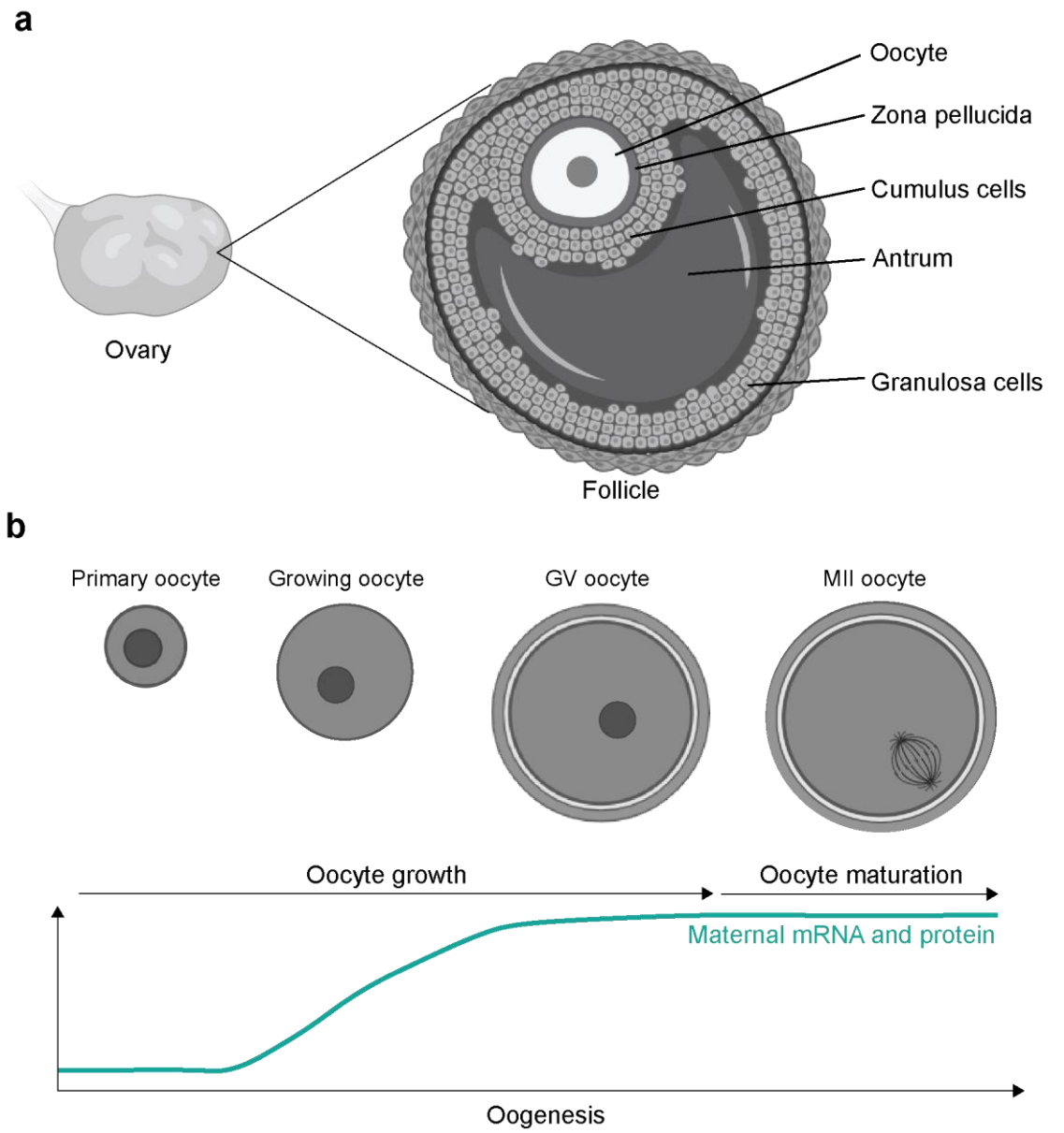


Figure 1.8 Oogenesis overview
a) Primordial follicle (51). **b)** Oocyte growth with increase in maternal mRNA and proteins as the oocyte grows. GV = germinal vesicle. MII = metaphase II arrested. Ovary, follicle and oocyte cartoons produced using BioRender.com.

Oocyte maturation is comprised of an initial growth phase where, within the follicle, the immature oocytes increase in size and accumulate all the necessary proteins and mRNA for the embryo to successfully navigate the early stages of development (**Figure 1.8b**) (43). Upon fertilisation, the embryo will be entirely dependent on these maternally deposited factors as its own genome will not yet be active (52). Many of these maternal factors are encoded by what are called maternal-effect genes. These maternal-effect genes are commonly only expressed in the oocyte and early embryo, with vital functions in oogenesis and early embryo development

(53). Maternal-effect-gene mRNA and proteins are accumulated and stored in the growing oocyte, building up large stores using precursor molecules provided to them by the surrounding cumulus cells. The functions of the maternal-effect-genes are diverse, and for many, poorly understood. Known functions include chromatin remodelling, transcription and mRNA degradation, DNA methylation and maintenance, organelle localisation, and embryonic genome activation (EGA) (52–60). *PADI6* is a maternal-effect gene and is expressed at high levels in the oocyte and early embryo up until the blastocyst stage (32,52,61). Like many other maternal-effect genes, expression of *PADI6* is limited in somatic tissues. Various maternal-effect genes including *PADI6* are not transcribed from the paternal genome, as such *PADI6* loss-of-function variants, or *Padi6* knock-out, only results in female infertility (20,32).

At the point where they are fully grown, the oocytes are termed germinal vesicle (GV) oocytes. They are still arrested in meiotic prophase, are significantly larger than immature oocytes and have a large and clearly visible nucleus in their centre (43,51). This growth phase is then followed by one meiotic division (meiosis I), before the oocyte arrests in the metaphase II stage of a second meiotic divison (meiosis II) (62). It is at this stage, named the metaphase II oocyte (MII oocyte), that the oocyte is mature and ready for fertilisation. During maturation, a mucified matrix develops and expands between the follicular cells, eventually rupturing the follicle, allowing the now mature oocyte to be released into the oviduct (43,63). It is here in the oviduct (fallopian tube in humans), that the mature MII oocyte can be fertilised by a sperm (**Figure 1.9**).

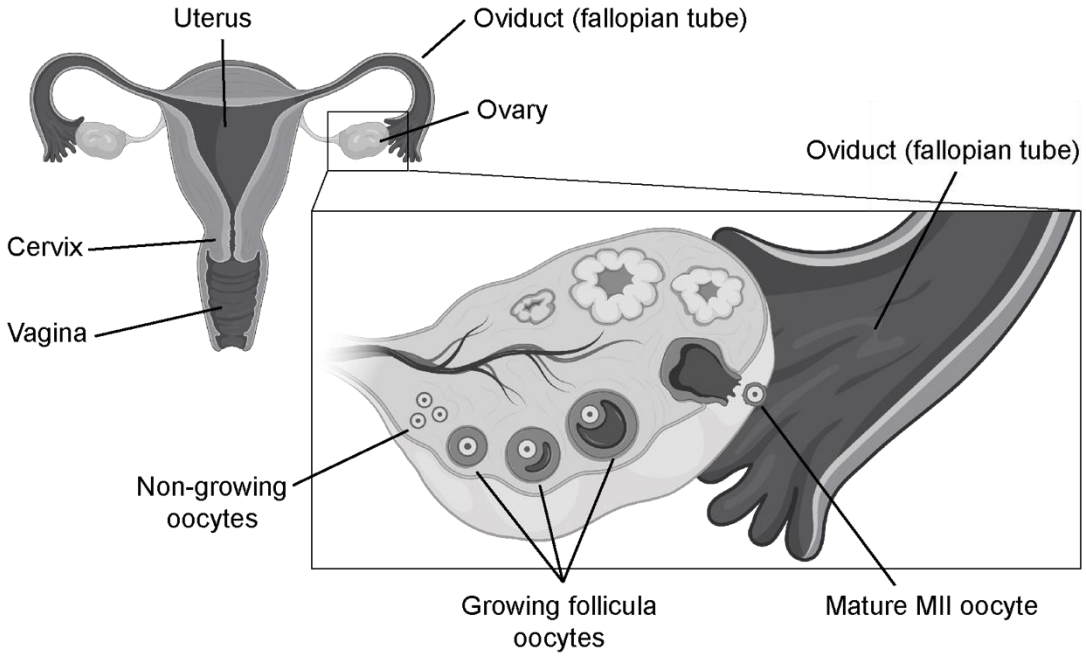


Figure 1.9 Ovulation
Human female reproductive system. Insert: ovary at various stages of oocyte growth and finishing in the ovulation of a mature MII oocyte into the oviduct (fallopian tube in humans). Cartoons produced using BioRender.com.

Despite high levels of PADI6 in the mouse oocyte, oogenesis appears to proceed normally in *Padi6* knock-out oocytes (32). The major difference from wt oocytes is that *Padi6* knock-out oocytes lack a structure unique to oocytes called the cytoplasmic lattices (CPLs).

Early work on PADI6 and *the cytoplasmic lattices*

The CPLs are a highly abundant cytoplasmic structure unique to the oocyte and early embryo, first identified in the 1960s (64–66). Until late 2023 they could only be visualised by electron microscopy (EM) (67). The composition and function of the CPLs has remained mostly unknown, with many theories posed over the years as to what they truly are. Visually, they appear as stacks of short fibres, distributed uniformly throughout the oocyte and early embryo (**Figure 1.14**) (67).

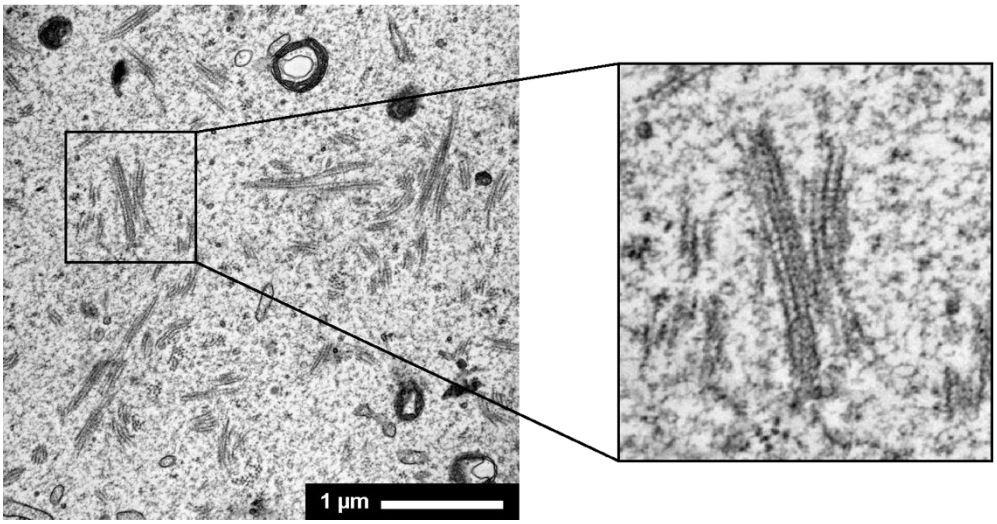


Figure 1.10 Cytoplasmic lattices
Transmission electron microscopy (TEM) image of CPLs in a mouse GV oocyte. Insert magnified view of representative lattices. Image taken as part of this project by Dr. Anne Weston of the Crick EM STP.

A link between PADI6 and the CPLs was first described by Coonrod and co-workers in 2003 when PADI6 appeared to co-localise with the CPLs in mouse oocytes by gold immune-electron microscopy (11). The next development came when Gossen and co-workers discovered that the CPLs were absent in oocytes from *Padi6* knock-out females, suggesting PADI6 was crucial for their formation or maintenance (32). This work was followed up in 2008 by Coonrod and co-workers who set out to characterise the function and composition of the CPLs using PADI6-deficient embryos as a model (68). It had previously been suggested that the CPLs were a storage site for maternal ribosomes (69). Oocyte lysate centrifugation experiments, combined with immunoblotting and TEM, indeed confirmed that maternal ribosomes were associated with a large supramolecular complex (68). This association was disrupted in PADI6-deficient oocytes although overall ribosomal levels were unchanged in these oocytes. It was therefore proposed that PADI6 is involved in the regulation and storage of maternal ribosomes, likely at the CPLs, and not in their production or maintenance.

Coonrod and co-workers published again on the relationship between PADI6 and the CPLs in 2011 (70). It was observed that during detergent extraction, CPL-associated components were not extracted. Thus, by subjecting both wild-type (WT) and PADI6-deficient oocytes to detergent extraction, targets that were specifically associated with the CPLs could be identified. Using this method, combined with immunofluorescence and immunoblotting, they identified α-tubulin as a CPL associated

protein. Interestingly, α-tubulin was co-immunoprecipitated with PADI6, suggesting an interaction with PADI6. Whether this interaction is direct or through the CPLs was not characterised. It was further shown that mitochondria and endoplasmic reticulum (ER) showed defects in their positioning and reorganisation in PADI6-deficient oocytes. Throughout oocyte maturation, the mitochondria and ER, along with other organelles, undergo dramatic reorganisation, mediated by microtubules (MTs) (71,72). This organelle reorganisation appeared defective in the absence of PADI6 (70). Furthermore, the levels of acetylated α-tubulin appeared to be reduced in GV and MII oocytes from *Padi6* knock-out females. Acetylated-α-tubulin has previously been reported to be important in MT stabilization, and involved in the trafficking of the ER, and mitochondria in somatic cells (73,74). Together these data led to the proposal that PADI6 and the CPLs are involved in proper MT-mediated organelle localisation and distribution during oogenesis. Further work by Coonrod and co-workers on the CPLs and PADI6 was published in 2017, where they reported that mRNA as well as mRNA binding protein Y-box-binding protein 2 (MSY2) were also associated with the CPLs (75). Furthermore, MSY2 appeared to localise with PADI6, and its localisation was disrupted in oocytes from *Padi6* knock-out females. MSY2 is a germ cell specific mRNA binding protein involved in the stability and translation repression of maternal mRNA during oogenesis in mice (76–78). Together, these works have led to the theory that the CPLs act as a hub to store and regulate maternal ribosomes and mRNA. During EGA, the translation and degradation of maternal factors must be tightly controlled, both temporally and spatially. It therefore follows that there would exist a specialised module to house maternal ribosomes and mRNA, possibly the CPLs.

The subcortical maternal complex

Very similar phenotypes to *Padi6* knock-out, including absence of the CPLs and organelle localisation disruption have been reported in knock-out models of proteins belonging to the subcortical maternal complex (SCMC). The SCMC is a complex of three core proteins (OOEP, TLE6, NALP5) and a possible four further non-core proteins (KHDC3L, ZBED3, NALP2, NALP7), encoded by maternal-effect genes (79–83). The complex was so named as it appeared to localise to the subcortex of early embryos, however this observation has recently been disproven and the SCMC appears to be distributed throughout the cytoplasm (67). The complex is conserved in both mice and humans, with the majority of experimental work characterising the

complex being performed in mouse oocytes and embryos (80). In 2008, immunoprecipitation of OOEP from mouse embryos followed by shotgun mass spectrometry identified PADI6 in the immunoprecipitated fraction (83). Since then, it has been widely theorised that PADI6 is also part of the SCMC, however, at the start of this work no experiments had been reported directly characterising an interaction between the two.

It has been demonstrated that like PADI6, the SCMC proteins are vital for early embryonic development. Loss of any of the core member proteins results in similar phenotypes to embryos from *Padi6* knock-out females (**Table 1.4**). Crucially, like PADI6, absence of core SCMC proteins, OOEP, TLE6 or NLRP5, results in a complete absence of the CPLs and defective organelle distribution (67,84–86). These observations suggest that the SCMC is also required for proper CPL formation or maintenance, and this function is likely tightly linked with PADI6. It has also been reported that the SCMC is required for the formation of the oocyte and embryo F-actin meshwork through Cofilin, a key protein involved in the assembly of F-actin (87). The F-actin meshwork is required for proper spindle positioning and as such is crucial for symmetric division in the zygote. As such, absence of SCMC proteins also results in defective spindle positioning and dramatically asymmetric mitotic division.

Table 1.4 SCMC genes and mouse knock-out phenotypes

Human gene (protein)	Mouse gene names	Mouse knock-out phenotype	Ref
OOEP (OOEP)	*Mater*	2-cell arrest	(84)
TLE6 (TLE6)	*Filia*	2-cell arrest	(87)
NLRP5 (NALP5)	*Floped*	2-cell arrest	(55)
KHDC3L (KHDC3L)	*Tle6*	Reduced fecundity	(88)
ZBED3 (ZBED3)	*Zbed3*	Reduced fecundity	(82)
NLRP2 (NALP2)	*Nlrp2*	Subfertility	(81)
NLRP7 (NALP7)	*Nlrp7*	2-cell arrest	(89)

The phenotypes of oocytes and embryos from *Padi6* and SCMC knock-out females share many similarities. Furthermore, like PADI6, numerous SCMC gene variants have been reported as causes of infertility in women (42,90,91). The exact functional relationship between PADI6 and the SCMC has remained unknown up until late 2023, when work by Schuh and co-workers showed they both come together to

form the CPLs in mouse oocytes (67). This work is described further in the following section.

The latest model of the CPLs, the SCMC and PADI6

Based on the previously described work, in 2017 the widely accepted theory in the field was that the CPLs act as a storage hub for maternal ribosomes and mRNA, are involved in the redistribution of organelles, and the proper localisation of the mitotic spindle. Whilst many proteins have now been reported to be associated with the CPLs (for example α-tubulin and MSY2), and others also required for their formation (PADI6 and various SCMC proteins), the exact composition and structure of the CPLs remained unknown. In 2023 however, an extensive characterisation of the CPLs was published by Schuh and co-workers that disputed many of the previously held theories on the composition and function of the CPLs (67).

Firstly, Schuh and co-workers confirmed that PADI6 and the SCMC proteins were not localised to the subcortex of the embryo as previously reported but distributed throughout the cytoplasm (67). PADI6 and the SCMC did not appear uniformly distributed however but localised in distinct structures throughout the oocyte. By using expansion microscopy, they further confirmed that PADI6, along with SCMC proteins TLE6, NLRP5, OOEP, and KHDC3L were arranged into protein fibres, of similar dimensions to the CPLs. Through comparisons with *Padi6* and *Tle6* knock-out oocytes, it was confirmed that the CPLs are indeed composed of fibres of PADI6 and the SCMC proteins. This is the first experimental evidence that the SCMC and PADI6 are present within the CPLs, and not just associated with them. The organisation and structure of the lattices was further elucidated by electron tomography: the CPLs were composed of twisted protein filaments, grouped together with defined interaction points between them (**Figure 1.15**). Interestingly, the structure and size of these filaments ruled out the previously held view that the CPLs contain ribosomes. Using proteomics, the function of the CPLs was then interrogated, and it was found that in oocytes lacking PADI6 or TLE6, and therefore lacking CPLs, many proteins, predominantly maternal factors, were down-regulated. Various down-regulated proteins were characterised and appeared to be enriched on the CPLs, leading to the theory that the CPLs act as a storage site not for maternal mRNA, but for maternal proteins. The depleted proteins in question include actin and microtubule associated proteins, consistent with previously observed phenotypes. On top of these, maternal-effect-genes, and proteins involved in

the cell cycle, chromatin modification, protein ubiquitination and the ribosome were also down-regulated.

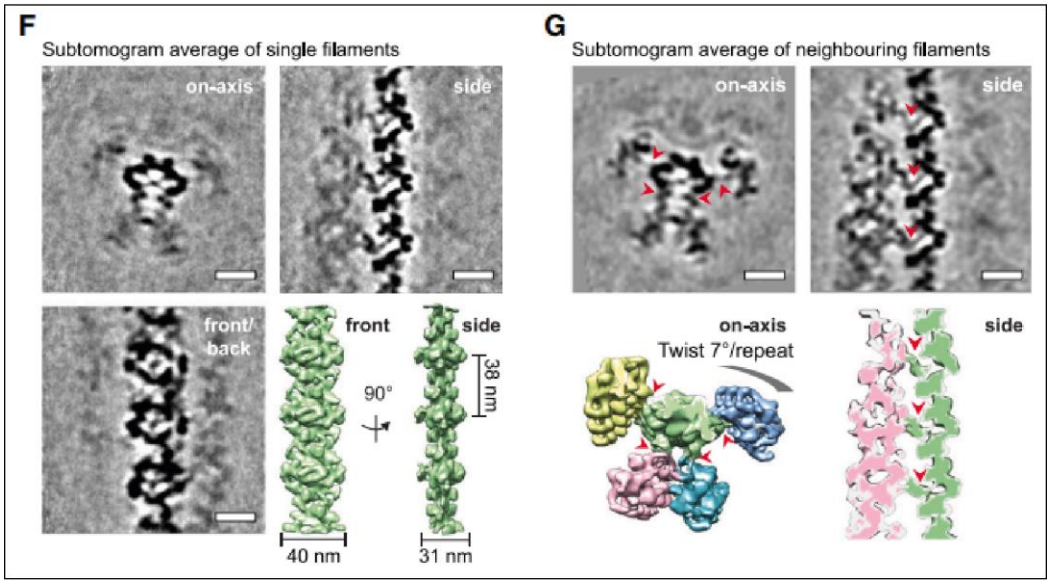


Figure 1.11 Electron tomography of CPL structure
Subtomogram average of cytoplasmic lattice filaments with a mask focusing the alignment on a single filament (**F**) and neighbouring filaments (**G**) with isosurface rendering of the front and side views. Figure reproduced from (67) with permission (**Appendix 20**).

The exact protein composition of the CPLs, and how they come together to form such ordered and abundant structures in the oocyte is unknown. The CPLs are absent when PADI6, and various SCMC genes are knocked-out out suggesting they are crucial structural components in the lattices. Ubiquitin-like with PHD and RING finger domains 1 (UHRF1) was identified as another part of the lattices in 2023 by Sasaki and co-workers (92). Oocyte specific *Uhrf1* knock-out disrupts CPL formation and results in a down-regulation of proteins independent of transcriptomic down-regulation. Furthermore, in the *Uhrf1* knock-out oocytes, both the mitochondria and ER were mis-localised, and the localisation of SCMC proteins NLRP5 (NALP5 in humans) and OOEP were disrupted. Together, these data suggests that UHRF1 may be a new integral component of the CPLs in mice.

Interestingly, UHRF1 was down-regulated by almost 90% in *Padi6* knock-out oocytes, but not *Tle6* knock-out oocytes in the 2023 work by Schuh and co-workers (67). UHRF1 is an important co-factor of DNA methyltransferase 1 (DNMT1) involved in maintaining DNA methylation during replication (93). The DNMTs, UHRF1 and further data suggesting a potential relationship between UHRF1 and PADI6 is discussed in **Chapter 1.3.**

Whether the CPLs are also present in human oocytes is an open question. Using a similar experimental set up to mouse oocytes, Schuh and co-workers showed that human oocytes do possess clusters of PADI6 and SCMC proteins distributed ubiquitously throughout the cytoplasm (67). The clusters did not possess the same fibrillar structure as mouse CPLs however and appeared more mesh-like. Whether these are indeed the same structures as the mouse CPLs is unknown, but these data do confirm that PADI6, SCMC and UHRF1 co-localise together in human oocytes in distinct structures.

1.2.2 Fertilisation and zygotic genome activation

Fertilisation is the fusion of sperm and egg, initiated by the binding of the sperm cell to the ZP of the mature MII oocyte in the oviduct (43). Upon binding, the sperm releases the acrosome, a secretory vesicle containing hydrolytic and proteolytic enzymes (94). Through a combination of ZP hydrolysis by the acrosome-released enzymes and sperm motility, the sperm penetrates the ZP and fuses with the plasma membrane of the egg. Sperm binding causes a release of Ca^{2+} from the ER, which triggers a cascade of events, including the completion of the second meiosis, and translation of maternal mRNAs (95–99). These processes are the first steps in the embryo completing the maternal-to-zygotic transition (MZT). The MZT involves a series of changes the embryo must successfully complete to transition from meiotic to mitotic division, including:

- Replacement of protamines with nucleosomes in the paternal genome (100,101)
- Global demethylation of the paternal genome (102,103)
- Substitution of the meiotic cohesion complex with the mitotic (104,105)
- Transition from acentrosomal to centrosomal spindle assembly (106,107)
- Transition from asymmetric to symmetric spindle positioning (108)
- Maternal factor degradation (106,107)

The large transient release of Ca^{2+} ions triggered by fertilisation could be a potential evolutionary explanation for the lack of conservation in calcium binding residues in PADI6. For the activity of PADI6 not to be affected by this calcium transient, then it makes sense that PADI6 would genetically diverge so as not to be activated by

calcium. The calcium binding capacity of PADI6 remains unknown, however, and remains a key question that must be answered.

One of the most important transitions the embryo must navigate during MZT, is the activation of its own genome, in a process known as embryonic genome activation (EGA) (111–113). After fertilization, the zygote's genome is transcriptionally silent. If it is to survive, it must progress through a programme of highly regulated chromatin modification, and transcriptional activation. EGA occurs at different times in mice and humans and is comprised of two phases: an initial minor wave of transcription (minor EGA), followed by a significant major wave in transcription (major EGA) (**Figure 1.12**) (113). In mice, minor EGA starts in the zygote, with major EGA occurring during the 2-cell stage (114). In humans, minor EGA occurs around the 4-cell stage with major EGA at the 8-cell stage. The coincidence of PADI6 attributed early developmental arrest in mice (2-cell stage) and humans (4 to 8-cell stage), with the timings of EGA in each species, has led to speculation that PADI6 has a vital function in EGA (20).

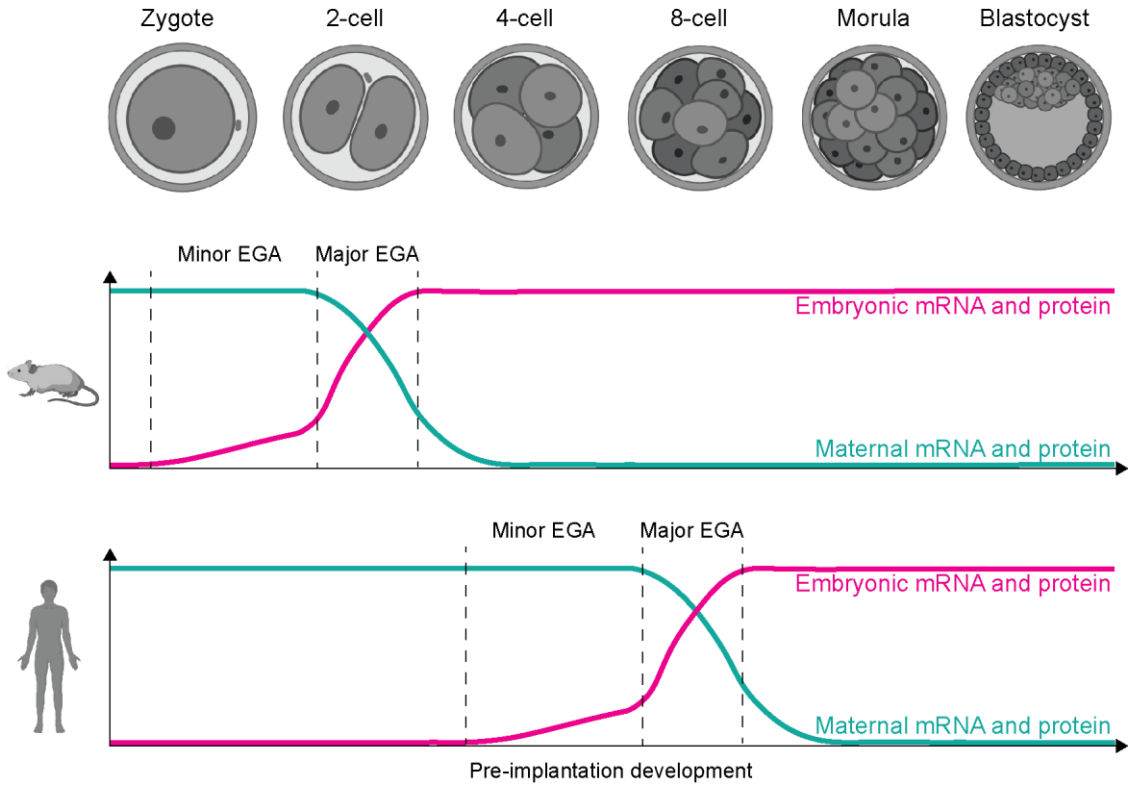


Figure 1.12 EGA timings in mouse and human pre-implantation embryos
Schematic of mammalian preimplantation embryo development with changes in maternal and embryonic mRNA levels. The timing of minor and major EGA in mice and humans are indicated. Embryo cartoons produced using BioRender.com.

Our understanding of the mechanisms regulating EGA are incomplete and remain a highly researched topic. However, there are two key known mechanisms involved in ensuring successful EGA: chromatin remodelling and specific transcriptional activation (113,115–126).

Chromatin remodelling

Chromatin modifications are one of the mechanisms through which gene transcription is regulated. Given the highly regulated wave of gene transcription occurring during EGA, it has therefore been proposed that chromatin remodelling through chemical modifications is a key factor in this regulated genomic activation (113). Studying chromatin remodelling in human embryos is challenging however, and as such, it has primarily been studied in mice. Various mechanisms have been reported as to how chromatin modifications are a key factor in EGA, with certain mechanisms appearing to be conserved between different species. These mechanisms include histone modifications, and nucleosome composition and positioning.

It has been shown that chromatin, specifically histone modifications, undergo dramatic changes during the early stages of MZT (127). Histone modifications have the capacity to dramatically alter gene accessibility, and by extension expression (128). As such, it follows that histone modifications would be a key mechanism in which the early embryo regulates and activates its own genome in a controlled manner. Chromatin immunoprecipitation sequencing-based (ChIP-seq) experiments have been conducted in mouse embryos, highlighting that histone 3 lysine 27 acetylation (H3K27ac) and histone 3 lysine 4 trimethylation (H3K4me3) could both have important functions in activating the mouse zygotic genome (115–118,129). Histone acetylation is known to promote chromatin accessibility and active transcription, and increases during EGA, in particular at actively transcribed EGA genes (113). In mice, this has been confirmed for H3K27ac and is a conserved mechanism in multiple species including zebrafish (130). H3K4me3 on the other hand is often found at the transcription start site (TSS) of genes and like histone acetylation, is associated with active transcription (131). The H3K4me3 signature of early mouse embryos undergoes a dramatic reorganisation, moving from broad domains to narrow regions of H3K4 trimethylation at TSSs (115,118,119). This reorganisation requires transcription, and it has been suggested that the broad H3K4me3 domains present before EGA are important in preventing premature transcription. Whilst H3K4me3 is deposited on the mouse genome in the

oocyte, these marks only appear to occur after fertilization in humans, suggesting the function of chromatin modification during EGA is not conserved between the two (115,132).

Chromatin can also be remodelled through altering the composition of nucleosomes (133). This is indeed the case in mouse EGA, where embryonic variants of linker histone H1 are substituted for somatic variants during major EGA (120,134). This substitution is conserved in flies and frogs and has been demonstrated to be vital for the proper EGA (135). The positioning of nucleosomes has also been shown to be an important modulator during genome activation (116,117). Observed in mice and humans, as well flies and zebrafish, nucleosomes form organised arrays and mark gene promoters with open chromatin regions during EGA (121,122).

Transcriptional activation

It has been estimated that approximately 12-15% of the genome is transcriptionally activated during EGA (114). As such, the specific binding and activation of transcription factors plays a vital role in the directed transcription of EGA genes. These transcription factors differ between species, for example POU domain, class 5, transcription factor 1 (OCT4) is vital for EGA in humans but not mice (113,123,136,137). One family that is conserved is the double homeobox protein (DUX) transcription factors (*Dux* in mice, *DUX4* in humans) which activates hundreds of genes during EGA (124,125). It has been proposed that these EGA transcription activators are expressed during the minor EGA wave and go on to activate other genes during the major wave, although the mechanism behind their initial transcription is unknown. Included within the genes transcribed in the major wave are many endogenous retroviral elements, notably MERVL and HERVL in mice and humans respectively (126). The activation and transcription of MERVL has been shown to be vital for EGA in mice, and its knock-down results in early embryonic developmental arrest (138). It is likely that many other transcription factors are required for the controlled activation of the embryo's genome. Given the lack of conservation between factors, the regulatory mechanisms in both mice and humans need to be elucidated.

PADI6 and EGA

The coincidence of reported arrest points (2-cell in mice, 4 to 8-cell in humans) with the timing of major EGA in humans and mice has led to speculation that PADI6 is

involved in genome activation (20). EGA is characterised by the highly regulated and significant wave of transcription and translation of early embryonic proteins, as well as the degradation of maternal factors. As such, the transcription and translation of the embryo globally, and of known early embryonic factors, can be used to characterise EGA in the early embryo. Furthermore, the level of phosphorylated RNA Pol II can be used as a proxy for measuring active transcription in a cell (139). The levels of both RNA Pol II and phosphorylated RNA Pol II are reduced by approximately 20% and 30% respectively in *Padi6* knock-out embryos compared to WT (68). It was also observed that RNA Pol II appeared to aggregate around the nuclear envelope in the knock-out embryos. A similar decrease in phosphorylated RNA Pol II staining is seen in arrested 4-cell embryos from women possessing either the homozygous Q381* or compound heterozygous H211Q/E670Gfs*48 PADI6 variants (39). Histone H4 acetylation, another marker of transcription, was also reduced by approximately 10% in 2-cell embryos from *Padi6* knock-out females (68). Looking at the translation of early embryonic factors, levels of the transcription requiring complex (TRC), one of the first protein complexes expressed from the early embryo's genome, were also reduced in 2-cell embryos from *Padi6* knock-out females (68,140). These observations suggest that PADI6 is important for transcription and/or translation of early embryonic factors during EGA, potentially through the nuclear translocation of RNA Pol II and/or by regulating its phosphorylation to the actively transcribing form. Although, it must be noted that these observations could be the result of upstream defects caused by the absence of PADI6 and CPLs in the oocyte, and PADI6 may not have a direct function in EGA.

1.2.3 Gastrulation

After successfully completing the MZT and activating its own genome, the totipotent mammalian embryo proceeds through a series of cleavage stages leading up to the first point of cellular differentiation, the blastocyst (**Figure 1.13a**) (141). Without increasing in overall volume, the embryo transitions through synchronised cleavage stages up to the point of 16-cells where the embryo compacts down, and cell-cell contacts disappear at the morula stage. It is at this stage that the first different cell populations are established (142). Driven by polarity and cell-cell contacts, the inner and outer cells start to differentiate from one another (**Figure 1.13b**). In a comprehensive study by Niakan and co-workers, it was shown that atypical protein kinase C (aPKC) in outer cells blocks angiomotin (AMOT) activation of large tumour suppressor kinase 2 (LATS2) (142). This in turn prevents LATS2's subsequent

phosphorylation of yes-associated protein 1 (YAP1), keeping the Hippo signalling pathway inactive. The un-phosphorylated YAP1 can then move into the nucleus along with its transcription factor partner TEAD4 and activate trophectoderm (TE) associated transcription factors including homeobox protein CDX-2 (CDX2) and GATA binding protein 3 (GATA3). This is the first event in lineage differentiation, with the outer cells of the morula going on to form the blastocyst TE, with the inner cells going on to form the blastocyst inner cell mass (ICM). It was also shown that this pathway is conserved between mice and humans. After this first point of differentiation, a fluid filled cavity begins to form, termed the blastocyst cavity. This cavity continues to expand to form the blastocyst, which is composed of an outer single layer of TE cells, with the ICM, a small cluster of cells attached to one end (**Figure 1.13c**). Later in development, the TE goes on to form the placenta while the ICM becomes the embryo proper. Once fully formed, the blastocyst hatches from the ZP and implants in the uterine wall marking the end of pre-implantation development.

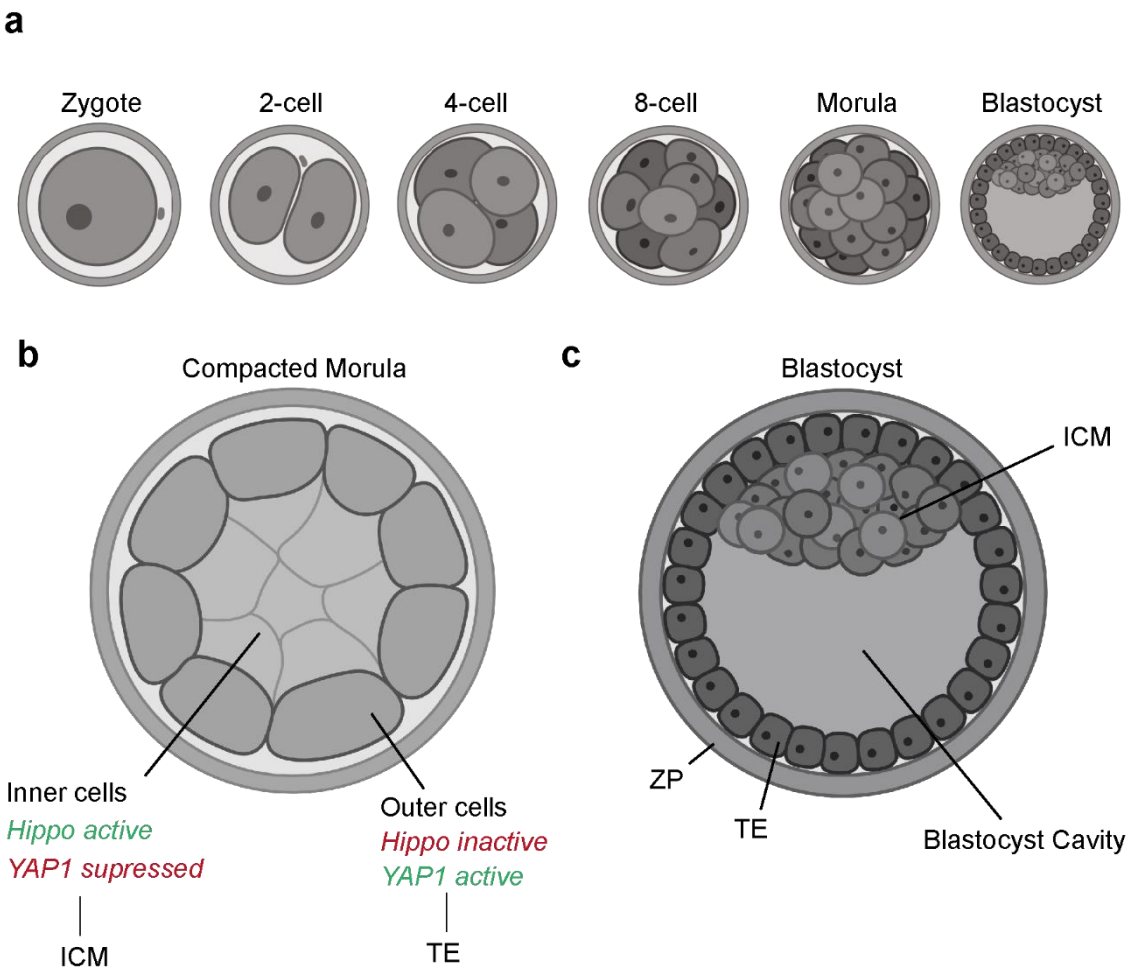


Figure 1.13 Mammalian pre-implantation embryo development
a) Mammalian pre-implantation embryo development. **b)** Cartoon of a compacted morula highlighting inner and outer cells which go on to form the inner cell mass (ICM) and trophectoderm (TE) of the blastocyst respectively. **c)** Cartoon a blastocyst with ICM, zona pellucida (ZP), TE, and blastocyst cavity labelled. Embryo cartoons produced using BioRender.com.

Interestingly, two of the 26 infertile women were noted to have extended pregnancies that ended prematurely in miscarriage or the formation of a hydatidiform mole (HM) (**Table 1.5**) (35,41). HMs, sometimes called a molar pregnancy, are characterised by a cancerous growth in the uterus during pregnancy caused by hyperproliferation of the trophectoderm, often as a result of a defect in the fertilised egg cell (143,144). The molecular mechanisms that result in the formation of a HM are poorly understood. The difference in *PADI6* variant phenotypes observed with the pregnancies of these women has been attributed to the variants resulting in single amino acid changes which conceivably negatively impact protein function without ablating it completely. It is also unknown whether the PADI6-attributed formation of

HMs is a result of a late stage PADI6 function, or upstream defects in the oocyte and early embryo.

Table 1.5 PADI6 variants resulting in late-stage pregnancy complications

PADI6 variants	Zygosity	Phenotype	Ref
N598S/R682Q	Compound heterozygous	6 miscarriages, 1 hydatidiform mole	(35)
I599N	Homozygous	3 miscarriages, 4 hydatidiform moles	(41)

1.2.4 PADIs in the early embryo

Characteristic of a maternal-effect gene being degraded post-EGA, scRNA-seq experiments have shown that *Padi6* mRNA is abundant in the MII oocyte and zygote but drops off at the 2-cell stage in mouse embryos (61). Mouse PADI6 protein, however, remained abundant until the blastocyst stage in a large-scale comparative proteomic experiment conducted by Gao and co-workers, and appeared to be one of the most abundant proteins in the mouse early embryo (61). Low input ribosome sequencing (LiRibo-seq) of MII oocytes and early embryos by Zhang and co-workers showed two *Padi6* translational spikes, once at the MII oocyte stage and the next at the 4-cell embryo stage (145). This second spike at the 4-cell stage raises the question of whether PADI6 has an additional function later in development beyond the oocyte, zygote and 2-cell embryo. Furthermore, as EGA will be complete by this point, could *Padi6* be being transcribed from the embryonic genome, and not the maternal mRNA store as the majority of maternal mRNA will have been degraded by this stage?

PADIs 1 to 4 on the other hand, appear to be negligibly expressed in the mouse oocyte and early embryo. In the large-scale proteomic profiling of mouse early embryo development conducted by Gao and co-workers, no peptides from PADIs 1 to 4 were identified (61). Furthermore, both scRNA-seq experiments, and the LiRibo-seq experiment conducted by Zhang and co-workers showed negligible RPKM values for PADIs 1 to 4 (145). Together these data suggest that the mRNA and protein levels of PADIs 1 to 4 is negligible in the mouse oocyte and early embryo. Contrastingly, it has been reported that PADIs 1 to 4 are present in the early embryo shown using immunofluorescence experiments (146).

1.2.5 Citrullination in the early embryo

Detecting citrullinated proteins in a cellular context is challenging. Common techniques used to study PTMs, such as mass spectrometry, are impeded by the 0.98 Dalton mass difference between citrulline and arginine (147). Furthermore, the similarity between the two residues makes it difficult to selectively raise antibodies to recognise citrulline, and as such, few selective citrulline antibodies are available. The antibodies that are available do not recognise citrullination in a site-agnostic manner but are raised to recognise specific citrullinated sequences. As such, the identification of novel citrullinated sites remains an ongoing challenge to the field and one that needs to be addressed if we are to fully elucidate the function of PADIs.

Data regarding the possible citrullination of proteins in the early embryo is inconsistent. Immunohistochemical staining of wild-type (WT) and *Padi6* knock-out ovaries with an antibody raised against a citrullinated histone 4 peptide (H4Cit3) showed a decrease in immunofluorescence intensity in the oocytes from *Padi6* knock-out ovaries versus WT (32). Immunofluorescence experiments in a later work however showed no difference in immunofluorescence intensity between WT and *Padi6* knock-out oocytes stained with antibodies to a range of citrullinated histone sequences, including H4Cit3 (148). The inconsistency in these experiments exemplifies the limitations of antibodies raised against citrullinated protein sequences and fails to conclusively answer whether protein citrullination occurs in the oocyte.

The use of PADI inhibitors is commonly used either in parallel, or in conjunction with, immunofluorescence experiments to interrogate either citrullination or protein phenotypes in the absence of PADI activity (1). The most commonly used PADI inhibitor is Cl-Amidine (**Figure 1.14**), which covalently inhibits PADIs 1 to 4 with similar efficacies (149). The development of specific PADI inhibitors is a highly researched field, out of which has come D-Cl-Amidine, a derivative of Cl-Amidine that specifically inhibits PADI1 (150). Zygotes treated with 100 μM D-Cl-Amidine, followed by staining with antibodies to H4Cit3 and H3Cit2+8+17 at the 4-cell stage showed decreased nuclear staining (146). Furthermore, none of the D-Cl-Amidine-treated zygotes successfully reached the morula stage compared to 75% of vehicle-treated zygotes; experiments have been conducted with the pan-PADI inhibitor Cl-Amidine, displaying similar results (148). There is currently no reported data describing whether Cl-Amidine, or D-Cl-Amidine, covalently bind to PADI6, and as PADI6 currently has no reported

catalytic activity, it has not been possible confirm whether these molecules have any inhibitory activity. As discussed in **Chapter 1.2.4**, PADIs 1 to 4 appear to be absent in the oocyte and early embryo in transcriptomic and proteomic analysis, which raises the question of why these phenotypes are observed upon treatment with a PADI inhibitor. Given the low specificity of Cl-Amidine, and other PADI inhibitors, it is conceivable that the observed effects in the early embryo upon treatment are due to inhibition of PADI6. On the other hand, as Cl-Amidine based inhibitors have a propensity to react with nucleophilic cysteine residues, it is also plausible that the observed phenotypes could be due to off-target reactions, ultimately resulting in a reduced developmental capacity (150,151). As the above-described data is inconsistent, further work is required to elucidate whether proteins are citrullinated in the early embryo and whether this citrullination is catalysed by PADI6.

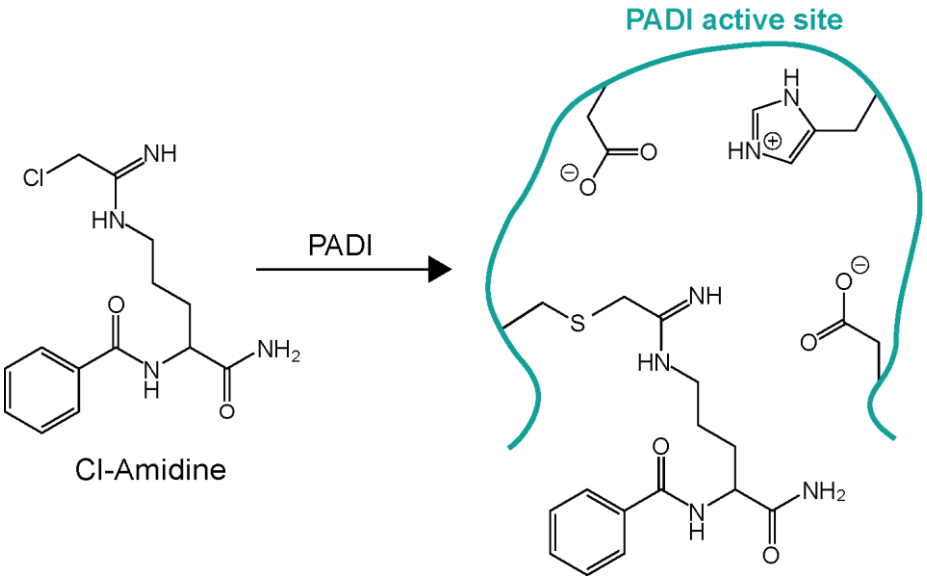


Figure 1.14 Cl-Amidine structure and PADI inhibition
Structure of pan-PADI inhibitor Cl-Amidine and PADI inhibition mechanism.

1.2.6 Is PADI6 regulated by 14-3-3 proteins in the early embryo?

It has been reported that PADI6 is not phosphorylated in the immature oocyte but is phosphorylated sometime during mouse oocyte maturation and remains phosphorylated in the mature egg and 2-cell embryo (152). While the specific phosphorylation sites have not been elucidated, this phosphorylated PADI6 was shown to interact with 14-3-3 proteins in the mature egg and 2-cell embryo. 14-3-3 proteins are a family of seven highly conserved proteins that can regulate target proteins

through interaction with specific phosphorylated sequences (ArgXXpSer or ArgXXXpSer motifs) (153–155). Following on from this work, Rose *et al.* used an *in silico* workflow to identify possible 14-3-3 binding sites in human PADI6 (156). Two canonical and six non-canonical possible sites were identified, of which the recombinant phosphorylated peptides of two, centred at Ser10 (canonical) and Ser446 (non-canonical), were shown to interact with 14-3-3σ using fluorescence polarization assays. The interaction of these two peptides with 14-3-3σ was then characterised structurally by X-ray crystallography, with both docking in the expected pocket. Whilst it appears that PADI6 may be phosphorylated and regulated by 14-3-3 proteins *in vivo*, the exact phosphorylation sites, the relevant kinases and the overall function of this regulation are unknown.

1.3 PADI6, DNA methylation and maternal imprinting

As well as the 32 pathogenic *PADI6* variants reported to cause infertility, a further 14 variants in nine fertile women have been reported (**Figure 1.15, Table 1.6**) (20). Whilst these women have been able to have children, it is often with a high occurrence of pregnancy complications such as recurrent miscarriages (35,157). Furthermore, children from these women are most often born with multi-locus imprinting disorders (MLID) (**Table 1.6**). Common MLID diagnosed in children from mothers carrying *PADI6* variants include Beckwith-Wiedemann syndrome (BWS), Silver-Russell syndrome (SRS) and Temple syndrome (TS) (158). As MLIDs are commonly caused by *trans*-acting factor variants, it has been suggested that PADI6 belongs to this class of imprinting factors and has an important function in either the establishment or maintenance of the maternally imprinted signature (158–160). Before discussing imprinting and MLID, DNA methylation in the oocyte and early embryo must be discussed.

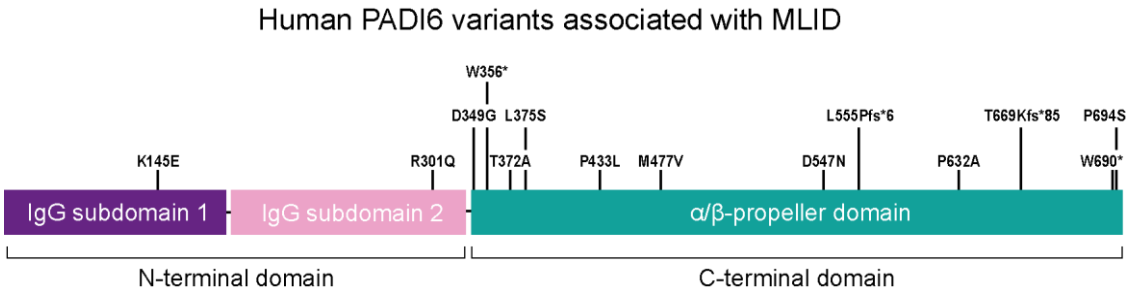


Figure 1.15 PADI6 variants associated with MLID
PADI6 variants that are attributed to MLID in offspring displayed along the PADI protein domains.

Table 1.6 PADI6 variants that cause MLID
BWS = Beckwith-Wiedemann syndrome, SRS = Silver-Russell syndrome, TS = Temple syndrome.

PADI6 variants	Zygosity	Phenotype	Ref
K145E/wt	Heterozygous	1 child with SRS	(159)
R301Q/P433L	Compound heterozygous	1 child with SRS and TS	(161)
D349G/wt	Heterozygous	1 child with SRS, 2 children with no MLID	(159,161)
W356*/P632A	Compound heterozygous	2 children with BWS	(162)
T372A/W690*	Compound heterozygous	2 miscarriages, 1 premature birth, 2 children with BWS	(157)
L375S/D547N	Compound heterozygous	1 child with BWS, reproductive history not reported	(161)
M477V/P694S	Compound heterozygous	1 miscarriage, 1 child with BWS, 1 child with no BWS	(162)
D547N/L555Pfs*6	Compound heterozygous	9 miscarriages, 1 child with BWS	(163)
T669Kfs*85/wt	Heterozygous	1 child with BWS, 2 children with no BWS	(159,162)

1.3.1 DNA methylation

DNA methylation is an epigenetic modification that refers to the methylation of the 5-position on the pyrimidine ring of cytosine, producing 5-methyl-cytosine (5mC) (**Figure 1.16a**). DNA methylation primarily functions in the transcription repression of genes and as such is a key epigenetic regulator (**Figure 1.16b**) (164). Specific functions of DNA methylation include X-chromosome inactivation and genomic imprinting (165–168). The mechanism by which DNA methylation represses gene regulation hasn't been fully elucidated, however possibilities include:

- DNA methylation interfering with the binding of proteins involved in transcription initiation to the DNA (169).
- Binding of proteins to methylated DNA regions blocking the binding of other proteins involved in transcription initiation (164).
- Alterations in chromatin structure induced by DNA methylation (170).

In mammals, methylation primarily occurs at CpG dinucleotides, and the methylation signature of the genome is carried forward during DNA replication (171). In humans, approximately 60-80% of all CpG dinucleotides are methylated (172,173). The frequency of CpG dinucleotides is lower than expected by random chance in the genome, and CpG dinucleotides frequently occur in what are called CpG islands (CGIs) (174). CGIs are regions of the genome, with an abnormally high frequency of CpG dinucleotides, they comprise about 1% of the genome, and make up less than 10% of all CpGs. These CGIs are prevalent at transcription start sites, primarily of housekeeping genes and are thus commonly unmethylated.

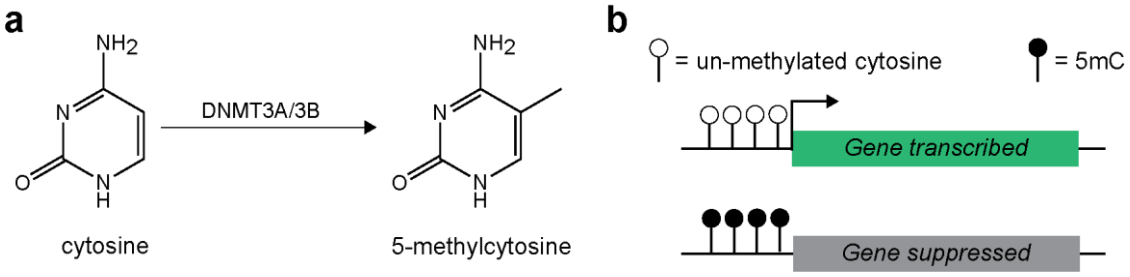


Figure 1.16 DNA methylation overview
a) DNMT3A/3B catalysed methylation of the 5 position on the pyrimidine ring of cytosine. **b)** Diagram of gene suppression by DNA methylation at a gene promoter. 5mC = 5-methylated cytosine.

DNA methyl transferases

DNA is methylated by DNA methyltransferase enzymes, the DNMTs (175). There are four main DNMT genes in humans, *DNMT1, DNMT3A, DNMT3B and DNMT3L*. Only DNMT3A and DNMT3B are capable of *de novo* methylation however, with DNMT1 responsible for the maintenance of methylation during DNA replication, and DNMT3L acting as a co-factor for DNMT3A, stimulating its DNA methylation activity (168,176–179). None of the DNMTs have reported DNA sequence specificity however, and their mechanism of targeting to specific regions of the genome to establish methylation is a highly researched question (171).

As well as their C-terminal catalytic domains, DNMT3A and DNMT3B both possess a Pro-Trp-Trp-Pro (PWWP) domain that can recognise trimethylated Histone 3 Lysine 36 (H3K36me3) (**Figure 1.17a**) (180). H3K36me3 is catalysed by SET domain containing 2 (SETD2) protein, a histone methyltransferase enzyme that can be recruited by actively transcribing RNA Pol II (181,182). This informed the theory that DNMTs are targeted to actively transcribing genes. In line with histones and their modifications being responsible for targeting DNMTs to genomic regions, DNMT3A and DNMT3B also possess an ATRX-DNMT3L-DNMT3A (ADD) domain, which has been reported to recognise unmethylated histone 3 lysine 4 (H3K4), with methylation at this lysine repelling *de novo* methylation (**Figure 1.17a**) (183,184). DNMT3L has a vastly different sequence, with only the ADD domain and a truncated C-terminal catalytic domain resulting in it being catalytically inactive (185). Crystallographic analysis has shown that DNMT3L forms a DNMT3L-DNMT3A-DNMT3A-DNMT3L tetramer which enhances DNMT3A catalytic activity (186).

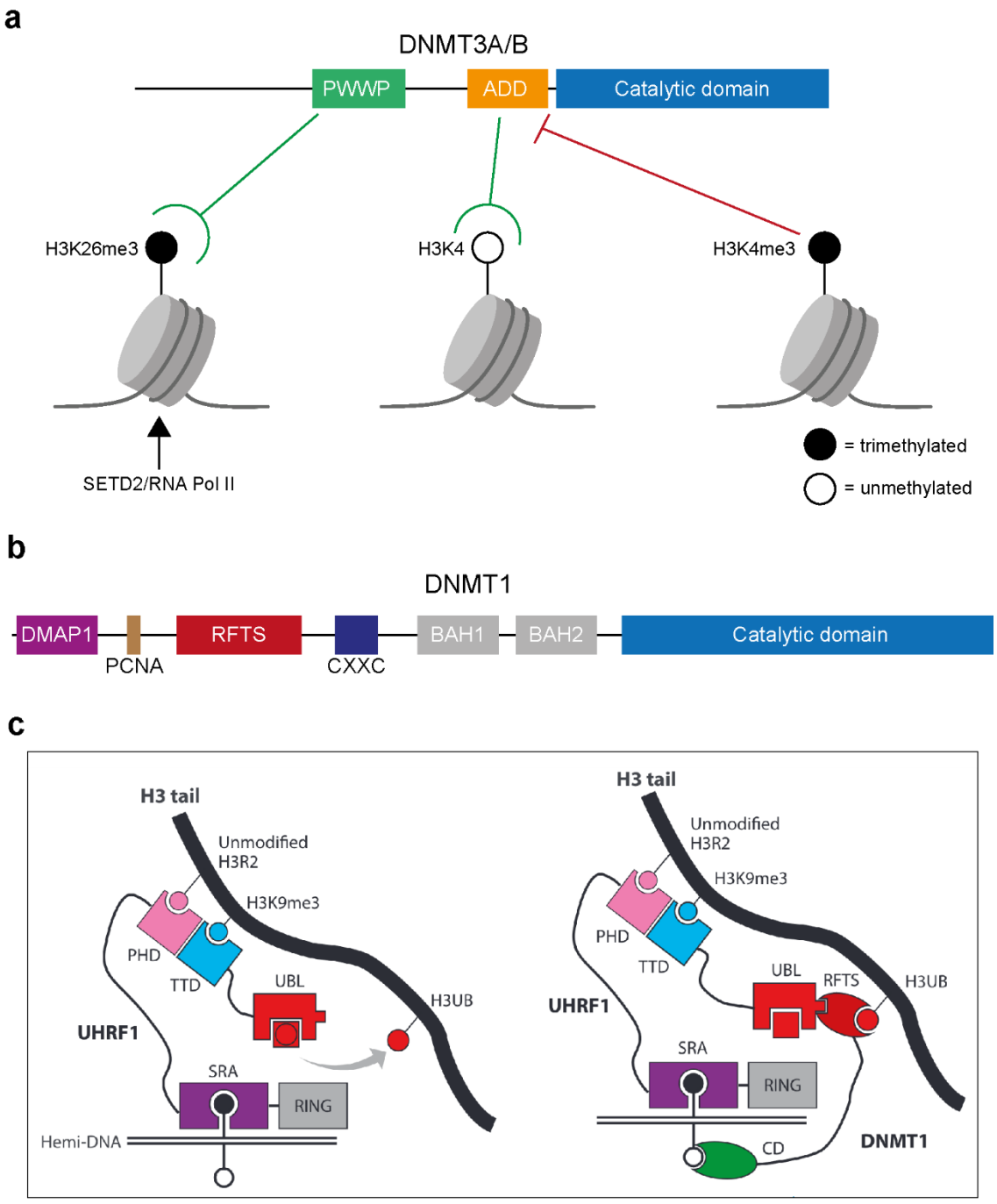


Figure 1.17 DNMT domains and recognition
a) DNMT3A/B domain architecture with histone modification recognition. **b)** DNMT1 domain architecture. **c)** UHRF1 mediated DNMT1 methylation of hemi-methylated DNA, figure reproduced from reference (171) with permission (**Appendix 20**).

DNMT1 on the other hand has a significantly different sequence to the other DNMTs, with neither a PWWP or ADD domain and a significantly longer catalytic domain (**Figure 1.17b**) (171). The proliferating cell nuclear antigen (PCNA) and

replication foci-targeting sequence (RFTS) are the two key domains for targeting DNMT1 to DNA for maintenance of methylation marks (187,188). The primary function of DNMT1 is to maintain the DNA methylation signature of the replicating genome during cell division. To do this, DNMT1 works with UHRF1 an E3 Ubiquitin ligase that recruits DNMT1 to hemi-methylated DNA at the replication foci during S phase (**Figure 1.17c**) (93,189,190). The underlying mechanism involves the binding of UHRF1 to hemi-methylated DNA, as well as both unmodified histone 3 arginine 2 (H3R2), and trimethylated histone 3 lysine 9 (H3K9) (191,192). This is followed by the ubiquitination by UHRF1 of multiple lysines on histone 3 that are then recognised by the RFTS domain of DNMT1 which in turn fully methylates the hemi-methylated DNA (187). Crucially, DNMT1 does not methylate unmethylated CpG dinucleotides.

1.3.2 Genomic imprinting

Genomic imprinting refers to the parent-of-origin dependent monoallelic expression of a subset of autosomal genes (167,193). Such genes are termed imprinted genes, with either the maternal or paternal allele being expressed, and as of 2019, approximately 100 had been identified in humans (160). A variety of epigenetic factors have been reported to be involved in this parent-of-origin expression, resulting in altered accessibility of gene promoters to transcription factors. One of the primary epigenetic factors implicated in imprinting is differential methylation signatures between the maternal and paternal alleles of imprinted genes, with methylation at an imprinted loci in one allele normally leading to transcriptional suppression (**Figure 1.18**) (194). Such loci, commonly CGIs, are termed germline differentially methylated regions (gDMRs) referring to the different methylation states between alleles, established in the germline (194).

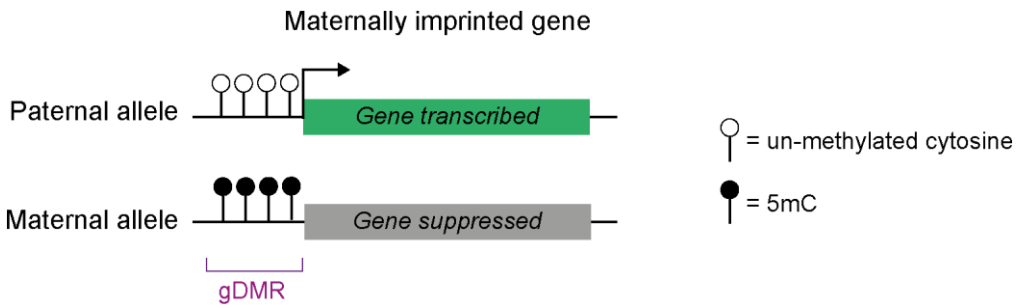


Figure 1.18 Imprinting gDMR overview
Methylation signatures at the gDMR of a maternally imprinted gene, methylated on the maternal allele and so only expressed from the paternal allele. gDMR = germline DMR.

Imprinted genes are commonly grouped together in the genome in small clusters, under the control of specific regions called imprinting control regions (ICRs) (195,196). These ICRs are normally differentially methylated and can regulate the imprinted expression of many different imprinted genes within the cluster up to several megabases away from the centre (167,197).

Other epigenetic factors reported to be involved in imprinting are histone modifications and long non-coding RNA. In the case of histone modifications, there is limited evidence to suggest they directly affect the expression of imprinted genes (167). Histone modifications, in particular H3K4me3, have been suggested to shield DNA from *de novo* methylation during the establishment of the imprinted signature (183). Instead of directly suppressing the expression of one allele, it has therefore been postulated that histone modifications function only in directing the establishment of the germline imprinted signature, and do not directly affect the expression of imprinted genes (167). Long non-coding RNA has a more direct effect on the expression of imprinted genes. It has been reported that the allele-specific expression of long non-coding RNAs in certain imprinting clusters can regulate the expression of other protein coding or protein non-coding imprinted genes within the same cluster (198,199).

1.3.3 DNA methylation through oogenesis and early embryo development

The maternal and paternal imprinted DNA methylation signatures are established during germline development and maintained throughout embryo development (171). This specific methylation signature is orthogonal to the substantial DNA methylation reprogramming that occurs throughout germline and early embryo development. During implantation, the embryo's methylation signature is fully established and remains primarily static throughout the rest of life (200).

In both mice and humans, DNA methylation is almost completely erased genome-wide upon primordial germ cell specification (201–205). This happens in developing embryos before birth and is achieved primarily through passive dilution caused by down-regulation of DNMT1 and UHRF1 resulting in the failure to maintain methylation marks during DNA replication (206,207). Methylcytosine dioxygenase TET1 (TET1) mutant mice have been reported to show aberrant hypermethylation at

meiotic genes in female germ cells however, suggesting a role for TET1 in this methylation erasure (208,209). Subsequent work has postulated that TET1 is not directly involved in DNA demethylation but prevents aberrant *de novo* methylation at specific loci (210).

After DNA demethylation in PGCs, the mechanisms and timings of establishment of the germline methylation signature diverges between male and females (171). In males, the methylation signature of germ cells is fully established at birth, and the sperm genome is almost fully methylated (~75% in humans, 80% in mice) (211–213). In female germ cells however, it occurs after birth during oocyte growth and has a much more specific signature (214). This oocyte methylation signature is established by DNMT3A and DNMT3L in mice (**Figure 1.19**), and by only DNMT3A in humans as DNMT3L is not expressed in human oocytes (168,179,211,214–217). Various mechanisms have been proposed to explain how the oocyte methylome is established in mouse oocytes. Interestingly, gene transcription correlates with methylation of the transcribed genes, suggesting that transcription is guiding *de novo* DNA methylation during oocyte growth (218). This could be explained by transcription promoting the recruitment of DNMT3A/3L for *de novo* methylation through trimethylation of H3K36 by RNA Pol II-recruited SET domain containing 2 (SETD2) (197,219). In support of this, histone modification profiling experiments combined with DNA methylation profiling has shown a strong correlation between H3K36me3 and hypermethylation (220,221). Consistent with this finding, depletion of SETD2 leads to global hypomethylation (219). In contrast, histone modifications shown to repel DNA methylation such as H3K4me3 show an inverse correlation with hypermethylation (115,119). By the end of oocyte growth, mouse oocytes display a global methylation level of approximately 40%, compared to human oocytes with approximately 54% (217). There is little direct experimental evidence to confirm whether the mechanisms shown to govern the establishment of the mouse methylome is conserved in human oocytes. Methylated regions are enriched at gene bodies, suggesting that the establishment of the human oocyte methylation signature is also guided by transcription (217). The reason for the absence of DNMT3L in human oocytes is not known, and mechanism by which DNMT3A *de novo* methylates the oocyte genome in the absence of DNMT3L remains to be elucidated (171).

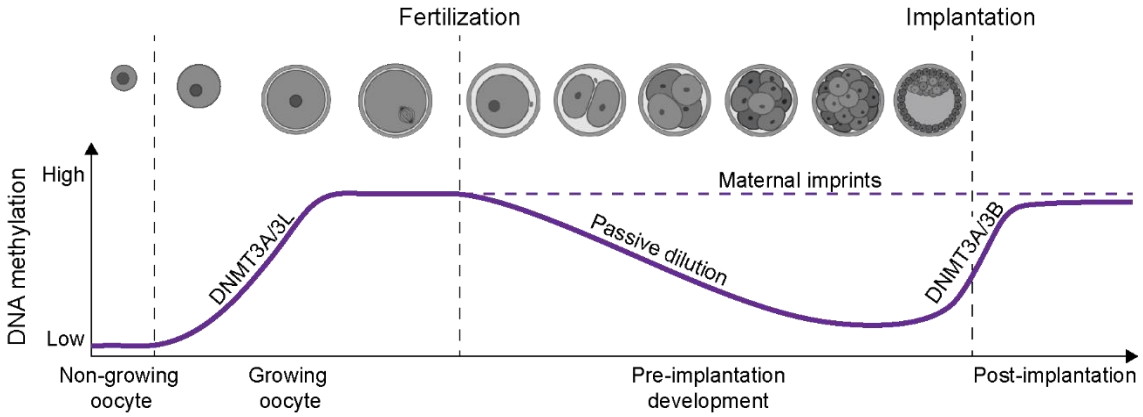


Figure 1.19 Maternal genome methylation dynamics in development
Schematic of the methylation dynamics of the growing oocyte, and the pre-implantation embryo. In non-growing oocytes, genome methylation has been largely erased. A unique maternally imprinted methylation signature, along with specific non-imprinted methylation, is then established during oocyte growth. After fertilisation, the non-imprinted methylation signature not maintained is slowly diluted whilst the maternal imprints are protected from passive demethylation. During implantation, a rapid re-methylation of the genome by DNMT3A and DNMT3B occurs. Figure adapted from (171).

After fertilisation, both the maternal and paternal genomes are demethylated at different rates (171). This demethylation is global, with the exception of the maternal and paternal imprinted gDMRs, which are protected by zinc finger proteins ZFP57 and ZFP445 (222,223). ZFP57 and ZFP445 recognise and bind to imprinted regions, recruit DNMT1 to maintain the methylation marks and promote H3K9 trimethylation to protect imprints from demethylation (224). The global demethylation during early embryo development begins at the zygote stage and is achieved primarily through passive dilution like in primordial germ cells (225). Tet methylcytosine dioxygenase 3 (TET3) has also been reported to be involved in demethylation of the maternal genome (226,227). The embryo reaches its minimum methylation level at the blastocyst stage, and it is at this point that the genome-wide re-methylation occurs to establish the signature found in somatic cells throughout the rest of life (200,216). At implantation, increased levels of DNMT3A and DNMT3B cause this re-methylation, with DNMT3B reported to be the dominant factor (177,228). DNMT3L is present at this stage, but not required, as zygotes depleted of DNMT3L only display a delay in methylation at limited sites, but methylation establishment occurs as normal (229).

1.3.4 Imprinting disorders

Imprinting disorders are congenital conditions that result from alterations in the function or regulation of imprinted genes (160). As many imprinted genes have important functions in development, such disorders are often accompanied by a range of developmental disorders and clinical syndromes, such as growth and metabolic disturbances and cognitive dysfunction (159). It is also believed that hydatidiform mole (HM) formation is often a result of imprinting disturbances, however in many cases the cause of HM formation remains unknown (194).

There are four main genetic or epigenetic alterations, known as *cis*-factors, that can lead to imprinting disorders (194):

- Copy number variations – the deletion of an imprinted allele.
- Single nucleotide variants – a point mutation in an imprinted allele resulting in a non-functional protein product.
- Uniparental disomy – the inheritance of both chromosomes, or chromosomal regions from a single parent.
- Epimutations:
    - Primary: alterations in the epigenetic modifications or interactions involved in imprinting themselves without genetic sequence alterations.
    - Secondary: genetic alterations in the gDMR (cis-acting) or proteins not encoded near the gDMR but that either interact or play an upstream role in its regulation (trans-acting).

MLID are defined as imprinting disorders with alterations at multiple imprinted loci (194). The epigenetic characteristics associated with patients living with MLID include either hypo- or hypermethylation at multiple imprinted loci (159). As the alterations occur in multiple locations across the genome, MLID are most commonly caused by deletions or variants of *trans*-acting factors (194). These *trans*-acting factors are genes often distant from gDMRs, though involved in the establishment or maintenance of the imprinted signature of the whole genome (159). As such, trans-acting factor variants often result in methylation defects at a wide range of imprinted loci, and not just a single site, as is the case with *cis*-factors.

PADI6 and MLID

It has been suggested that PADI6 is an imprinting *trans*-acting factor. Multiple women with *PADI6* variants have been reported to have children who are later diagnosed with MLID (**Figure 1.15, Table 1.6**) (20). Women with variants in SCMC protein genes have also been reported to have children with MLID, suggesting that the function of PADI6 in maternal imprinting is also linked with the SCMC, and potentially the CPLs (81,159,161,163,194,230). Numerous studies have been conducted where the methylation state of peripheral blood lymphocytes from MLID patients was characterised (159). MLID patients from mothers possessing *PADI6* and SCMC protein variants displayed dramatic hypomethylation at multiple maternal gDMRs, with varying levels of hypomethylation, and different gDMRs being hypomethylated depending on the maternal variants (159).

The mechanism behind PADI6 and maternal imprinting is unclear, however. The most important question that must be answered is whether PADI6 is involved in the establishment of the maternal imprinted signature during oocyte growth, its maintenance throughout early embryo development, or both. No methylation sequencing data has been reported for oocytes or embryos from women with *PADI6* variants. An experiment has been performed on five MII oocytes and a blastocyst like embryo that were donated for methylation sequencing by a patient homozygous for a loss-of-function *KHDC3L* variant, one of the SCMC proteins (231).The patient in this case was infertile, and thus did not give birth to any children with MLID. However, it is the only reported case of methylation sequencing being conducted on oocytes or embryos from women possessing SCMC gene variants. When compared with two separate reported datasets for WT oocytes, a significant reduction in methylation percentage was observed at maternal gDMRs in the *KHDC3L* variant oocytes. Interestingly however, this hypomethylation was not restricted to the maternal gDMRs. Hypomethylation was observed globally across the genome, suggesting loss of KHDC3L function resulted in global defects in the establishment of the oocyte methylome. Methylation sequencing of the blastocyst-like embryo also showed further defects in the maintenance of the maternal methylome. Together, this suggests that KHDC3L is important for both the proper establishment, and maintenance of the maternal imprinted signature in humans. Given the likely tight functional link between PADI6 and the SCMC, it is conceivable that PADI6 is also involved in this function.

DNMT3A, the key enzyme responsible for establishing the maternal imprinted signature in mice, is interestingly a substrate of PADI4 (6). It has been shown that citrullination of an arginine in DNMT3A by PADI4 stabilises the protein and upregulates DNA methylation. Given the apparent absence of PADI4 in the oocyte, and the overlapping substrate profiles of other PADIs, it is conceivable that DNMT3A is also a substrate of PADI6. The defects in DNA methylation could thus be attributed to an un-citrullinated and destabilised DNMT3A. It is yet to be confirmed whether PADI6 is a citrullinating enzyme, however.

On the other hand, as discussed in **Chapter 1.2.1**, UHRF1 is down-regulated 90% in PADI6 deficient mouse oocytes and appears to be a key component of the CPLs (67). In the early embryo, UHRF1 is involved in the maintenance of the DNA methylation signature and is vital for early embryo development after fertilisation. Whether UHRF1 is similarly down-regulated in the early embryo is unknown, and more crucially, it us unknown whether *PADI6* variants lead to down-regulation of UHRF1 in human oocytes and embryos. If *PADI6* knock-out or variation does indeed lead to UHRF1 down-regulation in the embryo, then this poses an additional mechanism through which DNA methylation defects, and ultimately MLID are caused.

1.4 Project Aims

PADI6 is clearly indispensable for early embryonic development in humans and mice. Despite efforts to characterise the function of PADI6, the molecular mechanisms behind its function remain largely unknown. One of the main unanswered questions surrounding PADI6, is whether it has a catalytic function or not. Whilst no substrates or *in vitro* catalytic activity have been reported, it is possible that to date neither the right substrates nor activating conditions have been trialled. Whether PADI6 is catalytically active or not is a central question to this thesis and is investigated throughout. Additional questions that I aimed to assess to fully understand the function and importance of PADI6 in early embryo development are:

- What effect does the lack of sequence conservation have on the overall structure of PADI6? What can we learn about the biochemical function of PADI6 from this structure? Investigated in **Chapter 3**.
- Does PADI6 interact with the SCMC? What is the nature and function of this interaction? Investigated in **Chapter 4**.
- What is the molecular function of PADI6 in early embryo development? Is PADI6 catalytic activity required for early embryo development? Investigated in **Chapter 5**.
- How is PADI6 involved in the establishment of the maternal imprinted methylation signature? Investigated in **Chapter 6**.

Chapter 2. Materials and Methods

2.1 General materials

2.1.1 Antibodies

Table 2.1 Antibodies

Antibody	Use: Working concentration	Source	Identifier
Anti-hPADI6, Rabbit	WB: 1 in 300 IF: 1 in 200	Covalab	Raised for this project
Anti-Myc, Rabbit	WB: 1 in 1000	Sigma-Aldrich	#SAB4301136
Anti-HA, Rabbit	WB: 1 in 1000	Cell Signalling	#C29F4
Anti-V5, Rabbit	WB: 1 in 1000	Cell Signalling	#D3H8Q
Anti-FLAG HRP Conjugate, Mouse	WB: 1 in 5000	Sigma-Aldrich	#A8592
Anti-GAPDH, Mouse	WB: 1 in 200	Santa Cruz Biotechnology	#0411, sc-47724
Anti-Rabbit HRP Conjugate, Goat	WB: 1 in 10000	Invitrogen	#32460
Anti-Mouse HRP Conjugate, Sheep	WB: 1 in 10000	Cytivia	#NA931
Anti-Rabbit AlexaFluor 488, Donkey	IF: 1 in 300	Invitrogen	#A21206

2.1.2 Bacterial strains

Table 2.2 Bacterial strains

Bacterial strain	Source	Identifier
NEB® 10-beta	New England Biolabs	#C3019H
NEB® 5-alpha	New England Biolabs	#C2987H
NEB® Nico21(DE3)	New England Biolabs	#C2529H
Rosetta(DE3)pLysS	Merck	#70596-3
ArcticExpress(DE3)	Agilent	#230191

2.1.3 Antibiotics

Table 2.3 Antibiotics

Antibiotic	Source	Identifier	Working concentration
Ampicillin	Scientific Laboratory Supplies	#A0166-25G	50 µg/mL
Kanamycin	Sigma-Aldrich	#60615	50 µg/mL
Penicillin/Streptomycin	Gibco	#15140122	100 µg/mL

2.1.4 Buffers

Table 2.4 Buffers

Buffer	Composition
Home-made KOD buffer (10X)	1.2 M Tris-HCl pH 8.0, 100 mM KCl, 60 mM $(NH_4)_2SO_4$, 1% Triton X-100, 0.01% BSA (w/v)
TAE	40 mM Tris-HCl pH 8.3, 20 mM Acetic Acid, 1 mM EDTA
TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA
E. coli lysis	50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM Imidazole, 1 mM TCEP, 5% Glycerol
Nickel IMAC elution	50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM Imidazole, 1 mM TCEP, 5% Glycerol
Cleavage dialysis	50 mM HEPES pH 7.5, 200 mM NaCl, 1 mM TCEP, 5% Glycerol
SEC standard	50 mM HEPES pH 7.5, 200 mM NaCl, 1 mM TCEP, 5% Glycerol
SEC crystallography	50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM TCEP
IEX buffer A	100 mM Tris-HCl, pH 7.5, 1 mM TCEP
IEX buffer B	100 mM Tris-HCl, pH 7.5, 1 M NaCl, 1 mM TCEP
Expi293 purification	50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM TCEP
HEK-293T lysis	50 mM Tris-HCl pH 7.4, 100 mM NaCl, 50 mM NaF, 1 mM EDTA, 1% Triton X-100
Transfer	25 mM Tris, 192 mM Glycine, 20% methanol (v/v)
NanoDSF	50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM TCEP
SEC-MALLS	50 mM HEPES, pH 7.5, 200 mM NaCl, 1 mM TCEP, 3 mM NaN_3
COLDER assay	50 mM HEPES pH 7.5, 150 mM NaCl, 2 mM DTT
PG-TAMRA assay	50 mM HEPES pH 7.5, 50 mM NaCl, 10 mM $CaCl_2$, 2 mM DTT
PG-TAMRA lysis	50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 50 mM NaF, 10 mM $CaCl_2$, 1% Triton X-100

2.1.5 Primers

Primers were synthesized by Integrated DNA Technologies or ThermoFisher as desalted lyophilised solids. A full list of all primers and their uses can be found in **Appendix 3**, **4** and **13**.

2.1.6 Plasmids

Plasmids were either purchased, donated or sub-cloned as part of this project via InFusion cloning. FLAG = 3x FLAG sequences. For InFusion cloning primers used see **Appendix 3**.

Table 2.5 Purchased, donated or InFusion cloned plasmids

Name	Notes
Purchased plasmids	
pENTR223.1_hPADI6	DNASU, #HsCD00297377
pTWIST_His-mPADI6	Synthesised by Twist Bioscience
pCMV6-Entry_Myc-DDK-OOEP	Origene, #RC216662
pDONR221_TLE6	DNASU, #HsCD00820922
pTWIST_Myc-NALP5	Synthesised by Twist Bioscience
pCMV6-Entry_Myc-DDK-KHDC3L	Origene, #RC215631
pCMV_HA empty vector	Addgene, #32530
pCMV_Myc empty vector	Addgene, #85759
pcDNA3.1_FLAG-V5 empty vector	Addgene, #87063
pCMV5-Entry_FLAG-ZBED3	Origene, NM_032367
Donated plasmids	
pMCSG_His-TEV-hPADI4	Previously cloned in the Walport Group
pGEX6P2_GST-3C-hPADI4	Previously cloned in the Walport group
pET28a_His-hPADI4	Previously cloned in the Walport group
pcDNA3.1_Strep-Strep-TEV empty vector	Aliquot provided by Dr. Dave Briggs of the McDonald group at the Crick
pcDNA3.1_FLAG-GFP	Previously cloned in the Walport group
pcDNA3.1_Myc-GFP	Previously cloned in the Walport group
pcDNA3.1_FLAG-hPADI4	Previously cloned in the Walport group
In-Fusion cloned plasmids	
pMCSG_His-TEV-hPADI6	Backbone source: pMCSG-His-TEV-hPADI4 Insert source: pENTR223.1-hPADI6

pGEX6P2_GST-3C-hPADI6	Backbone source: pGEX6P2-GST-3C-hPADI4 Insert source: pMCSG-His-TEV-hPADI6
pMCSG_His-3C-hPADI6	Backbone source: pMCSG-His-TEV-hPADI6 Insert source: pGEX6P2-GST-3C-hPADI6
pET28a_His-hPADI6	Backbone source: pET28a-His-hPADI4 Insert source: pMCSG-His-3C-hPADI6
pcDNA3.1_Strep-Strep-TEV-hPADI6	Backbone source: pcDNA3.1-Strep-Strep-TEV empty vector Insert source: pET28a-His-hPADI6
pMCSG_His-3C-mPADI6	Backbone source: pMCSG-His-3C-hPADI6 Insert source: pTWIST-His-mPADI6
pcDNA3.1_Strep-Strep-TEV-mPADI6	Backbone source: pcDNA3.1-Strep-Strep-TEV-hPADI6 Insert source: pMCSG-His-3C-mPADI6
pcDNA3.1_FLAG-hPADI6	Backbone source: pcDNA3.1-3XFLAG-hPADI4 Insert source: pMCSG-His-TEV-hPADI6 Insert source: pMCSG-His-3C-mPADI6
pCMV_HA-TLE6	Backbone source: pCMV-HA empty vector Insert source: pDONR221-TLE6
pCMV_HA-OOEP	Backbone source: pCMV-HA empty vector Insert source: pCMV6-Entry-Myc-DDK-OOEP
pcDNA3.1_V5-KHDC3L	Backbone source: pcDNA3.1-3xFLAG-V5 empty vector Insert source: pCMV6-Entry-Myc-DDK-KHDC3L
pET28a_His-NALP5	Backbone source: pET28a-His-hPADI6 Insert source: pCMV-Myc-NALP5
pET28a_His-TLE6	Backbone source: pET28a-His-hPADI6 Insert source: pCMV-HA-TLE6
pET28a_His-OOEP	Backbone source: pET28a-His-hPADI6 Insert source: pCMV-HA-OOEP
pET28a_His-TLE6_WD40	Backbone source: pET28a-His-hPADI6 Insert source: pCMV-HA-TLE6
pcDNA3.1_Strep-Strep-TEV-NALP5	Backbone source: pcDNA3.1-Strep-Strep-TEV-hPADI6 Insert source: pCMV-Myc-NALP5
pcDNA3.1_Strep-Strep-TEV-TLE6	Backbone source: pcDNA3.1-Strep-Strep-TEV-hPADI6 Insert source: pCMV-HA-TLE6
pcDNA3.1_Strep-Strep-TEV-OOEP	Backbone source: pcDNA3.1-Strep-Strep-TEV-hPADI6 Insert source: pCMV-HA-OOEP

2.2 General molecular biology methods

2.2.1 Plasmid transformation

1 µL of DNA was added to a 50 µL *E. coli* aliquot. The sample was gently mixed and incubated on ice for 30 mins. The sample was heat-shocked at 42 °C for 30 secs and incubated on ice for a further 2 mins. 250 µL of Super Optimal broth with Catabolite repression Media (S.O.C.) was added and sample incubated at 37 °C for 1 h. 50 µL of culture was plated on a pre-warmed Lysogeny Broth (LB) agar plate with appropriate antibiotic and incubated at 37 °C overnight.

2.2.2 DNA sequencing

Sanger sequencing was performed by GeneWiz after the InFusion cloning or mutagenesis of all plasmids.

2.2.3 In-Fusion cloning

Both vector backbones were linearized and fragment inserts amplified by polymerase chain reaction (PCR) or restriction enzyme digestion and fragment inserts were amplified by PCR. Primers were designed using SnapGene® software (Dotmatics) for InFusion cloning.

DNA fragments of interest were amplified by PCR. PCR mixes were made up as follows in a total volume of 50 µL:

- 1X Home-made KOD buffer
- 0.125 mM MgCl2
- 0.2 mM dNTP mix (NEB®, N0447L)
- 0.25 µM forward primer
- 0.25 µM reverse primer
- ~50 ng insert plasmid
- 1 µL KOD enzyme (recombinantly expressed by Dr. Tessa Bertran of the Walport group)

The standard PCR program used was as follows:

- Hot start: 95 °C for 30 sec
- Denaturation: 95 °C for 30 sec
- Annealing: 50–65 °C for 30 sec (normally 55 °C, adjusted to primers)
- Extension: 72 °C for 30 sec per kb of product
- Final extension: 72 °C for 7 min
- The denaturation, annealing and extension steps were repeated for 32 cycles.

DNA vector backbones were linearized by PCR. PCR mixes were made up as follows in a total volume of 50 µL:

- 5X Q5 Reaction Buffer (NEB®, B9027S)
- 0.2 mM dNTP mix (NEB®, N0447L)
- 0.5 µM forward primer
- 0.5 µM reverse primer
- ~50 ng backbone plasmid
- 0.5 µL Q5 Polymerase (NEB®, M0491L)

The standard PCR program used was as follows:

- Hot start: 98 °C for 30 sec
- Denaturation: 98 °C for 10 sec
- Annealing: 50–65 °C for 30 sec (normally 55 °C, adjusted to primers)
- Extension: 72 °C for 30 sec per kb of product
- Final extension: 72 °C for 2 min
- The denaturation, annealing and extension steps were repeated for 34 cycles.

Amplified fragment inserts and linearised vectors were resolved by gel electrophoresis in a 1% Agarose TAE gel (100 V, 30 mins) in TAE buffer. DNA was visualised with SYBR Safe DNA gel stain (Invitrogen, #S33102) and extracted and purified from the gel using the PureLink™ Quick Gel Extraction Kit (Invitrogen, #K210012) following the manufacturers protocol. Fragment (~50 ng) and linearised vector (~50 ng) were combined with 2 µL 5X In-Fusion HD Enzyme Premix (Takara Bio, #938910) and mixture made up to 10 µL with Nuclease Free Water (Fisher

Bioreagents, #10336503). The mixture was incubated for 15 mins at 50 °C and 5 µL transformed into 50 µL a *E. coli* NEB® 5-alpha aliquot.

2.2.4 Site-directed mutagenesis

Mutagenesis primers were designed to be ~35 bp in length, with a melting temperature >78 °C with a GC content of ~40% terminating in C or G. For full list of plasmid mutagenesis performed and primers used see **Appendix 4**. Mutagenesis was performed by PCR. PCR mixes were made up as follows in a total volume of 50 µL:

- 1X Pfu reaction buffer (Agilent, #600250)
- 0.2 mM dNTP (NEB®, N0447L)
- 2.5 ng/µL forward primer
- 2.5 ng/µL reverse primer
- ~25 ng plasmid
- 0.5 µL Pfu Polymerase (Agilent, #600250)

The standard PCR program used was as follows:

- Hot start: 98 °C for 30 sec
- Denaturation: 98 °C for 30 sec
- Annealing: 50 °C for 1 min
- Extension: 72 °C for 1 min per kb of plasmid
- The denaturation, annealing and extension steps were repeated for 15 cycles.

After PCR amplification, 1 µL Dnp1 restriction enzyme (NEB, #R0176S) was added to the PCR mixture and the sample incubated at 37 °C for 1 h. After 1 h, 5 µL were transformed into a 50 µL *E. coli* NEB® 5-alpha (High Efficiency) aliquot.

2.2.5 DNA plasmid purification

A single LB agar plate colony was either picked into 4 mL or 300 mL of LB media (Crick Media Preparation) with appropriate antibiotic, incubated at 37 °C overnight and miniprepped (QIAprep Spin Miniprep Kit, Qiagen, #27106), midiprepped (NucleoBond® Xtra Midi EF, Takara Bio, #740420.10) or maxiprepped (ZymoPure II Plasmid Maxiprep Kit, Zymo Research, #D4203) following manufacturers protocol.

2.2.6 Reverse transcription-quantitative PCR

mRNA was diluted in DEPC-treated Nuclease Free Water (Fisher Bioreagents, #10336503) to 1 µg/µL and was reverse transcribed, and no-RT and no template controls produced using LunaScript® RT Super Mix Kit (NEB, #E3010S) as follows:

- RT mix:
  - 4 µL LunaScript RT SuperMix (5X)
  - 1 µL of 1 µg/µL mRNA
  - 15 µL Nuclease Free Water
- No-RT mix:
  - 4 µL No-RT Control Mix (5X)
  - 1 µL of 1 µg/µL mRNA
  - 15 µL Nuclease Free Water
- No template mix:
  - 4 µL µL LunaScript RT SuperMix (5X)
  - 16 µL Nuclease Free Water
- PCR program:
  - Annealing: 25 ^{o}C for 2 min
  - cDNA synthesis: 55 ^{o}C for 10 min
  - Heat inactivation: 95 ^{o}C for 1 min

Transcribed mRNA and no-RT control were diluted in Nuclease Free Water to 5 ng/µL. Lyophilised QuantiTect qPCR primers (Qiagen) for *Padi6* and *B2m* were re-constituted in TE buffer to give a 10X primer solution. qPCR mixes were then made up as follows in a MicroAmp® Fast Optical 0.1 mL 96-Well Reaction Plate (Applied Biosystems™, #4346906):

- 10 µL 2X PowerUp™ SYBR™ Green Master Mix (Applied Biosystems™, #A25741)
- 2 µL 10X QuantiTect qPCR primer mix
- 2 µL cDNA, no-RT or no-template control
- 6 µL Nuclease Free Water

qPCR reactions were then performed in technical quadruplicates and biological duplicates on a QuantStudio™ 12K Flex Real-Time PCR System (Applied Biosystems™) instrument following the standard PowerUp™ SYBR™ Green PCR program.

RT-qPCR data was then analysed using the comparative C_T ($\Delta\Delta C_T$) method to determine relative mRNA quantification between WT and transgenic ovaries. Mean C_T and standard deviation values were calculated for each sample. ΔC_T values were then calculated by subtracting the *B2m* C_T from *Padi6* C_T. $\Delta\Delta C_T$ was then calculated by subtracting the ΔC_T of the WT ovary from the ΔC_T of the transgenic ovary. Fold-difference was then calculated relative to WT ovaries as $2^{-\Delta\Delta C_T}$. The experiment was repeated twice with ovaries from different females, and the $2^{-\Delta\Delta C_T}$ values were averaged and plotted using Graph Pad Prism.

2.2.7 SDS-PAGE

Samples were mixed with a 1/3 volume of 4X Laemli Sample Buffer (Bio-Rad, #1610747) and boiled at 95 °C for 5 mins. Samples were then centrifuged and loaded onto a mPAGE™ 4-20% pre-cast Bis-Tris gel (Merck, #MP42G12). PageRuler™ Prestained Protein Ladder, 10 to 180 kDa (Thermo Scientific™, #26616) was used as a ladder for all gels. Gels were run in NuPAGE™ MOPS SDS Running Buffer (Invitrogen™, NP0001) at 100 V for 1.5 h. Gels were then either processed for western blotting (see **Chapter 2.2.8**), or stained with Quick Coomassie Stain (Generon, #NB-45-00078-1L) following manufacturers protocol.

2.2.8 Western blot

SDS-PAGE gels were wet transferred to a PVDF membrane (AmershamTM HybondTM, 0.45 µm, GE Healthcare, #10600023) applying 100 V for 1 h. An ice pack was placed in the transfer tank and the tank placed on ice to prevent overheating. The membranes were blocked with PBS (5% milk, 0.1% Tween) at room temperature (RT) for 1 h. Membranes were then incubated with primary antibodies at 4 °C overnight, or at RT for 1 h with gentle rocking (source, working concentrations and incubation conditions detailed in **Table 2.1**). Membranes were washed thrice with PBS (0.1% Tween) and incubated with horseradish peroxidase-conjugated mouse or rabbit secondary antibodies at RT for 1 h with gentle rocking (source, working concentrations and incubation conditions detailed in **Table 2.1**). Membranes were washed thrice with PBS (0.1% Tween) and proteins were visualised by chemiluminescence using Pierce™ ECL Plus Western Blotting Substrate (ThermoFisher, #32132). Blots were imaged using a ChemiDocTM MP Imaging System (Bio-Rad).

2.2.9 Co-immunoprecipitation

2.2.9.1 Manual

Total protein quantity was determined using the Pierce™ BCA Protein Assay (ThermoFisher, #23225) and protein concentration was normalised to 1 mg/mL. 1 mg of total protein was immunoprecipitated with 25 µL Anti-FLAG® M2 Magnetic Beads (Millipore, #M8823-1ML) rotating end-over-end gently at 4 °C for 2 h. The beads were washed thrice with 500 µL HEK-293T lysis buffer rotating end-over-end gently at 4 °C for 5 mins per wash. 30 µL of 2X SDS loading buffer was added and the samples were heated at 95 °C for 10 min (Bio-Rad, #1610747). 5 µg total of protein input samples (0.5% IP) were prepared. 5% of IP fraction was taken to analyse bait protein, remaining 95% was taken to analyse co-IP protein. Input and IP samples were resolved by SDS-PAGE and characterised by western blot.

2.2.9.2 KingFisher Duo Prime

Parallel IPs were conducted using a KingFisher Duo Prime instrument (Thermo Scientific™). Total protein quantity was determined using the Pierce BCA Assay (Thermo Scientific™, #23225) and protein concentration was normalised to 1 mg/mL. A KingFisher 96 deep-well, v-bottom plate (Thermo Scientific™, #97003530) was then prepared as follows:

- Row A: 30 µL of 2X SDS loading buffer (Bio-Rad, #1610747)
- Row B: KingFisher 12-tip comb (Thermo Scientific™, #97003530)
- Row C: 25 µL AntiFLAG® M2 Magnetic Beads (Millipore, #M8823-1ML) in 50 µL HEK-293T lysis buffer
- Row D: 1 mL of 1 mg/mL cell lysate
- Row E: 500 µL HEK-293T lysis buffer
- Row F: 500 µL HEK-293T lysis buffer
- Row G: 500 µL HEK-293T lysis buffer

An IP protocol was then run using the protocol described in **Figure 2.1**.

Tip1		KingFisher Duo 12 tip comb	
	Pick-Up	IP plate	(B) - Tip
	CollectBeads1	IP plate	(C) - Dynabeads
		Collect count	3
		Collect time [s]	10
	Bind target	IP plate	(D) - Sample
	Beginning of step	Precollect	No
		Release time, speed	00:00:05, Fast
	Mixing / heating:	Mixing time, speed	02:00:00, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	10
		Post-temperature	No
	Wash I	IP plate	(E) - Wash I
	Beginning of step	Precollect	No
		Release time, speed	00:00:05, Fast
	Mixing / heating:	Mixing time, speed	00:05:00, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	10
		Post-temperature	No
	Wash II	IP plate	(F) - Wash II
	Beginning of step	Precollect	No
		Release time, speed	00:00:05, Fast
	Mixing / heating:	Mixing time, speed	00:05:00, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	10
		Post-temperature	No
	Wash III	IP plate	(G) - Wash III
	Beginning of step	Precollect	No
		Release time, speed	00:00:05, Fast
	Mixing / heating:	Mixing time, speed	00:05:00, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	10
		Post-temperature	No
	Elution Loading Buffer	IP plate	(A) - Elution in Loading Buffer
	Beginning of step	Precollect	No
		Release time, speed	00:00:05, Medium
	Mixing / heating:	Mixing time, speed	00:10:00, Medium
		Heating temperature [°C]	75
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	10
		Post-temperature	No
	Leave	IP plate	(B) - Tip

Figure 2.1 KingFisher IP Protocol
IP protocol used on KingFisher Duo Prime Instrument. BindIt software (Thermo Scientific™) was used to generate the protocol report.

2.3 Recombinant protein expression and purification

2.3.1 *E. coli* large scale expression

Expression plasmid was transformed into an *E. coli* aliquot. A single colony was picked into 100 mL LB (Crick Media Preparation) with appropriate antibiotic and incubated at 37 °C overnight with shaking. Terrific broth (Crick Media Preparation) with appropriate antibiotic was inoculated with overnight culture at a dilution of 20 mL culture per 1 L Terrific Broth. Cultures were incubated at 37 °C with shaking at 200 rpm until OD ~ 1 and induced with 0.1 mM IPTG (NovaChem, #M027626), at 18 °C overnight. Induced cultures were centrifuged at 4000 rpm for 30 mins, supernatant discarded, and bacterial pellets frozen at -80 °C.

2.3.2 *E. coli* cell lysis and clearing

Frozen bacterial pellets were re-suspended in *E. coli* Lysis with 1 tablet per 50 mL buffer protease inhibitor (cOmplete™ EDTA-free, Merck, #73567200), 1 mg/mL lysozyme (Sigma, #L6876-5G) and a micro-spatula full of DNAse (Roche, #10104159001). Bacteria was lysed on ice by sonication (3 secs on, 5 secs off, 30% intensity, 3 mins total on time). Crude lysate was centrifuged at 21,000 rpm for 45 mins on a JA-25.50 rotor (Beckman Coulter) and filtered through 0.45 µm PVDF filters (Merck, #SLHV033RS) using a syringe.

2.3.3 *E. coli* small scale trial expression

The protein expression plasmid was first transformed into *E. coli.* A single colony was picked into 10 mL LB (Crick Media Preparation) with appropriate antibiotic and incubated at 37 °C overnight with shaking. 10 mL Terrific Broth (Crick Media Preparation) with appropriate antibiotic was inoculated with 200 µL overnight culture and incubated at 37 °C with shaking at 200 rpm until OD ~ 1. Cultures were either induced with 0 mM, 0.1 mM, or 1 mM IPTG (NovaChem, #M027626), at 4 °C overnight. OD of cultures was measured, and 2 mL of each culture centrifuged at 13,000 rpm for 1 min. Supernatants were discarded and pellets resuspended in BugBuster® Protein Extracation Reagent (Millipore, #70584-3) at a volume in µL equal to the OD/12. 15 µL of each sample was taken (Uninduced and Total) and combined with 5 µL 4X Laemli Sample Buffer (Bio-Rad, #1610747). Induced samples were shaken at RT for 30 mins to lyse and centrifuged at 13,000 rpm for 30 mins. 15 µL supernatant sample was

taken (Soluble) and combined with 5 µL 4X Laemli Sample Buffer. Remaining supernatant discarded was and the pellet was resuspended in water. 15 µL sample (Pellet) were combined with 5 µL 4X Laemli Sample Buffer. Samples were analysed by SDS-PAGE as detailed in **Chapter 2.2.7**.

2.3.4 Expi293 expression

Expi293 cells (Gibco™, #A14527) were maintained and diluted twice a week by Chloe Roustan of the Crick Structural Biology Scientific Technology Platform (STP). For transfection, cells were counted using Vi-CELL XR counter (Beckmann Coultier) and diluted to 2 x 10^6 cells/mL in 400 mL pre-warmed Expi293™ Expression Medium (ThermoFisher, #A1435101). 24 hours later, cells were transfected as follows: 400 µg MaxiPrepped DNA was diluted in 20 mL Opti-MEM® (ThermoFisher, #11058021), 1.2 mL of 1 mg/mL PEI 25K™ (Polysciences, #23966-1) diluted in 18.8 mL Opti-MEM. Both mixtures were gently mixed by inversion and incubated for 5 mins at RT before being combined and incubated for 20 mins at RT. After incubation, DNA-PEI mixture was titrated dropwise into 400 mL Expi293 cell culture with constant agitation of the culture. Cultures were incubated with 150 rpm agitation, 8% CO_2, at 37 ^oC for 3 days before harvesting (see **Chapter 2.3.5**).

2.3.5 Expi293 cell lysis and clearing

Transfected cell cultures were centrifuged at 2000 rpm for 15 mins at 4 ^oC and the supernatant discarded. Cell pellets were then washed in cold PBS with 50 mL protease inhibitors (1 tablet per 50 mL PBS, cOmplete™ EDTA-free, Merck, #73567200) and centrifuged at 2000 rpm for a further 10 minutes. The supernatant was discarded, cell pellet flash frozen in liquid nitrogen and stored at -80 ^oC.

2.3.6 Nickel immobilised metal affinity chromatography

Cleared and filtered lysate was applied to a His IMAC column (HisTrap HP, 5 mL, GE Healthcare, #17-5248-02) on an ÄKTApure™ system (Cytivia). Step elution was performed with Nickel IMAC Elution Buffer. Elution fractions were characterised by SDS-PAGE and pooled.

2.3.7 In solution tag cleavage by TEV or 3C protease

Protein quantity of pooled IMAC fractions was determined using a NanoDrop ND-1000 Spectrophotometer and either TEV (Walport Group home-made) or 3C protease (donated by Stephane Mouilleron) was added to the sample (1 mg protease per 50 µg protein). Protein-protease sample was transferred into SnakeSkin™, 10K MWCO, 22 mM dialysis tubing (Thermo Scientific, #68100) and dialysed overnight at 4 ^{o}C in 2 L cleavage dialysis buffer. Cleaved sample was then applied to 2 mL Hi-Bind™ Ni QR Agarose Beads (BioVision, #6562) and gently rotated for 1 h at 4 ^{o}C to re-bind un-cleaved protein. After 1 h incubation, cleaved protein flow through was collected by gravity filtration and beads were washed with dialysis buffer (2 x 2 mL). Un-cleaved protein was eluted off the beads with Nickel IMAC Elution buffer (2 x 2 mL). 9 µL of each sample was combined with 3 µL of 4X Laemli sample buffer (Bio-Rad, #1610747) and cleavage efficiency was characterised by SDS-PAGE.

2.3.8 Strep-tag affinity purification and cleavage off resin

Cleared Expi293 cell lysate was added to StrepTactin® XT 4Flow® resin (Iba, #2-5010-025) at a ratio of 2 mL packed beads per 500 mL Expi293 cell culture and incubated at 4 ^{o}C for 4 hours with gentle rotation. After 4 h, the unbound fraction was removed by gravity filtration and the resin was washed with Expi293 purification buffer (5 times, double volume of resin). The resin was then re-suspended in twice its volume of Expi293 purification buffer and TEV protease (recombinantly produced in-house) added at a ratio of 2 mg protease per 500 mL Expi293 cell culture. The cleavage was then incubated at 4 ^{o}C overnight with gentle rocking. The cleaved fraction was then removed by gravity filtration and the resin was again washed with Expi293 purification buffer (3 times, double volume of resin). The cleaved and wash fractions were combined and taken forward for subsequent size exclusion chromatography (SEC) purification.

2.3.9 Size-exclusion chromatography

Protein samples were concentrated to 2 mL using a Vivaspin® 20, 30000 MWCO ultrafiltration centrifugal concentrator (Sartorius, #VS2022). The concentrated protein samples were then purified by size exclusion chromatography using a HiLoadTM 16/600 SuperdexTM 200 pg Gel Filtration Column (Cytivia, # 28989335), in SEC Standard or SEC Crystallography buffer on an ÄKTApure™ system (Cytivia).

Fractions were characterised by SDS-PAGE and pooled. Pooled fractions were concentrated again until no further decrease in volume was observed using a Vivaspin® 20, 30000 MWCO ultrafiltration centrifugal concentrator (Sartorius, #VS2022).

2.3.10 Ion-exchange chromatography

Size-exclusion chromatography purified protein was dialyzed overnight into IEX Buffer A using a Slide-A-Lyzer™ Dialysis Cassette with a 10K MWCO (Thermo Scientific™, #66380). Dialyzed samples were concentrated to 2 mL using an ultrafiltration centrifugal concentrator a Vivaspin® 20, 30000 MWCO ultrafiltration centrifugal concentrator (Sartorius, #VS2022). The concentrated protein samples were then purified by size exclusion chromatography using a 1 mL HiTrap Q FF anion exchange chromatography column (Cytivia, #17505301) on an ÄKTApure™ system (Cytivia). Protein was initially loaded onto the column in 100% IEX Buffer A. IEX Buffer B composition was increased to 50% over 20 column volumes, followed by an increase to 100% over 10 column volumes. Pooled fractions were concentrated again until no further decrease in volume was observed using a Vivaspin® 20, 30000 MWCO ultrafiltration centrifugal concentrator (Sartorius, #VS2022).

2.3.11 Band-ID

SDS-PAGE gel bands were excised using a scalpel and placed in 1.5 mL Protein LoBind® tubes (eppendorf, #0030108116) in minimal HPLC water (provided by Crick Proteomics STP) and handed off to the Crick Proteomics STP for Band-ID.

2.3.12 Intact-MS of purified proteins

Purified recombinant proteins were diluted to a final concentration of 20 µM in SEC crystallography buffer and handed to Sarah Maslen of the Crick Proteomics STP.

The following was completed by Sarah Maslen of the Crick Proteomics STP.

Denatured proteins were injected onto a C4 BEH 1.7µm, 1.0 x 100mm UPLC column using an Acquity I class LC (Waters™). Proteins were eluted with a 15 min gradient of acetonitrile (2% v/v to 80% v/v) in 0.1% v/v formic acid using a flow rate of 50 µl/min. The analytical column outlet was directly interfaced via an electrospray

ionisation source, with a time-of-flight (TOF) BioAccord mass spectrometer (Waters™). Data was acquired over a *m/z* range of 300 – 8000, in positive ion mode with a cone voltage of 40v. Scans were summed together manually and deconvoluted using MaxEnt1 (Masslynx, Waters™).

2.4 Biophysical methods

2.4.1 Circular dichroism

Proteins were diluted to an approximate concentration of 0.15 mg/mL in PBS (pH 7.4, Crick Media Preparation) and their exact diluted concentration was measured using V-550 Spectrophotometer. A circular dichroism spectrum was taken between 190 and 260 nm using a J-815 CD Spectrometer (Jasco), for a phosphate buffer background followed by the protein sample. The phosphate buffer background spectrum was subtracted from protein sample spectrum and CD signal was converted to mean residue CD extinction coefficient as described by Martin & Schilstra (232). CD data was plotted using GraphPad Prism.

2.4.2 Nano differential scanning fluorimetry (Nano-DSF)

Proteins were diluted to an approximate concentration of 0.15 mg/mL NanoDSF buffer. Melting experiments were performed using Prometheus™ NT.48 High Sensitivity Capillaries (NanoTemper, #PR-C006-200) on a Prometheus (NanoTemper) instrument with a 20 to 95 ^{o}C melt at 1 ^{o}C per minute. The 350/30 nm fluorescence ratio was plotted against temperature using GraphPad Prism. Melting temperatures were extracted as the first derivative peak in the melting curve calculated using the Panta Analysis software (NanoTemper).

2.4.3 NanoDSF calcium titration

Proteins were diluted to an approximate concentration of 0.15 mg/mL in NanoDSF buffer with an additional 0, 0.02, 0.06, 0.2, 0.6, 2, 6, or 20 mM $CaCl_2$. Proteins were melted as per **Chapter 2.4.2** and melting temperatures at each calcium concentration plotted using GraphPad Prism.

2.4.4 Size Exclusion Chromatography multi-angle laser light scattering (SEC-MALLS)

Protein samples were dialysed into SEC-MALLS buffer and diluted to 4, 2, 1 and 0.5 mg/mL. Protein samples were applied in a volume of 100 µL to a Superdex^{TM} INCREASE 200 10/300 GL column (GE Healthcare) equilibrated in SEC-MALLS buffer at a flow rate of 1.0 mL/min. The scattered light intensity and the protein concentration of the column eluate were recorded using a DAWN-HELEOS laser photometer (Waters™) and OPTILAB-rEX differential refractometer respectively (Waters™). The weight-averaged molecular mass of material contained in chromatographic peaks was determined from the combined data from both detectors using the ASTRA software version 6.0.3 (Wyatt Technology Corp).

2.4.5 Mass photometry

The following was conducted by Roger George of the Crick Structural biology STP.

Proteins were diluted to an approximate concentration of 20 nM in NanoDSF buffer. 2 µL of diluted protein was then added to 18 µL of PBS in the well of a gasket on a TwoMP instrument (Refeyn) and events recorded over the course of 1 min using the AquireMP software (Refeyn). Molecule sizes were then calculated using the AnalyseMP software (Refeyn) using BSA (66 kDa), ADH (150 kDa) and Urease (90, 272, 544 kDa) as standards. Mass normalised event histograms were then produced using GraphPad Prism with bin sizes of 5 kDa.

2.4.6 Buffer screening

The rubic 96-well buffer screen (Molecular Dimensions, #MD1-96) was completed by diluting recombinant protein in 2 µL of 1 mg/mL protein in 18 µL of each buffer from the 96-well screen and running NanoDSF to determine melting temperatures as per **Chapter 2.4.2**.

2.5 X-Ray crystallography

2.5.1 Reagents

Table 2.6 X-Ray crystallography reagents

Name	Supplier	Code
Primary screens		
JCSG Core I	NeXtal	#130924
JCSG Core II	NeXtal	#130925
JCSG Core III	NeXtal	#130926
JCSG Core IV	NeXtal	#130927
Pact Premier™	Molecular Dimensions	#MD1-36
Morpheus®	Molecular Dimensions	#MD1-47
PEG-Ion	Hampton Research	#HR2-139
Optimisation reagents		
50% w/v PEG 3350	Molecular Dimensions	#MD2-100-9
2.5 M Potassium thiocyanate	Molecular Dimensions	#MD2-250-58
1 M Bis-Tris propane	Molecular Dimensions	#MD-005-PH

2.5.2 Protein crystallisation

Sitting-drop crystallization trials were carried out at 25 and 4 °C by mixing 100 nL of protein and 100 nL of reservoir solution with a total volume of 75 µL in the reservoir. Crystallisation primary screens were laid using a mosquito® LCP (SPT labtech). Primary screens trialled are detailed in **Table 2.6**. Grid optimisation screens centred on a primary screen hit condition were formulated using a Formulator® (Formulatrix).

2.5.3 Preparation of multi matrix seed stock

The following was completed by Stephane Mouilleron of the Crick Structural Biology STP.

Crystals from a previous optimisation tray were harvested in their mother liquor in a Seed Bead (Hampton research) and crystals were crushed by vortexing. The seed stock was then diluted 1:10 in mother liquor and use for seeding experiments.

2.5.4 Seeded protein crystallisation

Seeded sitting-drop crystallization trials were carried out at 25 °C by mixing 300 nL of protein, 50 nL seed stock, and 250 nL reservoir solution with a total volume of 75 µL in the reservoir. Crystallisation trays were laid using a mosquito® LCP (SPT labtech). The Pact Premier™ and PEG-Ion screens were trailed. Grid optimisation screens centred on a primary screen hit condition were formulated using a Formulator® (Formulatrix).

Final crystal used to determine structure was present in a grid optimisation screen of PACT Premier™ Well F4. Crystal was full size after 21 days, final crystallisation conditions were as follows: 20.9% w/v PEG3350, 0.2 M KSCN, 0.1 M Bis-Tris propane, pH 6.5.

2.5.5 Crystal harvesting, cryoprotection and structure determination

The following was completed by Stephane Mouilleron of the Crick Structural Biology STP.

PADI6 crystallisation was achieved by micro-seeding as per **Chapter 2.5.4**. Four crystals were cryoprotected in perfluoropolyether cryo-oil (Hampton research, #HR2-814) and flash-frozen in liquid nitrogen. X-ray data were collected at 100 K at beamlines I03 (mx25587-64) of the Diamond Light Source Synchrotron (Oxford, UK). Data collection and refinement statistics are summarized in **Table 3.1**. Data sets were indexed, scaled and merged with xia2 (233). Molecular replacement used the atomic coordinates of an AlphaFold 2.0 model of PADI6. Refinement used Phenix (234). Model building used COOT (235) with validation by PROCHECK (236). Figures were prepared using the PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC).

2.6 Cryo-EM

2.6.1 Data acquisition

The following was completed by Donald Benton formerly of the Crick Structural Biology STP.

Cryo-EM samples were prepared by applying 4 µL of 1.5 mg/mL mPADI6, supplemented with ß-octylglucoside to reduce preferential orientation, to a glow discharged Quantifoil 200 mesh R2/2 grid (Electron Microscopy Sciences, #Q210CR2). Grids were blotted and plunge frozen using a Mark III Vitrobot (FEI). Data were then collected on a Titan Krios operating at 300kV, equipped with a Falcon 4i and Selectris energy filter, operating in zero-loss mode with a 10 eV slit width. Images were acquired for 6.1 sec and acquired in .eer format with a total dose of 28.0 e/Å2 and a calibrated pixel size of 0.95 Å.

2.6.2 Electron density map determination

The motion in the movies was corrected using MotionCor2 implemented in RELION (237), subdividing the .eer file in 52 fractions, each with 0.78 e/Å2. Contrast Transfer Function was estimated using CTFfind4. Particles were picked using Topaz and 2D classified using cryoSPARC (238). Particles belonging to classes with well-defined secondary structure were selected and an initial 3D model calculated using the ab-initio reconstruction in RELION. These selected particles were refined in RELION 3D Auto-refine prior to Bayesian polishing. Polished particles were classified using 3D classification in RELION. Particles belonging to the best-defined class were refined using non-uniform refinement in cryoSPARC coupled to CTF refinement. The maps obtained had their local resolution estimated using blocres implemented in cryoSPARC, followed by filtering by local resolution and global sharpening in cryoSPARC.

2.7 Cell culture methods

2.7.1 General cell culture methods

Human Embryonic kidney cells (HEK 293-T, Crick Cell Services) were cultured in DMEM (Gibco™, #11965092) supplemented with 10% Foetal Bovine Serum (FBS) (Sigma, #F7524-50ML) with penicillin and streptomycin in a 5% CO_2 atmosphere at 37 ^{o}C.

2.7.2 Transient transfection

DNA quantities in transfections were optimised based on numerous western blot experiments (data not shown). Cells were seeded at 1 x 10^6 cells per well of a Nunc™ cell-culture treated 6-well plate (ThermoScientific™, #140675) 24 h before transfection. Transient transfections were performed using Lipofectamine™ 2000 (Invitrogen™, #11668027) in 1 mL Opti-MEM per well (ThermoScientific™, #11058021) according to the manufacturer's protocol (DNA quantities used for transfection detailed in **Table 2.7**). After 4 h transfection, Opti-MEM was replaced with pre-warmed DMEM (10% FBS, penicillin/streptomycin). At 40-48 h post-transfection, cells were harvested as described in **Chapter 2.6.3**.

Table 2.7 Transient transfection DNA quantities

Plasmid	DNA quantity per well of 6-well plate /ng
pcDNA3.1_FLAG-hPADI6	500
pcDNA3.1_FLAG-GFP	20
pcDNA3.1_Myc-GFP	500
pTWIST_Myc-NALP5	100
pCMV_HA-TLE6	200
pcDNA3.1_V5-KHDC3L	400
pCMV_HA-OOEP	100
pcDNA3.1_FLAG-ZBED3	500
pcDNA3.1_FLAG-hPADI4	500

2.7.3 Cell harvesting

DMEM was discarded from the plate and cells were washed once with cold PBS buffer. Cold PBS was aspirated, and cells lysed with HEK 293-T lysis buffer on ice for 30 min. Lysates were clarified by centrifugation at 13,000 rpm for 15 min at 4 ^{o}C, and supernatants collected.

2.8 Citrullination characterisation methods

2.8.1 COLDER assays

Recombinant mPADI4 and hPADI4 expressed in *E. coli* were donated by Dr. Tessa Bertran of the Walport group. PADI activity was assessed using the Colour Development Reagent (COLDER) assay (239). Reactions were carried out on 96-well plates (Thermo Scientific #260895). Recombinant hPADI6 (final conc. = 500 nM), hPADI4 (final conc. = 50 nM), mPADI6 (final conc. = 500 nM), or mPADI4 (final conc. = 50 nM) in COLDER buffer with 10 mM $CaCl_2$ was combined with standard PADI substrate N-α-Benzoyl-L-arginine ethyl ester hydrochloride (BAEE, final conc. = 10 mM, Merck, #B4500). Reactions were quenched with 50 mM EDTA either before substrate and enzyme were combined, or after 1 h of incubation at RT. Citrulline level in the mixtures after quenching was then determined as follows. 1 volume of 80 mM Diacetyl monoxime/2,3-butanedione monoxime (Sigma, #31550-25G), 2 mM thiosemicarbazide (Thermo Scientific Chemicals, #138901000), and 3 volumes of 3 M H_3PO_4, 6 M H_2SO_4 and 2 mM $NH_4Fe(SO_4)_2$, was added to each well and the plate heated for 15 mins at 95 ^{o}C. After 15 mins, plates were cooled to RT and absorbance measured at a wavelength of 540 nm using a CLARIOstar Plus microplate reader (BMG Labtech). Absorbance bar charts were plotted using GraphPad Prism.

2.8.2 Phenyl glyoxal probe

The PG-TAMRA probe was synthesised and characterised by Dr. Ewen Calder of the Walport group.

2.8.2.1 In vitro histone citrullination assay

The *in vitro* histone citrullination assay was conducted as follows, based on (240). 1.5 µL of 1 mg/mL histone II-A from calf thymus (Merck, #H9250-100MG) and

1.5 µL of 10 µM recombinant hPADI4 (provided by Dr. Tessa Bertran of the Walport group, final conc. = 0.5 µM) were added to 27 µL of PG-TAMRA assay buffer and incubated at 37 °C. Reactions were quenched after 0, 1, 2, 5, 10, 30, or 60 mins by the addition of 3 µL of 0.5 M EDTA (final conc. = 50 mM). Labelling was performed by addition of 10 µL trichloroacetic acid (20% final), 3.125 µL of 1.6 mM PG-TAMRA probe (final conc. 0.1 mM), and 3.875 µL PG-TAMRA assay buffer and incubating for 30 mins at 37 °C. The labelling reaction was quenched by the addition of 12.5 µL of 500 mM L-Citrulline (Alfa Aesar, #A13316) in PG-TAMRA assay buffer (final conc. = 100 mM). Samples were cooled on ice for 30 mins and centrifuged at 14,000 rpm for 15 mins at 4 °C. Supernatants were removed and precipitated protein washed with cold acetone by vortexing and centrifuging at 14,000 rpm for 5 mins. Samples were then dried for 5 mins at 100 °C and re-suspended in 22.5 µL of 50 mM HEPES with 7.5 µL of 4X Laemli-SDS sample buffer before being resolved by SDS-PAGE (see **Chapter 2.2.7**). Gels were then imaged using a Typhoon FLA 9500 imager (GE Healthcare) set to excitation wavelength 542 nm and emission wavelength 568 nm. After in-gel fluorescent imaging, protein loading was characterised gels by staining with Quick Coomassie and imaging as in **Chapter 2.2.7**.

2.8.2.2 Cell lysate citrullination assay

The cell lysate citrullination assay was conducted as follows, based on (240,241). HEK-293T cells were transiently transfected as per **Chapter 2.6.2** in Nunc™ cell-culture treated 6-well plate (ThermoScientific™, #140675) with the following DNA (**Table 2.8**)

Table 2.8 Cell lysate transient transfection DNA quantities

Sample	DNA (quantity)
PADI6	pcDNA3.1_FLAG-hPADI6 (500 ng)
PADI6 + SCMC	pcDNA3.1_FLAG-hPADI6 (500 ng), pTWIST_Myc-NALP5 (100 ng), pCMV_HA-TLE6 (200 ng), pCMV_OOEP (100 ng)
PADI4	pcDNA3.1_FLAG-hPADI4 (500 ng)
PADI4 + SCMC	pcDNA3.1_FLAG-hPADI4 (500 ng), pTWIST_Myc-NALP5 (100 ng), pCMV_HA-TLE6 (200 ng), pCMV_OOEP (100 ng)
GFP	pcDNA3.1_FLAG-GFP (10 ng)
GFP + SCMC	pcDNA3.1_FLAG-GFP (10 ng), pTWIST_Myc-NALP5 (100 ng), pCMV_HA-TLE6 (200 ng), pCMV_OOEP (100 ng)

48 hours after initial transfection, cells were harvested into 100 µL PG-TAMRA lysis buffer, protein quantity was determined using the Pierce™ BCA Protein Assay (ThermoFisher, #23225) and protein concentration was normalised to 2 mg/mL. 33 µL of each 2 mg/mL sample was combined with 10 µL TCA (20% final), 3.125 µL of 1.6 mM PG-TAMRA (final conc. = 0.1 mM) and 3.875 µL PG-TAMRA assay buffer and incubated for 30 mins at 37 °C. The labelling reaction was quenched by the addition of 12.5 µL of 500 mM L-Citrulline (Alfa Aesar, #A13316) in PG-TAMRA assay buffer (final conc. = 100 mM). Samples were cooled on ice for 30 mins and centrifuged at 14,000 rpm for 15 mins at 4 °C. Supernatants were removed and precipitated protein was washed three times with cold acetone by vortexing and centrifuging at 14,000 rpm for 5 mins. Samples were then dried for 5 mins at 100 °C and re-suspended in 15 µL of 50 mM HEPES with 15 µL of 4X Laemli-SDS sample buffer before being resolved by SDS-PAGE (see **Chapter 2.2.7**). Gels were then imaged using a Typhoon FLA 9500 imager (GE Healthcare) set to excitation wavelength 542 nm and emission wavelength 568 nm. After in-gel fluorescent imaging, protein loading was characterised gels by staining with Quick Coomassie and imaging as in **Chapter 2.2.7**.

2.9 Mouse work

2.9.1 General husbandry

Mice maintained in the Francis Crick Institute BRF colony were housed in individually ventilated cages (GM500, Tecniplast) with automatic water, air, and humidity management systems, and a 12:12 light: dark cycle. Adult females and males were housed in separate cages in groups of 3-4 animals. Mice had free access to water and food, and enrichment activities including rodent balls and nesting boxes were provided. Pups were genotyped by either Andrew Ojrarikre of the Turner lab or by the BRF using assays for the *Padi6*^{C663A} and *Padi6* knock-out alleles using Transnetyx. Only females homozygous for the desired genetic mutations after Transnetyx genotyping were used in experiments.

2.9.2 Humane culling procedure

Mice were culled by cervical dislocation, and the death confirmed by observation of the onset of rigor mortis.

2.9.3 Timed natural matings

Natural matings for the collection of MII oocytes, zygotes or 2-cell embryos were conducted by Andrew Ojarikre of the Turner Group as follows: females were placed into the cage of males at approximately 1700 h. The following morning females were examined for the presence of a vaginal plug. If a plug was observed, this was taken as evidence that a mating had occurred during the night prior. 1200 h on the day a plug was observed was assumed to be E0.5.

2.9.4 Collection of GV oocytes

Female mice between the ages of 8 and 10 weeks were culled, and ovaries removed and placed in 2 mL of room temperature M2 medium (Sigma-Aldrich, #M7167) with 250 µM Dibutyryl cAMP (DBcAMP, Sigma-Aldrich, #D0260-5MG) in a 35 mm Cell Culture Dish (Corning, #430165). Using a Leica MC80 dissecting microscope, ovarian follicles were repeatedly punctured with an 30G BD Microlance™ 3 needle (Camlab, #304432) to release oocytes. GV oocytes were identified by presence of a clear nuclei in the centre of a round cell. Oocytes were aspirated using a Stripper pipette with a 275 µm tip (Cooper Surgical, #MXL3-STR and #MXL3-275), and washed five times through droplets of M2 medium, followed by three times through droplets of Dulbecco's PBS (Gibco, #14190-094). GV oocytes were then processed for different experiments as described in the relevant sections of this chapter.

2.9.5 Collection of MII oocytes

Female mice between the ages of 8 and 10 weeks were set up in timed natural matings with sterile XXY males, as previously described in (242), provided by the Turner group at the Crick. At approximately 1700h on the day the plug was found, females were culled, and their uteri were removed and placed a 50 µL droplet of room temperature M2 medium (Sigma-Aldrich, #M7167) in a 60 mm dish (Corning, #430196). Using a Leica MC80 dissecting microscope, uteri were dissected using surgical forceps (Dumont, #0208-5-PO) and cumulus masses were aspirated using a Stripper pipette with a 275 µm tip (Cooper Surgical, #MXL3-STR and #MXL3-275). Cumulus cells removed by incubation in a droplet of 3 mg/mL Hyaluronidase (Sigma-Aldrich, #H4272-30G) in FHM for 5 seconds and MII oocytes were then washed five times through droplets of M2 medium, followed by three times through droplets of Dulbecco's PBS

(Gibco, #14190-094). MII oocytes were then processed for different experiments as described in the relevant sections of this chapter.

2.9.6 Collection of zygotes

Female mice between the ages of 8 and 10 weeks were set up in timed natural matings with experienced wild-type C57BL/6 males. At approximately 1700h on the day the plug was found, females were culled, and their uteri were removed and placed in a 50 µL droplet of room temperature Embryo Max® FHM medium (Sigma Aldrich, #MR-024D) in a 60 mm dish (Corning, #430196). Using a Leica MC80 dissecting microscope, uteri were dissected using surgical forceps (Dumont, #0208-5-PO) and cumulus masses were aspirated using a Stripper pipette with a 275 µm tip (Cooper Surgical, #MXL3-STR and #MXL3-275). Cumulus cells removed by incubation in a droplet of 3 mg/mL Hyaluronidase (Sigma-Aldrich, #H4272-30G) in FHM for 5 seconds and zygotes were identified by the presence of one or two polar bodies, and clearly visible pronuclei. Zygotes were then washed five times through droplets of FHM medium, followed by three times through droplets of Dulbecco's PBS (Gibco, #14190-094). Zygotes were then processed for different experiments as described in the relevant sections of this chapter.

2.9.7 Collection of 2-cell embryos

Female mice between the ages of 8 and 10 weeks were set up in timed natural matings with experienced wild-type males. At approximately 900h on the second day after the plug was found, females were culled, and their uteri were removed and placed a 50 µL droplet of room temperature Embryo Max® FHM medium (Sigma Aldrich, #MR-024D) in a 60 mm dish (Corning, #430196). Using a Leica MC80 dissecting microscope, uteri were dissected using surgical forceps (Dumont, #0208-5-PO) and 2-cell embryos were aspirated using a Stripper pipette with a 275 µm tip (Cooper Surgical, #MXL3-STR and #MXL3-275). 2-cell embryos were then washed five times through droplets of FHM medium, followed by three times through droplets of Dulbecco's PBS (Gibco, #14190-094). 2-cell embryos were then processed for different experiments as described in the relevant sections of this chapter.

2.9.8 Dissociation of 2-cell embryos into blastomeres

For scRNA-seq, 2-cell embryos were dissociated into individual blastomeres. 2-cell embryos were initially extracted as per **Chapter 2.9.7** and after initial FHM washes were transferred to a droplet of 1X acidic Tryode's solution (Sigma-Aldrich, #T1788) for approximately 1 min or until ZP visibly disappeared. 2-cell embryos were then washed three times through droplets of FHM before being added to room temperature 1X StemPro™ Accutase™ (Gibco™, #A1110501) for 5 mins. 2-cell embryos were then washed in three further droplets of FHM medium before being dissociated into single blastomeres by repeatedly aspirating and forcefully ejecting the embryo using a Stripper pipette with a 100 µm tip (Cooper Surgical, #MXL3-STR and #MXL3-100). Individual 2-cell blastomeres were then further processed for scRNA-sequencing as per **Chapter 2.12.1.**

2.9.9 Time-lapse imaging of zygote development to blastocyst

Mouse zygotes were extracted by Dr. Afshan McCarthy of the Niakan group at the Crick and were cultured in drops of pre-equilibrated Global medium (LifeGlobal, #LGGG-20) supplemented with 10% human serum albumin supplement (LifeGlobal, #GHSA-125) and overlaid with mineral oil (Origio, #ART-4008-5P). Pre-implantation embryos were incubated at 37 °C and 5.5% CO2 in an EmbryoScope+ time-lapse incubator (Vitrolife) and over a period of five days. Time-lapse videos of embryo development were then analysed using Fiji. Pro-nuclear fading time was used to normalise the time at which the embryos reached the following stages: 2-cell, 4-cell, 8-cell, morula, cavity formation, and blastocyst.

2.9.10 mRNA extraction from mouse ovaries

Female mice between the ages of 8 and 10 weeks were culled, and ovaries removed, flash frozen in liquid nitrogen and stored at -80 ^{o}C. On the day of extraction, ovaries were thawed and resuspended in buffer RLT from the RNeasy® Mini Kit (Qiagen, #74104). Ovaries were homogenised using a 21G BD Microlance™ 3 needle (camlab, #304432) and syringe and mRNA extracted using RNeasy® Mini Kit (Qiagen, #74104) following manufacturers protocol.

2.9.11 Ovary mRNA amplification and sequencing

Primers were designed to amplify a region of the *Padi6* mRNA of length 1101 bp, spanning the mutation that results in the C663A substitution in the protein. These primers did not fall within the deleted exon in the *Padi6*$^{-/-}$ strain. Forward primer: CCCAAGGGTTTCCAAGC, reverse primer: CCTCCAGAAGTGTCAAGC.

Ovary mRNA was extracted as per **Chapter 2.9.10**. mRNA was diluted in DEPC-treated Nuclease Free Water (Fisher Bioreagents, #10336503) to 1 µg/µL and was reverse transcribed using LunaScript® RT Super Mix Kit (NEB, #E3010S) as follows:

- RT mix:
  - 4 µL LunaScript RT SuperMix (5X)
  - 1 µL of 1 µg/µL mRNA
  - 15 µL Nuclease Free Water
- PCR program:
  - Annealing: 25 °C for 2 min
  - cDNA synthesis: 55 °C for 10 min
  - Heat inactivation: 95 °C for 1 min

cDNA product was diluted in DEPC-treated Nuclease Free Water to 50 ng/µL and PCR mixtures made up as follows in a total volume of 50 µL:

- 5X Q5 Reaction Buffer (NEB®, B9027S)
- 0.2 mM dNTP mix (NEB®, N0447L)
- 0.5 µM forward primer
- 0.5 µM reverse primer
- ~50 ng backbone plasmid
- 0.5 µL Q5 Polymerase (NEB®, M0491L)

The standard PCR program used was as follows:

- Hot start: 98 °C for 30 sec
- Denaturation: 98 °C for 10 sec
- Annealing: 50–65 °C for 30 sec (normally 55 °C, adjusted to primers)
- Extension: 72 °C for 30 sec per kb of product (45 secs)
- Final extension: 72 °C for 2 min
- The denaturation, annealing and extension steps were repeated for 34 cycles.

Amplified fragments were resolved by gel electrophoresis in a 1% Agarose TAE gel (100 V, 30 mins) in TAE buffer. DNA was visualised with SYBR Safe DNA gel stain (Invitrogen, #S33102). PCR mixes were then purified using the EasyPure® PCR Purification Kit (Tran, #EP101-01) and sent for sanger sequencing as per **Chapter 2.2.2** with the following in-house designed sequencing primer: CATCATCCTGATCCCGC.

2.10 Immunofluorescent imaging

2.10.1 Zygote fixation and staining

All fixation, washing and staining steps were performed in 4-well non-adherent Nunc cell culture dish (ThermoFisher, #176740). All wash steps were conducted with 1X Dulbecco's PBS (Gibco, #14190-094) with 0.1% Tween. Zygotes were harvested as per **Chapter 2.9.6**. Zygotes were placed in a 500 µL 4% paraformaldehyde (PFA, Sigma, #158127-500G) solution and incubated for 1 h at 4 ^{o}C on a rocking shaker. Zygotes were then washed quickly through three wells of wash buffer before being permeabilized for 20 mins at RT in 1X PBS with 0.5% Tween. Zygotes were then blocked in 10% donkey serum diluted in 1X PBS with 0.1% Tween for 1 h at RT. Zygotes were then incubated with primary antibodies (**Table 2.1**) in blocking solution overnight at 4 ^{o}C and washed quickly three times before incubating in wash buffer for 20 mins at RT. Zygotes were then stained with secondary antibodies (**Table 2.1**) in blocking solution for 1 h at RT before being washed quickly three times and incubated in wash buffer for 20 mins at RT. Finally, zygotes were transferred to a solution VECTASHIELD® with DAPI (Vector Laboratories, #H-1200-10) diluted 1 in 30 in wash buffer and stored at 4^{o}C until imaging.

2.10.2 Image acquisition and analysis

On the day of imaging, zygotes were transferred to a solution VECTASHIELD® with DAPI (Vector Laboratories, #H-1200-10) diluted 1 in 30 in wash buffer, in an 8-well glass bottomed µ-Slide plate (ibidi, #80827) and imaged using an Olympus FV3000-Invert microscope. Captured images were processed using Fiji software (243) to adjust brightness and contrast.

2.11 Transmission electron microscopy of GV oocytes

2.11.1 GV oocyte harvesting and initial fixation

GV oocytes were harvested as per **Chapter 2.9.4**. After PBS washes, oocytes were transferred into a solution of 5% (v/v) glutaraldehyde (Taab Laboratory Equipment Ltd, #G002) / 4% (v/v) formaldehyde (Taab Laboratory Equipment Ltd, #F017) in 0.1M Phosphate buffer for 1 hour. Fixed oocytes were then handed to Anne Weston of the Crick Electron Microscopy STP for further processing.

2.11.2 Full fixation, sectioning and imaging

The following was completed by Anne Weston of the Crick Electron Microscopy STP.

For electron microscopy, oocytes were processed using a Pelco BioWave Pro+ microwave (Ted Pella Inc, Redding, USA). Each step, up to and including 75% resin in acetone was performed in the Biowave, except for the phosphate buffer (PB) and water wash steps, which consisted of two washes on the bench followed by two washes in the Biowave without vacuum (at 250W for 40s). All the chemical incubations were performed in the Biowave for 14min under vacuum in 2min cycles alternating with/without 100W power. The SteadyTemp plate was set to 21ºC. In brief, the samples were stained with 1% (v/v) osmium tetroxide (Taab) / 1.5% (v/v) potassium ferricyanide (Sigma), incubated in 1% (w/v) tannic acid in 0.05M PB pH 7.4 (Sigma), followed by 1% sodium sulphate in 0.05 M PB pH 7.4 (Sigma). The samples were then washed in dH_2O and dehydrated in a graded ethanol series (25%, 50%, 70%, 90%, and 100%, twice each), and in acetone (3 times) at 250W for 40s without vacuum. Exchange into Epon 812 resin (Taab) was performed in 25% 50% and 75% resin in acetone, at 250 W for 3 min, with vacuum cycling (on/off at 30 sec intervals). The samples were transferred to 100% resin overnight before embedding at 60ºC for 48 h. 70 nm sections were sliced with a diamond knife on a RMC Powertome Ultramicrotome, and sections analysed by a 1400 Flash (Jeol) with Jeol Matataki Flash sCMOS camera.

2.12 Single cell RNA sequencing

2.12.1 Sample preparation

MII oocytes, zygotes and dissociated 2-cell blastomeres were harvested as described in **Chapter 2.9.5** to **Chapter 2.9.8**. After the final PBS wash, cells were dispensed into 12.5 µL CDS sorting solution from the Takara Smart-SeqHT kit (TakaraBio, #634439) composed of 0.95 µL 10X Lysis Buffer, 0.05 µL RNase Inhibitor, 1 µL 3' SMART-Seq CDS Primer II A, 10.5 µL Nuclease-Free water (Invitrogen, #AM9937) in a 96-well eppendorf twin.tec PCR LoBind plate (Eppendorf, #0030129512). Plates were then spun down at 1000 rpm for 1 min at 4 °C and samples snap frozen on dry ice before being stored at -80 °C. Samples were harvested over 18 sessions producing 18 x 96-well plates each with varying numbers of cells. On the first day of library preparation, plates were thawed in batches on ice and single cell solutions consolidated into a single PCR submission plate before being handed to Daniel Leonce of the Crick Advanced Sequencing STP for library preparation and sequencing.

2.12.2 Library preparation and sequencing

The following was completed by Daniel Leonce of the Crick Advanced Sequencing Facility.

Sequencing libraries were prepared using the Takara Smart-SeqHT Kit (Takara Bio, #634439) according to the manufacturer's instructions. Briefly, cells were lysed and polyadenylated RNA captured in a total volume of 12.5 µL. One-Step first strand synthesis and double stranded cDNA amplification were carried out with 18 PCR cycles. Amplified cDNA was purified via a SPRISelect (Beckman Coulter, #B23317) 1x bead clean-up and eluted in 15 µL Elution Buffer. Nextera Dual Index libraries were prepared with the Nextera XT kit (Illumina, #FC-131-1096). Purified cDNA was normalised to 0.2 ng/µL input and tagmentation was carried out by adding 10 µL tagment DNA Buffer + 5 µL Amplicon Tagment mix to 5 µL cDNA. The tagmentation reaction was run for 10 min at 55 °C. A final index PCR was performed after adding 10 µL Nextera compatible dual indices (at 10 µM) + 15 µL NPM PCR Mix. Amplified libraries were purified via a SPRIselect (Beckman Coulter, #B23317) 0.9x bead clean-up. The quality and fragment size distributions of the purified libraries was assessed by

a 4200 TapeStation Instrument (Agilent). Libraries were pooled and sequenced on the Illumina NovaSeq6000 in PE100 configuration to a read depth of at least 5M reads per sample.

2.12.3 Data processing

Fastq files were mapped to the GRCm38 genome and converted to gene count files using Nextflow nf-core/rnaseq pipeline, version 22.10.3 (244) with parameters: aligner = star_rsem, genome = mm10.

2.12.4 PCA

Transcriptomic analysis was performed using the Python ScanPy package (245). Gene count files were imported using ScanPy function read_csv. Counts were normalised to counts per million (CPM) with the ScanPy function normalize_total with parameter target_sum = 1e6. CPM values were then converted to log2 values using the ScanPy function log1p with parameter base = 2. PCA analysis was performed using the ScanPy function pca on the full dataset, as well as oocyte, zygote, and 2-cell split data. Principle components were extracted and plotted using GraphPad Prism.

2.12.5 Gene counts per million box plots

Transcriptomic analysis was performed using the Python ScanPy package. Gene count files were imported using ScanPy function read_csv. Counts were normalised to counts per million (CPM) with the ScanPy function normalize_total with parameter target_sum = 1e6. For each gene of interest, counts per million at each stage and genotype was extracted and plotted using GraphPad Prism.

2.12.6 MA plots

Transcriptomic analysis was performed using the Python ScanPy package. Gene count files were imported using ScanPy function read_csv. Counts were normalised to counts per million (CPM) with the ScanPy function normalize_total with parameter target_sum = 1e6. CPM values were then converted to $\log_2$ values using the ScanPy function log1p with parameter base = 2. For each stage, replicate $\log_2$CPM values for each gene were averaged, and the mat-*Padi6*$^{-/-}$ or mat-*Padi6*$^{C663A/C663A}$ vs. WT $\log_2$CPM ratios were then calculated for each gene and plotted against average gene CPM across all genotypes using the Python Seaborn package function scatterplot.

Those genes that had a CPM fold-change > 2 or < 0.5 with a false discovery rate of 0.05 (calculated using the Python statsmodels package function fdrcorrection) were highlighted as up and downregulated respectively. Venn diagrams were produced using the Python matplotlib_venn function venn2. GO analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (246), and Benjamini-Hochberg adjusted p-values extracted and plotted in Prism.

2.12.7 Minor and major EGA gene identification

Minor and major EGA genes were identified using a Python script similar to that described in **Chapter 2.12.6**, however comparing WT zygote vs. WT oocyte replicate data for minor EGA genes, and WT 2-cell vs. WT zygote replicate data for major EGA genes. Additionally, a CPM fold-change > 5 was used, and only those genes that were up-regulated were extracted. MA plots were then produced using the Python Seaborn package function scatterplot.

2.12.8 EGA gene heatmap

For minor and major EGA genes identified in **Chapter 2.12.7**, log_2CPM mat-*Padi6*$^{-/-}$or mat-*Padi6*$^{C663A/C663A}$ vs. WT ratios for each gene were calculated as previously described. The log_2Ratios were then clustered into 5 clusters using the Python scipy function fcluster, and then plotted using the Python seaborn function clustermap.

2.13 Embryo proteomics

2.13.1 Sample preparation and processing

GV oocytes were harvested as per **Chapter 2.9.4**. After PBS washes, single GV oocytes were dispensed into 5 µL of lysis buffer (see **Chapter 2.13.2, Table 2.9**) in 0.5 mL Protein LoBind tubes (eppendorf, #022431064) using a Stripper pipette with a 275 µm tip (Cooper Surgical, #MXL3-STR and #MXL3-275). Tubes were spun down briefly using a bench-top microcentrifuge and stored at -80 °C. On the day of sample preparation samples were thawed on ice and handed to Tania Auchynnikava of the Crick Proteomics STP for further processing and data acquisition.

2.13.2 Lysis buffers

Table 2.9 Proteomic lysis buffers

Buffer	Composition
Lysis buffer 1	100mM TEAB pH 8.5 (Sigma, #T7408-100 mL), 10% SDS
Lysis buffer 2	100mM TEAB pH 8.5 (Sigma, # T7408-100 mL), 0.2% DDM, 40 ng/µl Trypsin (Pierce™, #90057)

2.13.3 Initial method

A modified S-Trap™ (Protifi) sample processing protocol was initially employed. In brief, samples were resuspended in Lysis Buffer 1, acidified using formic acid, and loaded onto an S-trap™ column. Following standard S-trap washes, samples were digested with 0.5 ng of Trypsin (Pierce™, #90057) for one hour at 37°C. Following digestion, peptides were eluted from the S-trap™ using 0.1% formic acid and loaded onto Evotips (Evosep, #EV2013) and prepared using the manufacturers protocol. Peptides were separated using an Evosep+ 40SPD Whisper method (Evosep) and mass spectrometry data acquisition was performed using TIMS TOF Bruker Pro2 (Bruker) operated in data independent acquisition (DIA) parallel accumulation-serial fragmentation (PASEF) mode.

2.13.4 Adapted One-Tip method

A modified version of the published protocol was used (247). Oocytes were lysed using 5 µl Lysis Buffer 2 by brief sonication (10 sec in standard water bath sonicator). Evotips (Evosep, #EV2013) were pre-conditioned with 5 µl of lysis buffer before the sample was loaded and proteins digested for 1h at 37°C. Samples were then acidified using 2 µl of 20% formic acid, briefly washed with 0.1% formic acid and prepared for LC MS/MS analyses. Peptides were separated using an Evosep+ 40SPD Whisper method (Evosep) and MS/MS data acquisition was performed using TIMS TOF Bruker Pro2 (Bruker) operated in DIA PASEF mode.

2.13.5 Data analysis

Raw data was analysed using Spectronaut 18 (Biognosys) in DirectDIA mode using standard settings. Both protein and peptide q-values cut offs were set to 0.01.

2.14 Low input bisulphite sequencing

2.14.1 GV oocyte sample harvesting and preparation

GV oocytes from 3-week-old females were harvested as per **Section 2.9.4** by Dr. Sugako Ogushi of the Turner group and myself. Batches of 20 GV oocytes were dispensed into 1.5 mL DNA LoBind® tubes (eppendorf, #0030108051) in minimal Dulbecco's PBS (Gibco, #14190-094) using a mouth pipette. 20 µL RLT+ buffer (Qiagen, #1053393) was added and samples mixed gently by pipetting before being flash frozen in liquid nitrogen and stored at -80 °C.

2.14.2 HeLa gDNA harvesting

One 10 cm dish of HeLa cells was harvested, and DNA extracted using the QIAmp DNA Mini Kit (Qiagen, #51304) following the manufacturers protocol. Extracted DNA was then diluted to 1 ng/µL and 1 µL samples were combined with 19 µL RLT+ buffer (Qiagen, #1053393) to make 20 µL aliquots containing 1 ng of HeLa DNA and frozen at -80 °C.

2.14.3 In-house sequencing primers

Sequencing primers (listed in **Appendix 13**) based on the IDT for Illumina-TrueSeq UDIs, were synthesised by Integrated DNA Technologies. All primers were initially dissolved in Ambion™ Nuclease-Free Water (Invitrogen, #AM9937) to a concentration of 100 µM. i5/i7 mixes were then prepared by diluting 10 µL of each with 80 µL nuclease free water to give a final concentration of 10 µM. Combined primer mixes were then stored at -20 °C.

2.14.4 Library preparation

Bisulphite conversion was completed using the Zymo EZ DNA methylation-Direct Kit (Zymo, #D5020) and low input bisulphite sequencing libraries were prepared using the Zymo Pico Methyl-Seq™ Library Prep Kit (Zymo Research, #D5456) following the manufacturers protocol with several adaptions.

Briefly, 20 µL HeLa gDNA or GV oocyte samples were thawed and subjected to Proteinase K digestion followed by bisulphite conversion using the Zymo CT

Conversion Reagent. Bisulphite converted libraries were then subjected to a column clean up consisting of an initial wash, a desulphonation wash, and two further washes before DNA was eluted off the column in 8 µL M-Elution buffer. 7 µL of the bisulphite converted DNA was then subject to an initial random amplification with 40 or 10 µM PreAmp primer initial concentration, and two ramping PCR cycles, with addition of further PreAmp polymerase after the first cycle. PreAmp amplified DNA was then subjected to a column clean-up. Adapter primers were then added in a second amplification round consisting of 10 PCR cycles, followed by a column clean-up step, eluting 12.5 µL of adapterized DNA. Finally, sequencing primers were added in a third amplification round: 12 µL of adapterized DNA was combined with 13 µL Library Amp Master mix and 1 µL of 10 µM combined i5/i7 primers (see **Chapter 2.14.3**) and amplified by 10 PCR cycles, and a final volume of 26 µL. Libraries were then subjected to either one round of 0.9X, or two rounds of 1X, SPRIselect bead clean-up (see **Chapter 2.14.5**). Final library concentrations were quantified using a Qubit Fluoremeter and fragment size distribution was evaluated using an Agilent Tapestation 4200 with High Sensitivity D1000 ScreenTape, Ladder and Reagents (Agilent, #5067-5584, #5067-5587, #5067-5585). Libraries were pooled and sequenced on the Illumina NovaSeq6000 in PE100 configuration to a read depth of at least 20M, 50M or 200M reads per sample.

2.14.5 SPRIselect bead clean-up

SPRIselect beads (Beckman Coulter, #B23317) were placed at RT 30 mins before clean-up. Fresh 80% ethanol prepared using Ambion™ Nuclease-Free Water (Invitrogen, #AM9937). Either 22.5 µL (0.9x) or 25 µL (1x) SPRI select beads added to 25 µL column cleaned library sample and mixed slowly by pipetting until homogenously brown. Samples were rotated at 200 rpm for 15 mins before magnet used to immobilise beads and 45 µL of the clear supernatant removed and discarded from each sample without disturbing beads. 200 µL 80% ethanol was added to each tube, gently mixed and removed. Washing step was repeated, residual supernatant was removed from the tube and tubes were left open to air dry for 1 min at RT. Beads were gently re-suspended in 15 µL Zymo DNA Elution buffer (part of Zymo Pico Methyl-Seq™ Kit) and rotated at 200 rpm for 5 mins. After 5 mins, a magnet was used to remove supernatant and transfer to fresh tubes and frozen at -20 ^{o}C.

2.14.6 Data processing

Quality control of raw sequencing data was carried out using FastQC. Using TrimGalore! (248), fastq files were trimmed in single-end mode to remove adapter sequences and low-quality bases with the command:

```
trim_galore --clip_r1 10 --three_prime_clip_r1 10
```

Trimmed data were mapped to reference genomes (either GRCh38 or GRCm39) using Bismark (249) with the command:

```
bismark --non_directional
```

Mapped reads were then deduplicated using Bismark with the command:

```
deduplicate_bismark
```

Deduplicated methylation calls were then extracted using Bismark with the command:

```
bismark_methylation_extractor --bedGraph --multicore 2 --comprehensive --cytosine_report
```

Data processing statistics including mapping efficiency, deduplication rates, unique sequences after deduplication, and CpG, CHG and CHH methylation levels were extracted from Bismark output reports.

2.14.7 CpG coverage analysis

CpG coverages were calculated from the Bismark extracted CpG report Python. Reads at each CpG for each sample were extracted from the Bismark CpG report after deduplication, and the number of CpGs that were covered 1, 2, 3, or 4, times were counted. These counts were then divided by the total number of CpGs in the genome as calculated from the CpG report and multiplied by 100 to get a percent coverage at a depth of 1, 2, 3, or 4. CpG coverages were then plotted using GraphPad Prism.

For the CpG coverage of combined datasets, replicate data from each genotype was combined *in silico* using Python by combining the Bismark CpG reports of each. The CpG coverage was then calculated and plotted as above.

2.14.8 PCA of datasets

CpG reports for each sample were imported into R using the methRead function from the package methylKit (250). PCA was then performed using the methylKit function PCASamples and PC1 vs. PC2 plotted using the autoplot function in the R ggbio package.

2.14.9 Analysis of methylation at different genomic features

Due to poor processing statistics and lack of clustering after PCA, WT replicate 2 and *Padi6*$^{C663A/C663A}$ replicate 4 were not included. CpG reports for each sample (excluding WT replicate 2 and *Padi6*$^{C663A/C663A}$ replicate 4) were imported into R using the methRead function from the package methylKit. Data were then pooled using the commands unite (with the parameter min.per.group = 0L) and pool from methylKit. Methylation percentage at non-overlapping tiles and genomic loci were then calculated as follows:

2000 bp non-overlapping tiles, minimum coverage 100 CpGs

Methylation counts were extracted from the pooled data using the methylKit command tileMethylCounts with the parameters: win.size = 2000, step.size = 2000, cov.bases = 100. Percent methylation at each tile was then calculated from the counts using the methylKit command percMethylation and extracted to a .csv file. A violin and scatter plot were then produced using GraphPad Prism. Pearson correlation coefficients were calculated using GraphPad Prism.

Genomic loci, minimum coverage 20 CpGs

Genomic loci coordinates were imported and defined as follows. For genes, Ensembl GTF files were transformed to GRanges format in R using the import function from the package rtracklayer (251). GTF version GRCm39.109 was used. The imported GRanges object was then filtered include only genes, but to exclude pseudogenes and Y chromosome genes. For promoters, Ensembl GTF files were transformed to GRanges format in R using the import function from the package rtracklayer. GTF version GRCm39.109 was used. The GRanges object was the filtered to only include promoters. For intergenic regions, Ensembl GTF files were transformed to GRanges format in R using the import function from the package rtracklayer. GTF version GRCm39.109 was used. Intergenic regions were defined as those bases not

encompassed by a gene annotation, using the function gaps from the R package GenomicRanges. For CGIs, the mouse CGI coordinates were kindly provided by Dr. Bryony Leeke, formerly of the Turner group at the Francis Crick Institute. This included a bedGraph file of CGI locations in a GRanges format, converted from CGI annotation files downloaded from UCSC.

CpG counts at each genomic loci were then calculated using the methylKit function regionCounts with the parameter: cov.bases = 20. Percent methylation at each loci was then calculated from the counts using the methylKit function percMethylation and extracted to a .csv file. Violin plots and scatter plots were then produced using GraphPad Prism. Pearson correlation coefficients were calculated using GraphPad Prism.

Maternal gDMRs, minimum coverage 5 CpGs

Mouse maternally imprinted DMR coordinates were kindly provided by Dr. Bryony Leeke, formerly of the Turner group at the Francis Crick Institute. They were initially downloaded from the Web Atlas of Murine Imprinting and Differential Expression (WAMIDEX) (252) and transferred onto the GRCm38 genome. These coordinates were then transferred to the GRCm39 genome using the USCS Genome Browser liftOver tool (253), imported into R and converted to GRanges using the makeGRangesFromDataFrame command in the GRanges package. CpG counts at maternal imprinted DMRs were then calculated using the methylKit function regionCounts with the parameter: cov.bases = 20. Percent methylation at each DMR was then calculated from the counts using the methylKit command percMethylation and extracted to a .csv file. A violin and scatter plot, and heatmap were then produced using GraphPad Prism.

2.15 Statistical analysis

The range was used for error bars in box plots. For two-way comparisons, statistical significance based on two-tailed Student's t-tests were calculated in GraphPad Prism. All data are from at least two independent experiments. Statistical significance in the form of a two-tailed p-value of the embryonic survival Kaplan-Meier plot was calculated using GraphPad Prism. scRNA-seq PCA was performed using the Python ScanPy package. scRNA-seq false discovery rates were calculated using the Python statsmodels package. scRNA-seq EGA gene clustering was performed using

the Python scipy package. Single embryo proteomics protein p-values were calculated using Spectronaut 18 (Biognosys). Low input BS-seq PCA was performed using R methylKit package. Low input BS-seq Pearson coefficients were calculated using GraphPad Prism. Specific statistical analysis for each experiment is discussed in the experimental methods above.

Chapter 3. Structural characterisation of PADI6

3.1 Introduction

As mentioned in **Chapter 1.1**, PADI6 is less conserved with the other PADIs than they are with each other. Crucially, calcium binding residues are not conserved, and the key catalytic cysteine conserved in PADIs 1-4 is replaced by an alanine flanked by two cysteines, for human PADI6 (hPADI6) or a cysteine and serine for mouse PADI6 (mPADI6). Owing to these differences, the current widely accepted hypothesis is that PADI6 is not an active citrullinating enzyme like the other PADIs. One of the most common ways PADI *in vitro* activity is characterised is using the colour developing reagent (COLDER) assay, commonly performed with the pan PADI substrate benzoyl arginine ethyl ester (BAEE) (239). One COLDER assay has been reported in which mPADI6 displayed no catalytic activity against BAEE (254), however hPADI6 *in vitro* activity has not been characterised. COLDER assays are often performed in 10 mM $CaCl_2$, conditions that activate PADIs 1-4 *in vitro*. Whilst it is conceivable that PADI6 could require alternate activating conditions, it is first important to both confirm that PADI6 does not bind calcium and to validate reported data that under these conditions it has no citrullination activity. It is also of interest to determine whether PADI6 forms a stable dimer, a property of PADIs 2-4 but interestingly not PADI1 (255,256). Finally high-resolution structural information, either X-Ray Crystallographic or Cryo Electron Microscopy (Cryo-EM), will be invaluable in not only understanding the structural effects of the variation in PADI6 amino acid sequence, but also providing explanations for the observed biochemical and biophysical characteristics.

This chapter describes efforts to answer the above questions. hPADI6 and mPADI6 were recombinantly prepared from both *E. coli* and mammalian cell expression systems. The dimerization state, ability to bind calcium, and catalytic activity were characterised. Finally, a high-resolution X-ray crystal structure of hPADI6 was solved, providing significant structural insight its biochemical behaviour.

3.2 Human PADI6

3.2.1 Initial expression and purification of hPADI6 in *E. coli*

Human PADIs 1-4 have all been efficiently expressed in bacterial cell culture previously in the Walport group. Commonly, the desired protein is expressed with a peptide tag of six histidine residues (His-tag). This tag allows initial purification by immobilised metal affinity chromatography (IMAC) where the cleared bacterial cell lysate is applied to a Sepharose resin holding immobilized divalent metal ions, most commonly Ni^{2+}. The His-tag chelates to the immobilized metal ions whilst unwanted bacterial proteins are washed away. The desired protein can then be eluted off the resin with high concentrations of imidazole. Often, to prevent having unwanted peptide extensions in the protein sequence, purification tags are cleaved off. Protease-specific target sequences can be included in between the His-tag and protein sequence, allowing proteolytic cleavage with the corresponding protease, leaving just a short extension on the protein. IMAC is commonly followed by a size-exclusion chromatography (SEC) purification step, where the IMAC eluted proteins are purified by size. SEC columns consist of a matrix containing pores of varying sizes. Proteins larger than the pores pass straight through the column and are eluted first. Smaller proteins that can penetrate the pores are delayed in relation to their size, with the smallest proteins being delayed the longest and eluting last. Protein purifications can be tracked by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE), where purified fractions are denatured and separated by size using gel electrophoresis, followed by visualisation with a protein stain such as Coomassie. Only fractions containing protein bands of the desired molecular weight (M_W) are then taken forward to subsequent purification steps.

hPADI6 was initially expressed in *E. coli* with an N-terminal His-tag and a TEV protease cleavage sequence (pMCSG_His-TEV-hPADI6). Cultures were grown and expression induced overnight at 18 ^oC before bacteria were lysed by sonication. Recombinant His-TEV-hPADI6 was subject to an initial purification by Ni^{2+} IMAC using an AKTA Pure purification system, and the N-terminal His-tag was cleaved overnight with His-tagged TEV protease (**Figure 3.1a**). The sample was re-applied to nickel beads to remove the protease, and un-cleaved protein was then eluted off the beads with imidazole to assess cleavage efficiency (**Figure 3.1b**). A large amount of protein was not successfully cleaved and re-bound to nickel beads, indicating a low cleavage

efficiency. The successfully cleaved fractions were taken forward and further purified by SEC yielding 1.1 mg of purified protein from 8 L of bacterial culture (0.14 mg/L) (**Figure 3.1c-d**).

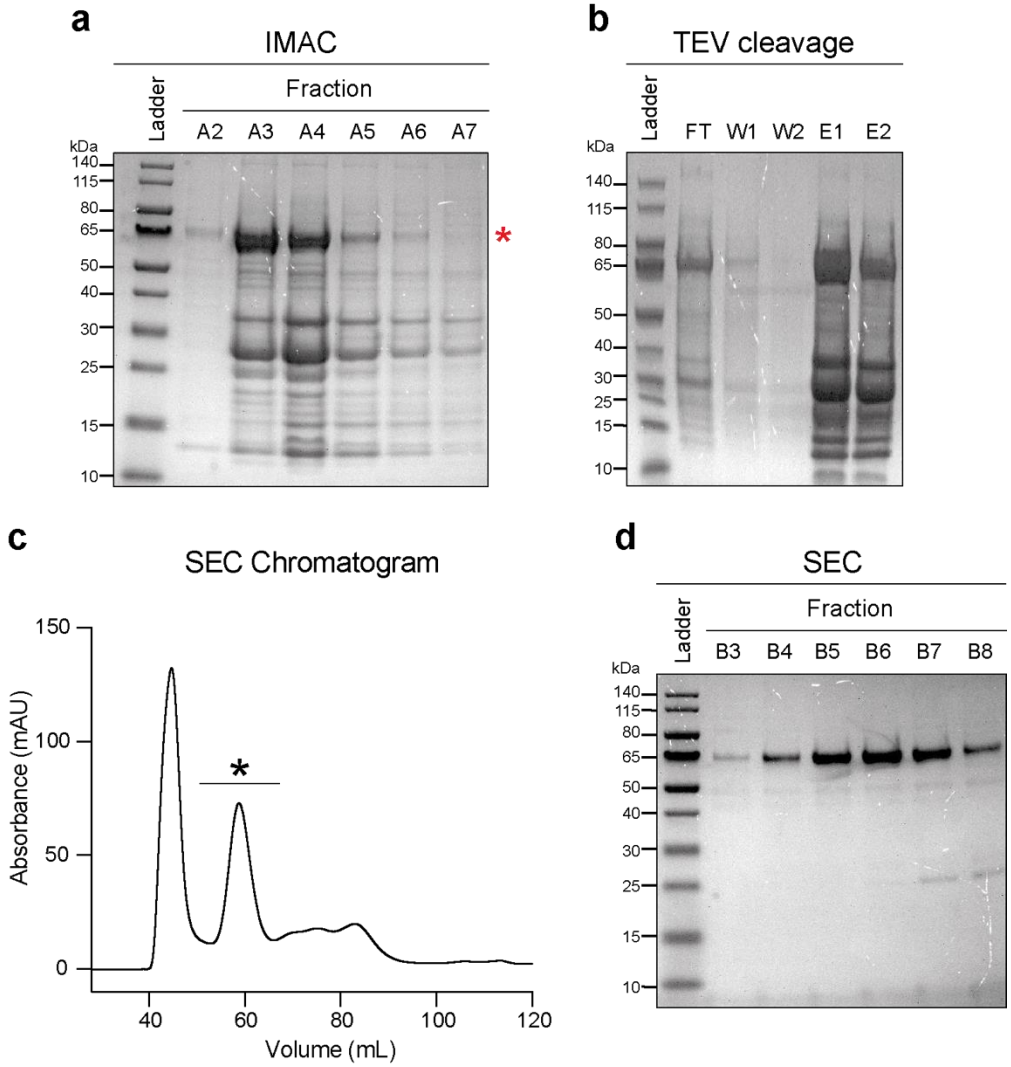


Figure 3.1 Expression and purification of pMCSG_His-TEV-hPADI6 from *E. coli*
a) SDS-PAGE gel of IMAC purification of pMCSG_His-TEV-hPADI6. Red star indicates band at approximate hPADI6 M_w. **b)** SDS-PAGE gel of His-tag cleavage by TEV protease from hPADI6. FT = flow through of cleaved protein, W1 W2 = washes, E1 E2 = elution of un-cleaved protein. **c)** Chromatogram of SEC purification of hPADI6. Fractions taken for SDS-PAGE characterisation are labelled with black star. **d)** SDS-PAGE gel of SEC purified fractions. Fractions B3 to B7 were combined and taken forward. SEC = size-exclusion chromatography.

Band-ID mass spectrometry was used to confirm that the purified protein was hPADI6. Whilst hPADI6 was present, common bacterial contaminant protein Bifunctional polymyxin resistance protein ArnA (ArnA, UniProt: P77398) was also present in high quantities. ArnA is particularly concerning due to its propensity to bind metal ions, thus being difficult to separate during IMAC, and its similar molecular

weight (74 kDa) to hPADI6 (78 kDa), rendering SEC separation challenging. Contamination by ArnA and other bacterial metal binding proteins is common, and one potential remedy is to express the desired protein in the NiCo21(DE3) *E. coli* strain engineered by NEB (257). In the NiCo21(DE3) strain, common bacterial contaminant protein ArnA, as well as the sensitive to lysis D (SlyD) protein and carbonic anhydrase 2 (Can) are tagged to allow removal by chitin affinity chromatography. Glutamine-fructose-6-phosphate aminotransferase (GlmS) is also mutated to eliminate its ability to bind to metal ions.

3.2.2 Optimisation of hPADI6 expression and purification in *E. coli*

The hPADI6 yield in the purification of pMCSG_His-TEV-hPADI6 was significantly lower than those achieved by other members of the research group when expressing other hPADIs (commonly >2 mg/L culture). Furthermore, as the Band-ID showed, the resulting protein was not of a desirable purity. One issue is likely to be the inefficient TEV cleavage. It has been reported that 3C protease can more efficiently cleave peptide tags for certain proteins (258). To improve upon the low yield of hPADI6 obtained, the TEV protease cleavage sequence was therefore replaced by a 3C-protease cleavage sequence in an attempt to improve cleavage efficiency (pMCSG_His-3C-hPADI6). Bacterial expression and purification was completed as for the previous construct, however using the NiCo21(DE3) *E. coli* strain and with tag cleavage by 3C protease. An additional chitin resin purification step was included to remove tagged bacterial contaminant proteins ArnA, SlyD and Can. His-tag cleavage with 3C protease was markedly more efficient than with TEV (**Figure 3.2b**), with an approximate final yield of 3.7 mg of purified protein from 6 L of bacterial culture (0.612 mg/L), an improvement in the final yield of approximately 4-fold. Representative SDS-PAGE gels and SEC chromatogram are displayed in **Figure 3.2a-d**.

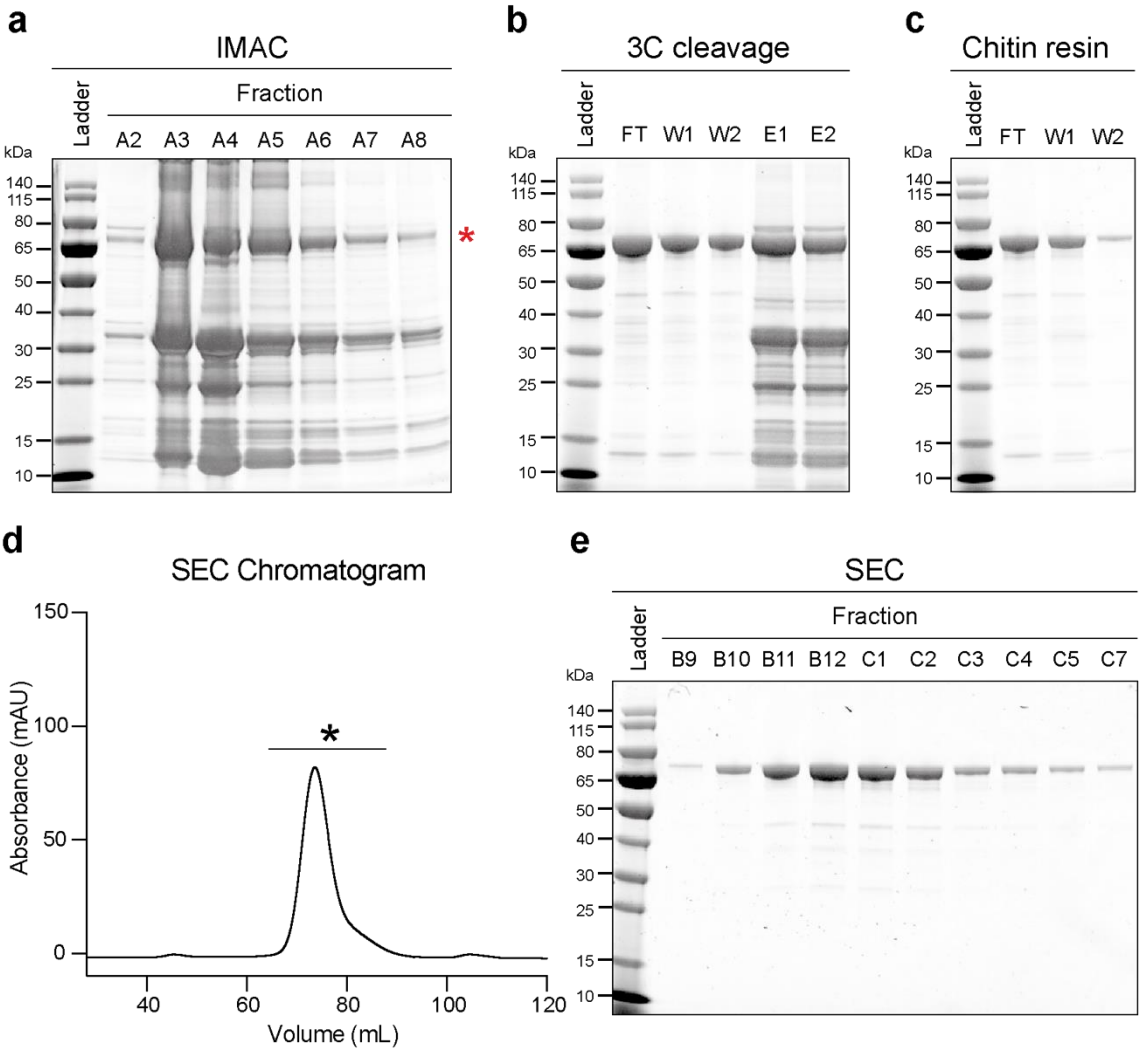


Figure 3.2 Expression and purification of pMCSG_His-3C-hPADI6 from *E. coli*
a) SDS-PAGE gel of IMAC purification of pMCSG_His-3C-hPADI6. Red star indicates band at approximate hPADI6 M_w. **b)** SDS-PAGE gel of His tag cleavage by 3C protease from hPADI6. FT = flow through of cleaved protein, W1 W2 = washes, E1 E2 = elution of un-cleaved protein. **c)** SDS-PAGE gel of sample after application to chitin resin. FT = flow through, W1 W2 = washes. **d)** Chromatogram of SEC purification of hPADI6. Fractions taken for SDS-PAGE characterisation are labelled with black star. **e)** SDS-PAGE gel of SEC purified fractions. Fractions B9 to C7 were combined and taken forward. SEC = size-exclusion chromatography.

3.2.3 Biophysical characterisation of recombinant hPADI6 from *E. coli*

Biophysical techniques are commonly employed in structural biology to elucidate key structural features such as protein secondary structure composition, stability and oligomeric state. Circular dichroism (CD) involves measuring the differential absorption of left and right circularly polarized light by a target protein (232). By using wavelengths in the far-UV range (190-250 nm), protein secondary structure

can be interrogated as peptide bonds exhibit different spectra depending on their local conformation in the protein. It must be noted however that qualitative nature of CD data is not enough to definitively define the folded state of the purified protein and only through techniques such as X-ray crystallography or Cryo-EM can we confidently determine a protein's secondary structure. Nano differential scanning fluorimetry (NanoDSF) on the other hand, is commonly used to determine protein melting temperatures, thus giving an indication of stability (259). In NanoDSF, protein samples are subjected to increasing temperatures and the fluorescence of tryptophan residues at 350 and 330 nm is measured. Tryptophan fluorescence is strongly associated with its immediate surroundings and thus as temperature increases and the protein unfolds, the 350 and 330 nm fluorescence ratio changes. The temperature at which 50% of the protein is unfolded is known as the melting temperature (T_m) of the protein. Size exclusion chromatography multi angle laser light scattering (SEC-MALLS) and mass photometry both probe the size and stoichiometry of a protein or protein complex by employing the linear, directly proportional relationship between a molecule's mass and its light scattering signal (260). In SEC-MALLS, an initial SEC step is employed to separate proteins in solution by their hydrodynamic volume, followed by passing through a MALLS detector that measures the light scattered at many angles, allowing the calculation of the absolute M_w of a sample (261). Mass photometry involves measuring the interference light scattered by the molecule and the light reflected by the measurement surface, which is directly correlated with M_w (262).

These biophysical techniques were used to investigate the structure and oligomerisation state of hPADI6 expressed in *E. coli*. CD spectra signal intensity of the purified protein was weak and noisy, potentially indicating an incompatible buffer used in the assay and as such limited information about the secondary structure of hPADI6 could be obtained (**Figure 3.3a**). NanoDSF was used to determine the T_m, measured at 48.5 ^oC, slightly higher than that of hPADI4 at 46.5 ^oC used as a control (**Figure 3.3b, Appendix 5**). The oligomeric state of purified recombinant hPADI6 was interrogated by SEC-MALLS (**Figure 3.3c**) and mass photometry (**Figure 3.3d**) and was found to be present entirely as a dimer by SEC-MALLs, and as a ratio of approximately 1:2 monomer: dimer by mass photometry. The presence of both monomer and dimer by mass photometry is likely caused by the significantly lower protein concentrations used for mass photometry experiments versus SEC-MALLS. Together, these data indicate

that recombinant hPADI6 is folded and suggests it exists predominantly as a dimer in solution.

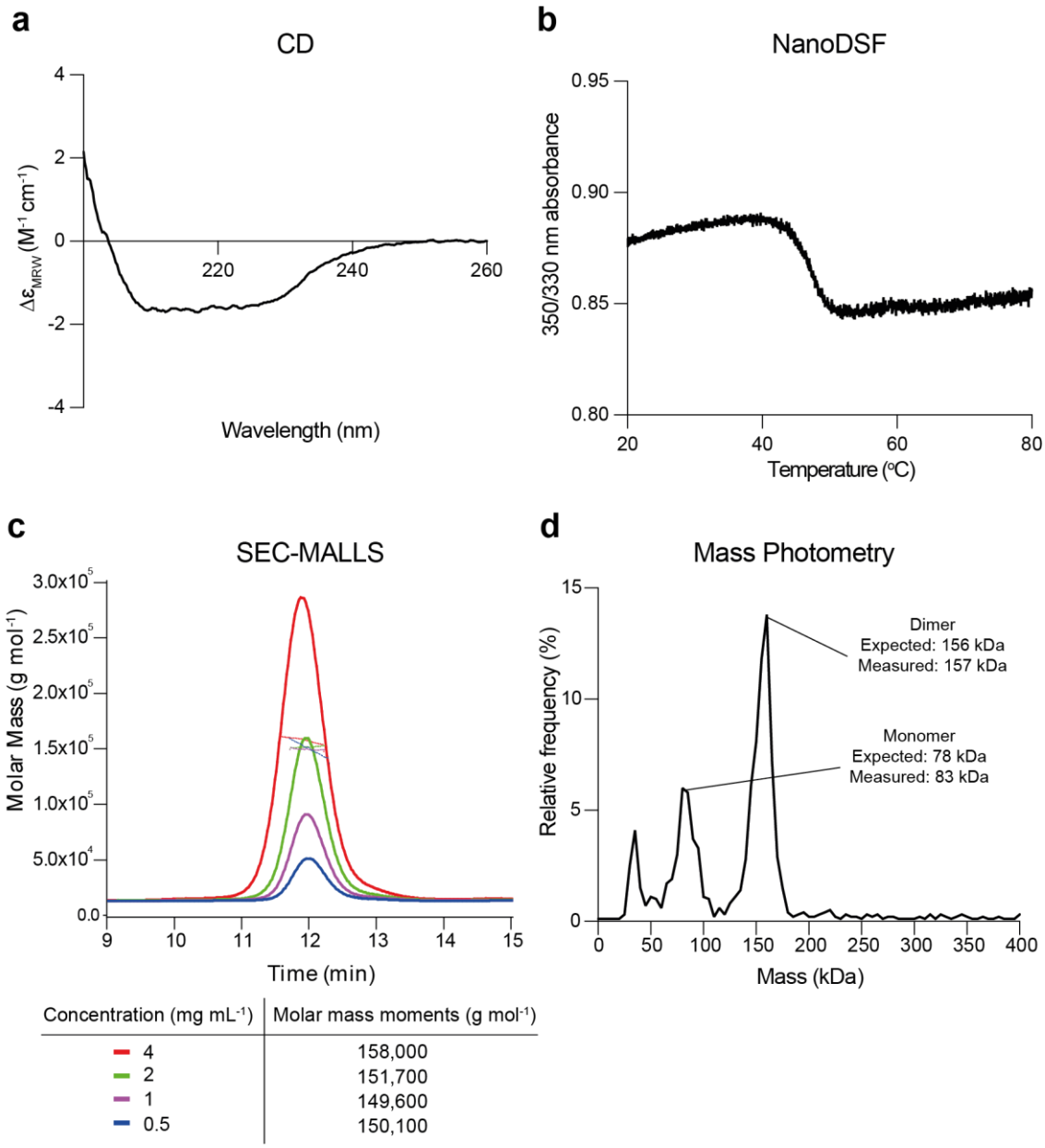


Figure 3.3 Biophysical characterisation of recombinant hPADI6 from *E. coli*
a) CD spectrum of hPADI6 purified from *E. coli*. Observed CD Signal in millidegrees was converted to mean residue CD extinction coefficient, as per (232). **b)** NanoDSF melting spectrum of hPADI6 purified from *E. coli*. Absorbance was measured at 350 and 330 nm and the ratio plotted. T_m = 48.5 °C. Experiment was performed in triplicate. **c)** SEC-MALLS spectrum of hPADI6 purified from *E. coli*. Four concentrations of hPADI6 were measured, 4 mg/mL (red), 2 mg/mL (green), 1 mg/mL (purple) and 0.5 mg/mL (blue). Calculated molar mass moments correspond to the predicted M_w of hPADI6 dimer (156 kDa). **d)** Mass photometry histogram of hPADI6 expressed from *E. coli*. (Bin size = 5 kDa). CD = circular dichroism. NanoDSF = nano differential scanning fluorimetry. SEC-MALLS = size exclusion chromatography multi angle laser light scattering.

3.2.4 COLDER assay of recombinant hPADI6 from *E. coli*

As mentioned in **Chapter 3.1**, the COLDER assay is commonly employed to investigate PADI citrullination activity. Purified hPADI6 was examined for citrullination activity in the COLDER assay using the standard PADI substrate BAEE and free arginine. The same assay conditions of calcium concentration, buffer, temperature and time were used as have been previously optimised for PADI4. A higher concentration of hPADI6 was employed in the assay, however. No citrullination was observed after one hour incubation of BAEE with hPADI6, in contrast to the significant citrullination observed with hPADI4 (**Figure 3.4a**). Similarly, no citrullination was observed using free arginine as the substrate (**Figure 3.4b**). Citrulline is detected when free arginine was incubated with hPADI4, this is likely auto-citrullination as free arginine cannot be citrullinated by hPADI4 (263). Both assays were performed with 10 mM $CaCl_2$ which activates hPADIs 1-4 *in vitro*. The absence of citrullination activity for hPADI6 is not surprising given that this has also previously been reported for mPADI6 in a COLDER assay (254). However, given the other PADIs require high concentrations of calcium for activity *in vitro*, and it is unlikely PADI6 even binds calcium ions, it is possible that hPADI6 requires alternate activating conditions that remain unknown.

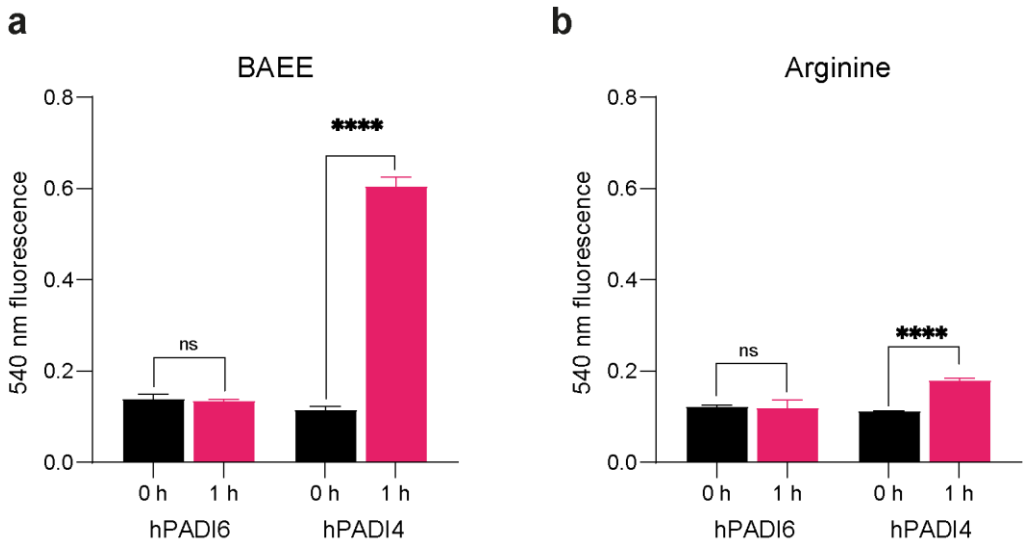


Figure 3.4 COLDER assays of recombinant hPADI6 and hPADI4
a) COLDER assay plot of hPADI6 and hPADI4 against BAEE. Reactions were performed in 10 mM $CaCl_2$ and quenched with EDTA at 0 and 1 h. hPADI6 concentration = 500 nM, hPADI4 concentration = 50 nM, BAEE concentration = 10 mM. Unpaired parametric t-test, **** < 0.0001. **b)** Same as **a)** except free arginine at concentration 10 mM instead of BAEE

3.2.5 Initial crystallisation of recombinant hPADI6 from *E. coli*

A high-resolution structure of hPADI6 would be invaluable to elucidate structural differences between hPADI6 and the other PADIs, particularly those caused by the variation in sequence. Two possible approaches can be used to generate such a structure, X-Ray crystallography and Cryo-EM. Cryo-EM is often used to interrogate large protein complexes, and it can be difficult to generate high resolution Cryo-EM structures for single proteins that are not part of a complex (264). It was therefore decided that the more commonly used X-Ray crystallography would initially be employed to produce a structure of hPADI6.

X-Ray crystallography involves the crystallisation of a target protein followed by exposure to X-rays and measurement of the diffraction pattern. A 3D electron density map can then be calculated from this diffraction pattern and the protein sequence modelled to fit the electron density. The production of high-quality crystals requires the initial screening of a wide variety of precipitant solutions covering as much chemical space as possible. If initial hit crystals are obtained, this is followed by targeted screens to optimise the crystallisation conditions, e.g., by varying precipitant concentration or the addition of other additives. Crystallisation conditions are often sensitive to the smallest changes in buffer concentration, pH, or protein concentration, and therefore it is common to have significant batch to batch variability. As such, purifying protein in high yields is often vital to ensure that crystallisation screens can be conducted from the same protein preparation.

Proteins need to be stable to form crystals and as such the storage buffer often needs to be optimised. This can be done by screening the protein of interest against a panel of buffers, with varying pH, salts, and additives to identify conditions in which the protein is most stable. Stability can be assessed by diluting the protein of interest into a range of buffers and measuring the T_m by NanoDSF. This was done using the Slice pH buffer screen to determine optimal hPADI6 buffer conditions for crystallisation. Tris buffer at pH 7.6 gave the highest melting temperature measured by NanoDSF (**Appendix 6**). Subsequently, the effect of NaCl concentration on hPADI6 stability was determined, with little difference observed in melting temperature between 50, 100 and 200 mM NaCl (**Appendix 7**). A final buffer composition of 20 mM Tris at pH 7.6 with 100 mM NaCl and 1 mM TCEP was chosen for crystallisation trials. hPADI6 was incubated in primary crystallisation trays JCSG Core 1-4, PACT Premier and Morpheus

at 4 and 20 °C at an initial concentration of 10 mg/mL, yielding hit crystals in PACT Premier well E4 (20% w/v PEG 3350, 0.2 M KSCN) (**Figure 3.5**). UV imaging confirmed that the crystals were indeed protein, however they appeared small, opaque, and irregular. A grid screen to optimise precipitant and salt concentrations, along with various additive screens were trialled, however none of the trialled conditions improved crystal size or morphology. Protein quantity became a limiting factor in trialling crystallisation conditions, and thus it was decided that the protein preparation should be further optimised to allow larger scale screening of crystallisation conditions.

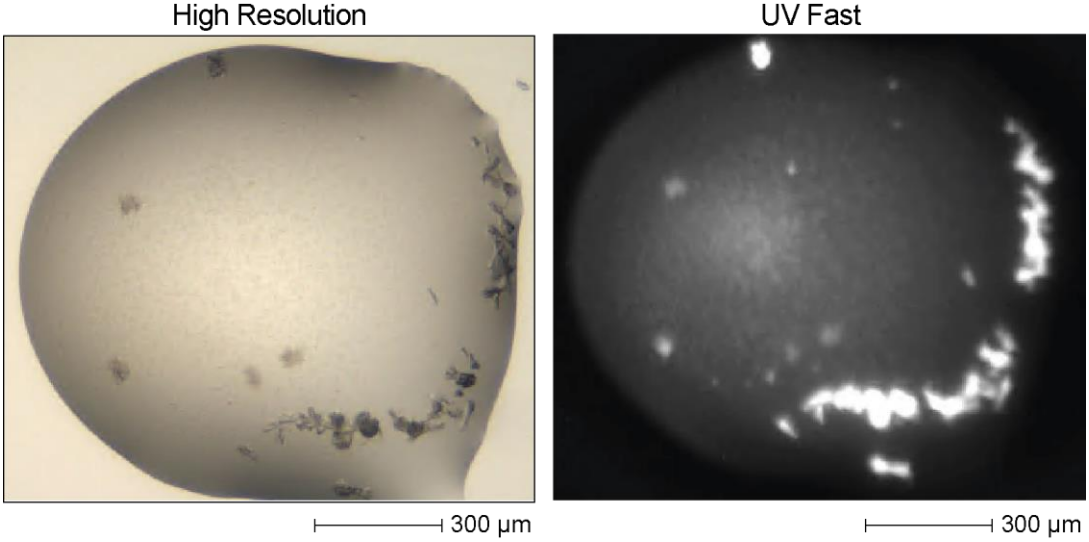


Figure 3.5. Initial crystal hit of recombinant hPADI6 purified from *E. coli*
Crystals observed in PACT Premier primary screen well E4 (20% w/v PEG 3350, 0.2 M KSCN). Left panel, high resolution image of droplet. Right panel, UV fast image of droplet indicating observed crystals are protein.

3.2.6 Optimisation of hPADI6 expression in *E. coli* using pET28a vector

The hPADI6 coding sequence was sub-cloned into a pET28a bacterial expression vector with an N-terminal His-tag and no cleavage site (pET28a_His-hPADI6). By eliminating the tag cleavage step, it was hypothesised that higher protein yields could be achieved as inefficient tag cleavage had been a problem during previous hPADI6 purifications. A trial expression comparing the pMCSG_His-3C-hPADI6 and pET28a_His-hPADI6 constructs showed markedly increased expression of a protein with an approximate M_w of 75 kDa when using the pET28a construct, which was confirmed as hPADI6 by western blot (**Figure 3.6**).

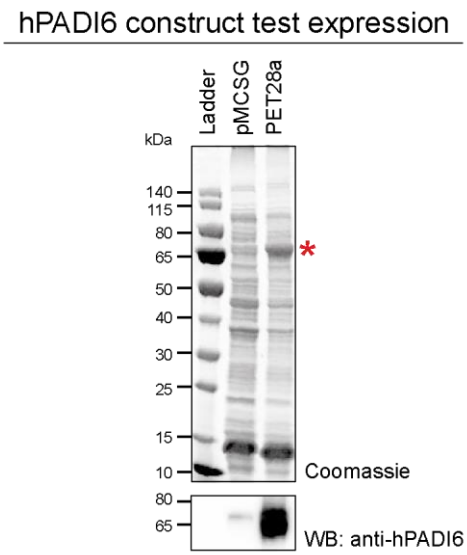


Figure 3.6. hPADI6 test expression comparison of pMCSG_His-3C and pET28a_His constructs
Top Panel: Coomassie stained SDS-PAGE. Bottom Panel: western blot against hPADI6. A large hPADI6 band is present in that pET28a sample at approximately 70 kDa that is absent in pMCSG sample. Red star indicates band at approximate hPADI6 M_w.

pET28a_His-hPADI6 was expressed following a similar procedure as for the previous two constructs but eliminating the cleavage step. This yielded 17 mg of protein after gel filtration from 6 L bacterial culture (**Figure 3.7a-d**). The protein, however, was significantly less pure than obtained when using the pMCSG_His-3C-hPADI6 construct, likely due to the absence of the tag cleavage step and subsequent additional nickel purification present in previous preparations. To remedy this, 5 mg of the SEC purified protein was dialysed into a zero-salt buffer and further purified by ion-exchange chromatography (IEX), yielding 3.2 mg of significantly purer protein (**Figure 3.7e**).

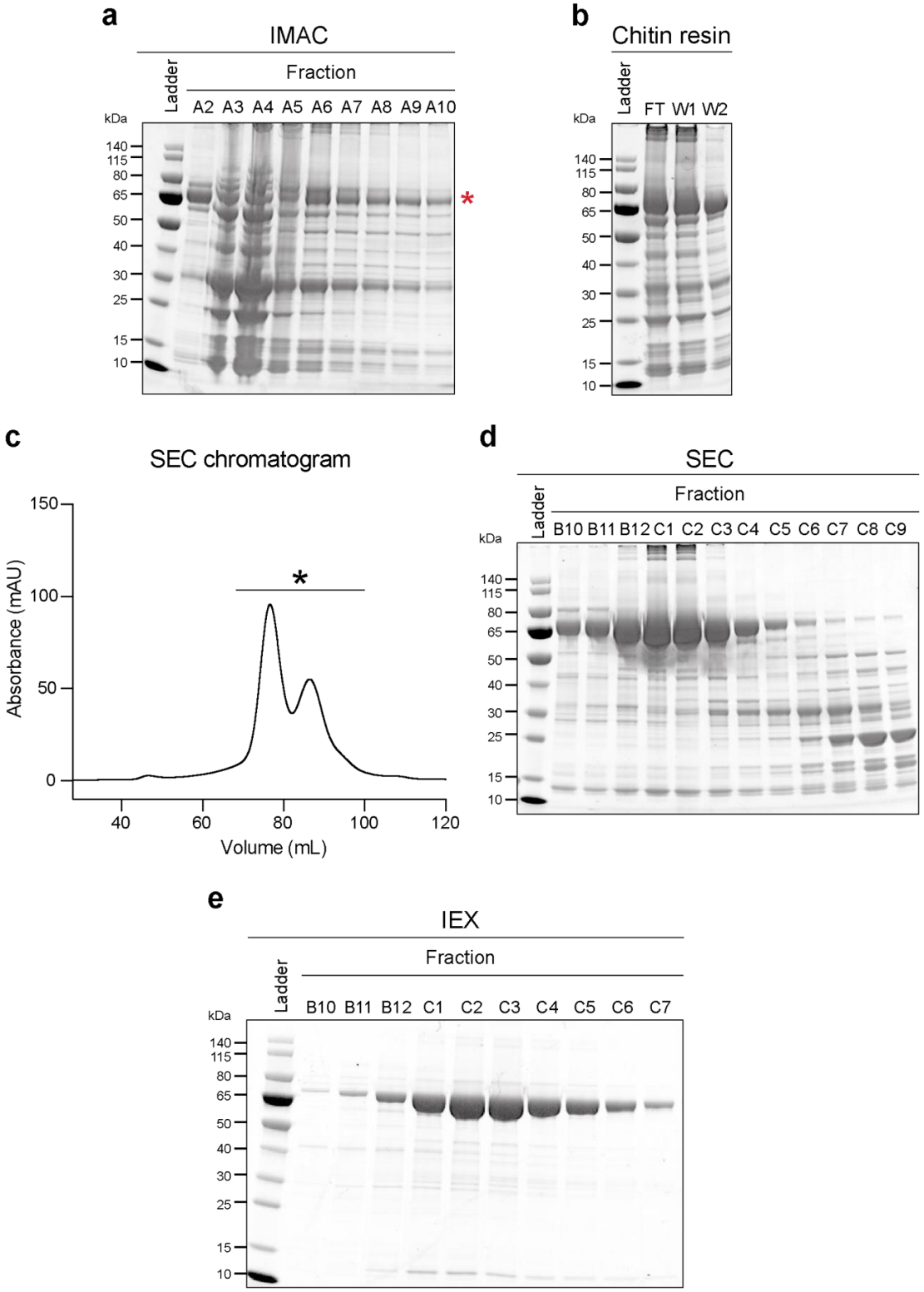


Figure 3.7. Expression and purification of pET28a_His-hPADI6 from *E. coli*
a) SDS-PAGE gel of IMAC purification of pET28a_His-hPADI6. Red star indicates band at approximate hPADI6 M_W. **b)** SDS-PAGE gel of sample after application to chitin resin. FT = flow through, W1 W2 = washes. **c)** Chromatogram of SEC purification of hPADI6. Fractions taken for SDS-PAGE characterisation are labelled with black star. **d)** SDS-PAGE gel of SEC purified fractions. Fractions B10 to C2 combined and taken forward. **e)** SDS-PAGE gel of final IEX

purified fractions. Fractions B10 to C7 were combined and taken forward. SEC = size-exclusion chromatography. IEX = ion-exchange chromatography.

3.2.7 Final *E. coli* expressed hPADI6 crystallisation

As before, primary crystallisation trays were trialled with this newly purified His-hPADI6 generated using pET28a_His-hPADI6, and hit crystals were noted in PACT Premier well E4, the same conditions as the previous hit. However, a grid screen to expand this condition again failed to produce higher quality crystals. Further protein preparations were completed; however, the purified protein yields were highly variable, and different protein batches were inconsistent in replicating crystal hits during optimisation. To understand these variabilities, small-scale test expressions were conducted to characterise the solubility of the hPADI6 (**Figure 3.8a**). When using both the pMCSG_His-3C-hPADI6 and pET28a_His-hPADI6 expression constructs, a large band was visible in the insoluble pellet fraction at approx. 75 kDa, indicating that the majority of expressed hPADI6 was insoluble (**Figure 3.8b-c**). In an attempt to optimise protein solubility, further test expressions were conducted with the pET28a_His-hPADI6 construct, co-expressing hPADI6 with chaperone proteins GroEL/GroES or using the ArcticExpress cell line, reported to improve yields of insoluble proteins (**Figure 3.8d-e**) (265). None of the trailed conditions succeeded in optimising hPADI6 solubility and it was therefore decided that alternate expression strategies should be investigated, namely mammalian protein expression systems.

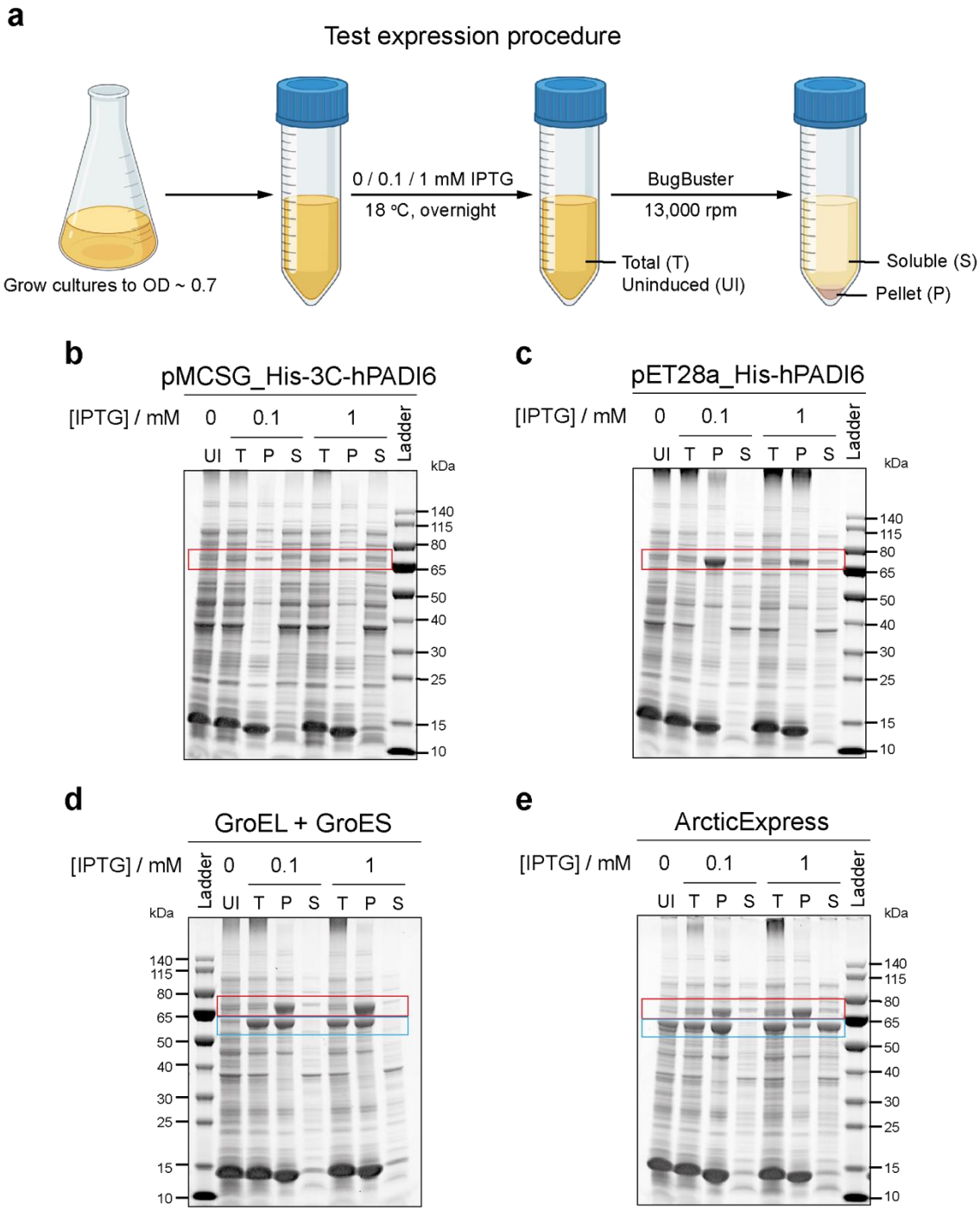


Figure 3.8. Test expressions to optimise hPADI6 solubility
a) Small-scale test expression scheme. T = Total, UI = Uninduced (equivalent to Total however when cultures were not induced with IPTG), S = Soluble, P = Pellet. **b)** SDS-PAGE gel of pMCSG_His-3C-hPADI6 test expression. **c)** SDS-PAGE gel of pET28a_His-hPADI6 test expression. **d)** SDS-PAGE gel of pET28a_His-hPADI6 test expression when co-expressed with GroEL and GroES chaperones. **e)** SDS-PAGE gel of pET28a_His-hPADI6 test expression when expressed in the ArcticExpress *E. coli* strain. Bands at approximate M_W of hPADI6 indicated by red box, bands at approximate M_W of GroEL/GroES chaperones indicated by blue box. None of the conditions tested were able to improve the solubility of hPADI6 in these expression systems.

3.2.8 Expression and purification of hPADI6 in mammalian Expi293 cells

It was decided that the Expi293 mammalian cell line would be trialled owing to its success in the McDonald lab at the Crick in producing proteins that were insoluble when expressed in *E. coli*. The hPADI6 coding sequence was sub-cloned into a pcDNA3.1_Strep-Strep-TEV vector (pcDNA3.1_SS-TEV-hPADI6). Protein expressions were completed, followed by purification by binding to StrepTactin resin, cleavage of the bound protein off the resin with TEV protease and final SEC purification, yielding approximately 5 mg protein per L of culture. Pleasingly, both the yield and purity of hPADI6 obtained from Expi293 expression were dramatically better than when expressing in *E. coli*. Intact-MS was used to confirm that the obtained protein was hPADI6 (**Appendix 8**). Representative SDS-PAGE gels tracking the purification are displayed in **Figure 3.9a-d**.

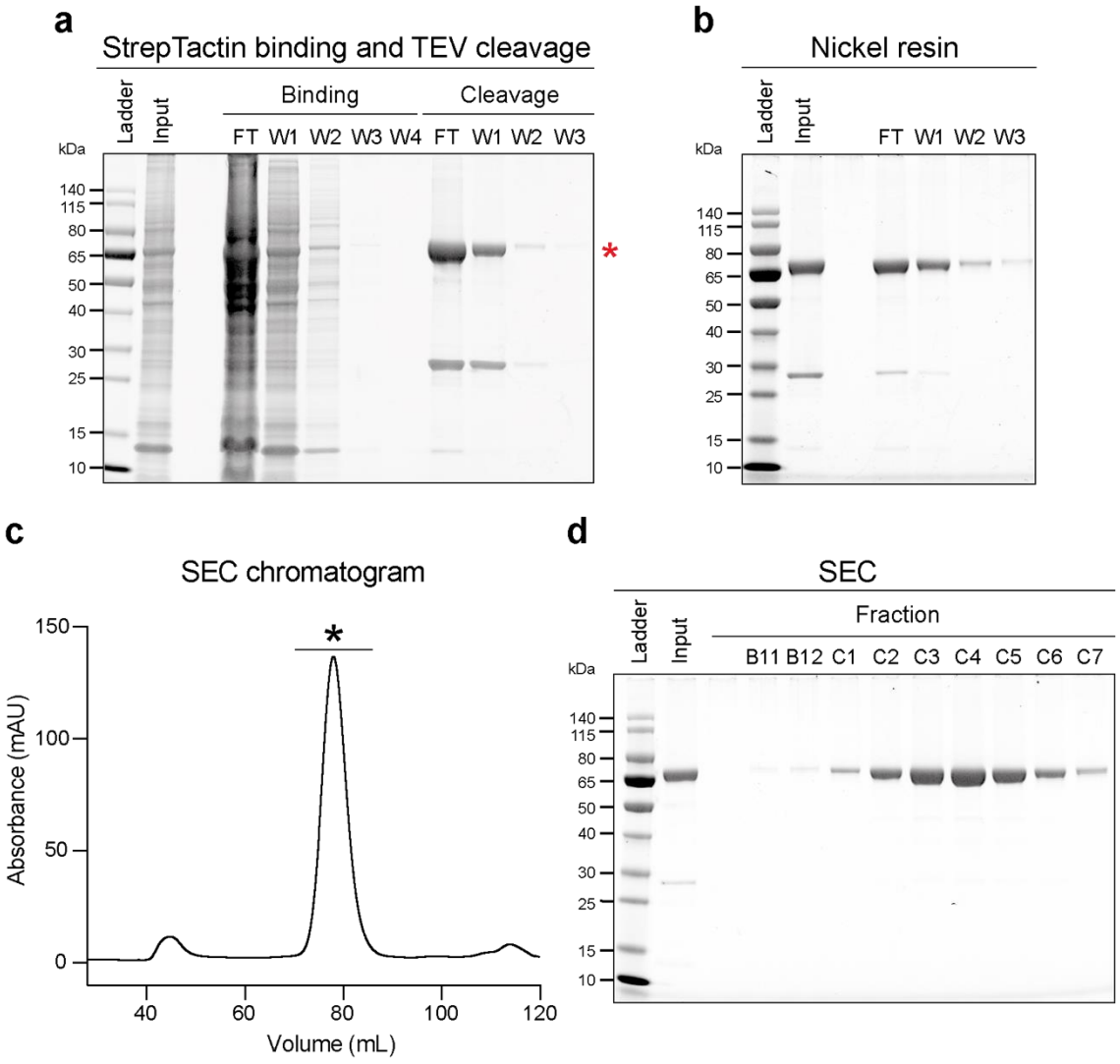


Figure 3.9. Expression and purification of pcDNA3.1_SS-TEV-hPADI6 from Expi293 cells
a) SDS-PAGE gel after binding of Expi293 lysate to StrepTactin resin, followed by on-resin cleavage of hPADI6 with TEV protease. FT = Flow through, W1-W3/4 = washes. Red star indicates band at approximate hPADI6 M_w. b) SDS-PAGE gel of sample before and after application to Nickel resin to remove TEV protease. FT = Flow through, W1-W3 = washes. **c)** Chromatogram of SEC purification of hPADI6. Fractions taken for SDS-PAGE characterisation are labelled with a black star. **d)** SDS-PAGE gel of SEC purified fractions. Fractions C1 to C7 were combined and taken forward. SEC = size-exclusion chromatography.

3.2.9 Biophysical characterisation of recombinant hPADI6 from Expi293 cells

CD, NanoDSF and mass photometry were conducted as before to investigate the stability and oligomerisation state of hPADI6 expressed in Expi293 cells. The mammalian expressed protein melted at 45.0 ^{o}C, slightly lower than the *E. coli* expressed protein, and formed a dimer in solution as previously observed (**Figure 3.10a-c**). Additionally, as with previously prepared hPADI6, the mammalian expressed protein possessed no citrullination activity against BAEE in the COLDER assay under the previously employed conditions (**Figure 3.10d**).

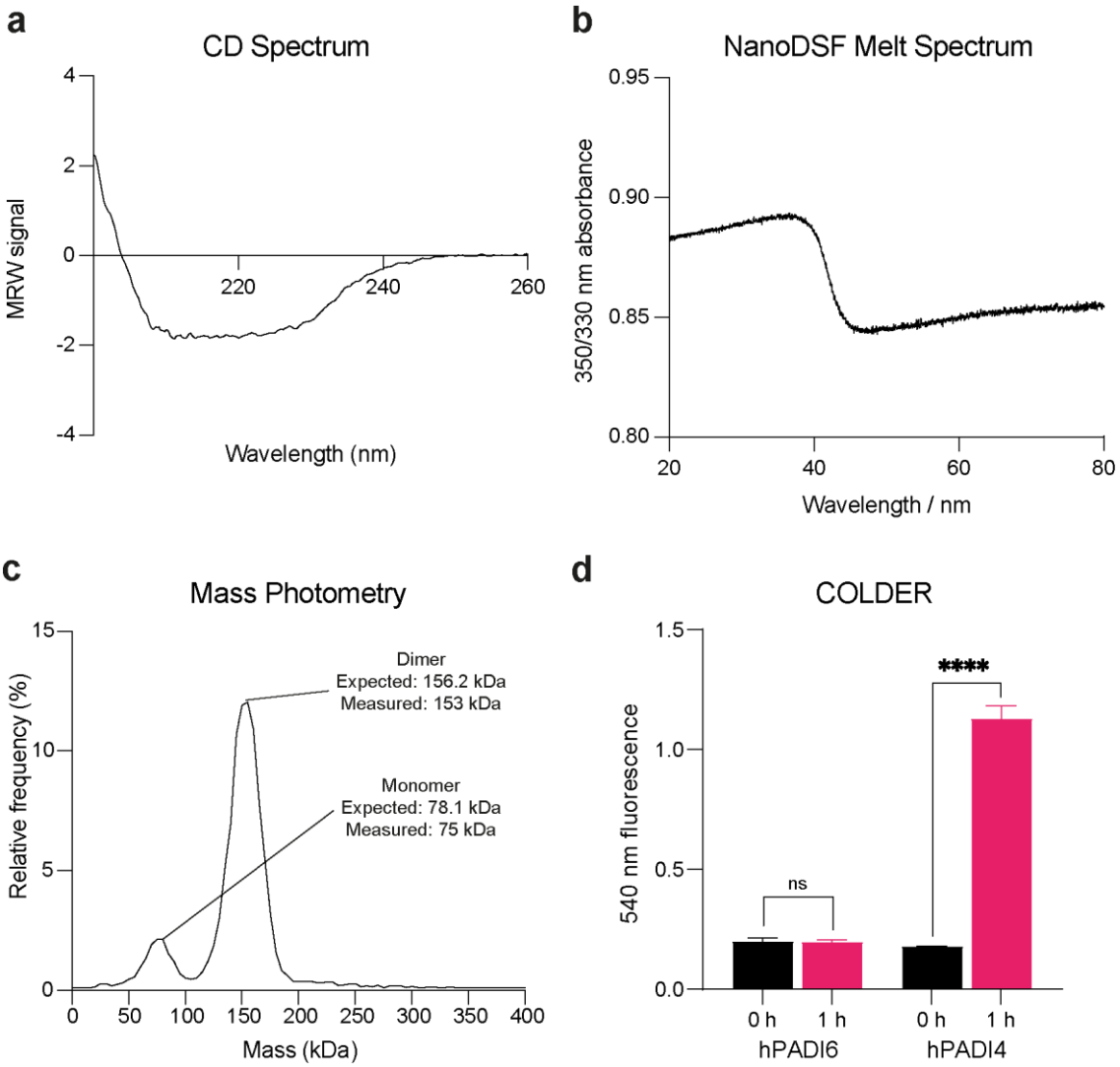


Figure 3.10. Biophysical characterisation of hPADI6 from Expi293 cells
a) CD spectrum of hPADI6 purified from Expi293 cells indicating that the protein is folded. Observed CD Signal in millidegrees was converted to mean residue CD extinction coefficient, as per (232). **b)** NanoDSF melting spectrum of hPADI6 purified from Expi293 cells. Absorbance was measured at 350 and 330 nm and the ratio plotted. T_m = 45.0 °C. Experiment was performed in triplicate. **c)** Mass photometry histogram of hPADI6 expressed from Expi293 cells showing that hPADI6 mainly exists as a dimer *in vitro*. Bin size = 5 kDa. **d)** COLDER assay plot of mammalian expressed hPADI6 and *E. coli* expressed hPADI4 (donated by Dr. Tessa Bertran of the Walport Group) against BAEE. hPADI6 does not exhibit citrullination activity. Reactions were performed in 10 mM $CaCl_2$ and quenched with EDTA at 0 and 1 h. hPADI6 concentration = 500 nM, hPADI4 concentration = 50 nM, BAEE concentration = 10 mM. Unpaired parametric t-test, **** < 0.0001. CD = circular dichroism. NanoDSF = nano differential scanning fluorimetry

3.2.10 Calcium binding of recombinant hPADI6

As previously discussed, many of the hPADI calcium binding residues are not conserved in hPADI6. As such, we have hypothesised that it is highly unlikely hPADI6 binds calcium, however, no direct experimental evidence is currently available to validate this hypothesis. It has previously been shown in the Walport group that hPADI4 calcium binding stabilises the protein, increasing its T_m by more than 20 °C. To confirm the hypothesis that hPADI6 is unable to bind calcium, NanoDSF was employed to test whether increasing calcium concentrations had any stabilising effect on hPADI6, compared with hPADI4. Increasing calcium concentrations up to 10 mM had no effect on hPADI6 melting temperature, whereas the T_m of hPADI4 was increased by up to 24.0 °C. This suggested that hPADI6 does not bind calcium ions like the other PADIs (**Figure 3.11**). It is therefore highly unlikely that calcium would have an activating effect on hPADI6 enzymatic activity and provides an explanation as to why hPADI6 is inactive in COLDER assays performed with 10 mM $CaCl_2$.

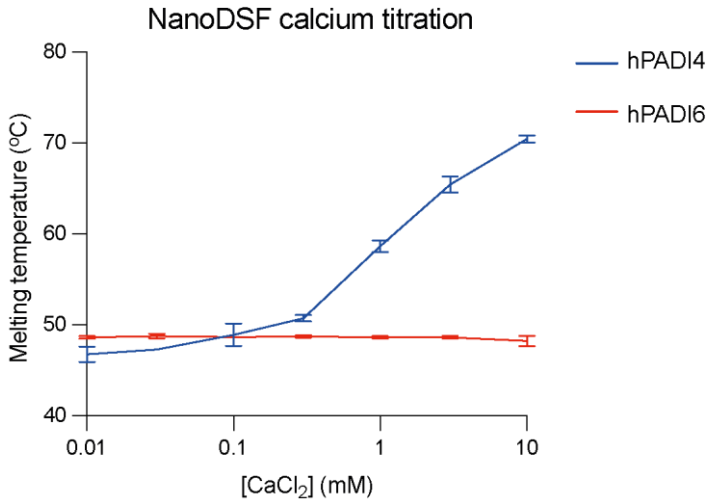


Figure 3.11 hPAD6 is not stabilised by $CaCl_2$
Melting temperatures of recombinant hPADI6 and hPADI4 in the presence of increasing calcium concentration, measured by NanoDSF. Experiment was performed in triplicate. NanoDSF = nano differential scanning fluorimetry.

3.2.11 Crystallisation of recombinant hPADI6 from Expi293 cells

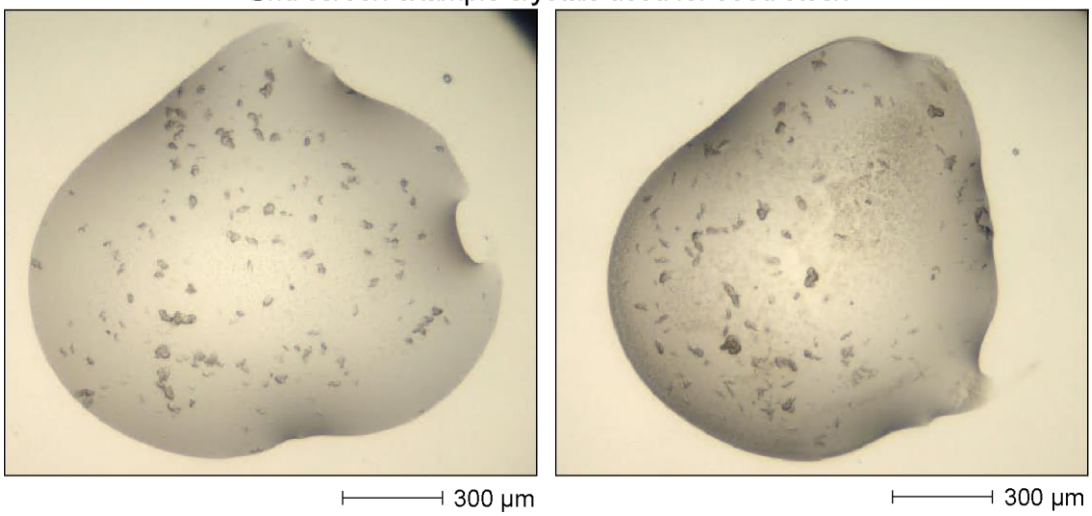
Towards crystallising mammalian expressed hPADI6, PACT Premier and PEG Ion primary crystallisation screens were trialled. These primary screens were chosen firstly to replicate crystal hits obtained with hPADI6 expressed from bacterial culture. Additionally, they were chosen due to their high PEG composition. PEG3350 was

present in the condition that succeeded in crystallising hPADI6 in **Chapter 3.2.5**, as such the high PEG composition of these screens allowed the screening of a wide variety of PEG precipitants. These trials generated crystals in PACT Premier wells G4 and H4, both containing 20% w/v PEG 3350 and 0.2 M KSCN with the addition of BIS-TRIS propane (BTP) at pH 7.5 and 8.5 respectively. Two grid screens were designed to improve crystal morphology by optimising BTP pH and % w/v of PEG 3350. At BTP pH 8.5, numerous crystals were observed in 8 drops spanning 19.1 and 21.3 %w/v PEG 3350, examples of which are displayed in **Figure 3.12a**. The observed crystals were still of too poor quality to attempt diffraction, however as large quantities of low-quality crystals could now be consistently reproduced, it was decided that micro-seed matrix seeding would be employed to improve their quality. Micro-seeding is a method used to improve crystal morphology by providing a nucleation point for crystals to grow from (266). Crystals from 8 wells were harvested by Stephane Mouilleron of the Crick Structural Biology STP and crushed using Hampton Research Seed Beads.

For the seeding experiment the PACT Premier and PEG-Ion primary screens were laid using this seed stock with a ratio of 3:1:2, protein: seed: reagent well. This produced crystals with significantly improved quality in many wells, notably PACT Premier Well F4 (**Figure 3.12b**). These crystals were larger, with a better morphology and fewer to no obvious defects. A total of 16 of these crystals were harvested by Stephane Mouilleron and sent to the Diamond Light Source in Oxford for X-ray diffraction, however none of the crystals diffracted. Following this, a grid optimisation screen centred on PACT Premier Well F4 was laid with 0.1 M BTP at pH 6.5, 0.2 M KSCN and PEG 3350 % w/v spanning 5 and 35. From this plate a further 4 crystals were obtained and tested at the synchrotron for X-ray diffraction. One crystal from optimisation well C2 with 20.957 %w/v PEG3350, 0.2 M KSCN and 0.1 M BTP at pH 6.5 diffracted to a good resolution (**Figure 3.12c**) and the structure of hPADI6 was able to be solved by Stephane Mouilleron at 2.44 Å using molecular replacement with the AlphaFold 2.0 precited hPADI6 structure as a search model. The structure parameters are displayed in **Appendix 9** and the structure is described in the following section.

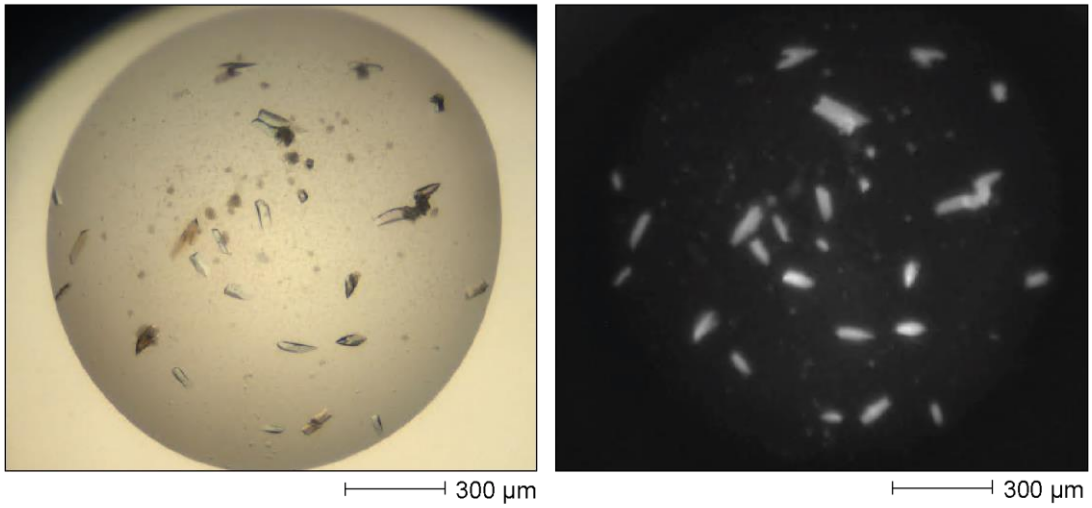
a

Grid screen example crystals used for seed stock



b

Initial seeded screen hit example in PACT Premier Well F4



c

Final crystals from seeded grid optimisation of PACT Premier Well F4

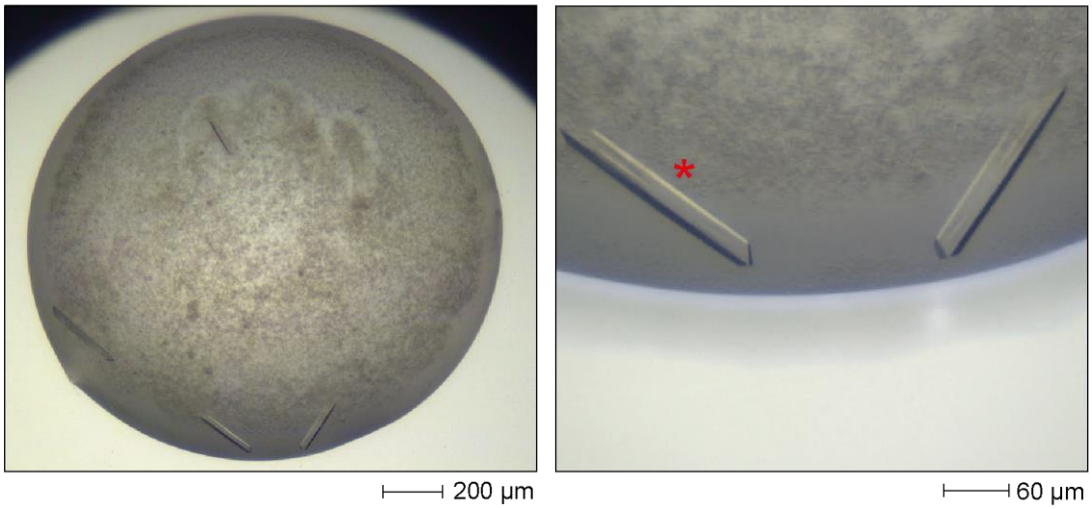


Figure 3.12. hPADI6 crystal optimisation
a) Examples of hit crystals used to produce seed stock from a grid optimisation screen of PACT Premier well G4. Left panel: 19.468% w/v PEG3350, 0.2 M KSCN, 0.1 M Bis-Tris propane, pH 8.5. Right panel: 20.745% w/v PEG3350, 0.2 M KSCN, 0.1 M Bis-Tris propane, pH 8.5. **b)** Crystal hit in PACT Premier Well F4 (20% w/v PEG3350, 0.2 M KSCN, 0.1 M Bis-Tris propane, pH 6.5) after microseed matrix seeding with hPADI6 crystals from initial primary screens. Left panel = high resolution image, Right panel = UV fast image. **c)** Final crystals from seeded grid optimisation screen of PACT Premier well F4. Crystals full size after 21 days in well containing 20.957% w/v PEG3350, 0.2 M KSCN, 0.1 M Bis-Tris propane, pH 6.5. Left panel = zoomed out high resolution image of droplet, Right panel = zoomed in image of two crystals. Crystal that diffracted to allow calculation of high-resolution crystal structure is labelled by a red star.

3.2.12 Structural analysis of hPADI6

Top and side view, ribbon and surface profiles of the hPADI6 X-ray crystal structure determined in **Chapter 3.2.11** are displayed in **Figure 3.13**. hPADI6 crystallised as a head-to-tail homodimer, consistent with mass photometry data, and as is the case for previously published hPADI2 and hPADI4 structures (27,28). As discussed in the introduction, each PADI monomer is composed of two N-terminal IgG subdomains, and a C-terminal αβ-propeller catalytic domain. Interaction contacts between monomers are predominantly between the N-terminal IgG subdomain 1 and C-terminal αβ-propeller domains of each monomer, with little contact between the middle regions. The surface of interaction between both monomers was calculated as 2075 $Å^2$ using the PDBePISA software (267). This closely matches the surface of interaction between hPADi4 monomers, reported to be 2020 $Å^2$ (28).

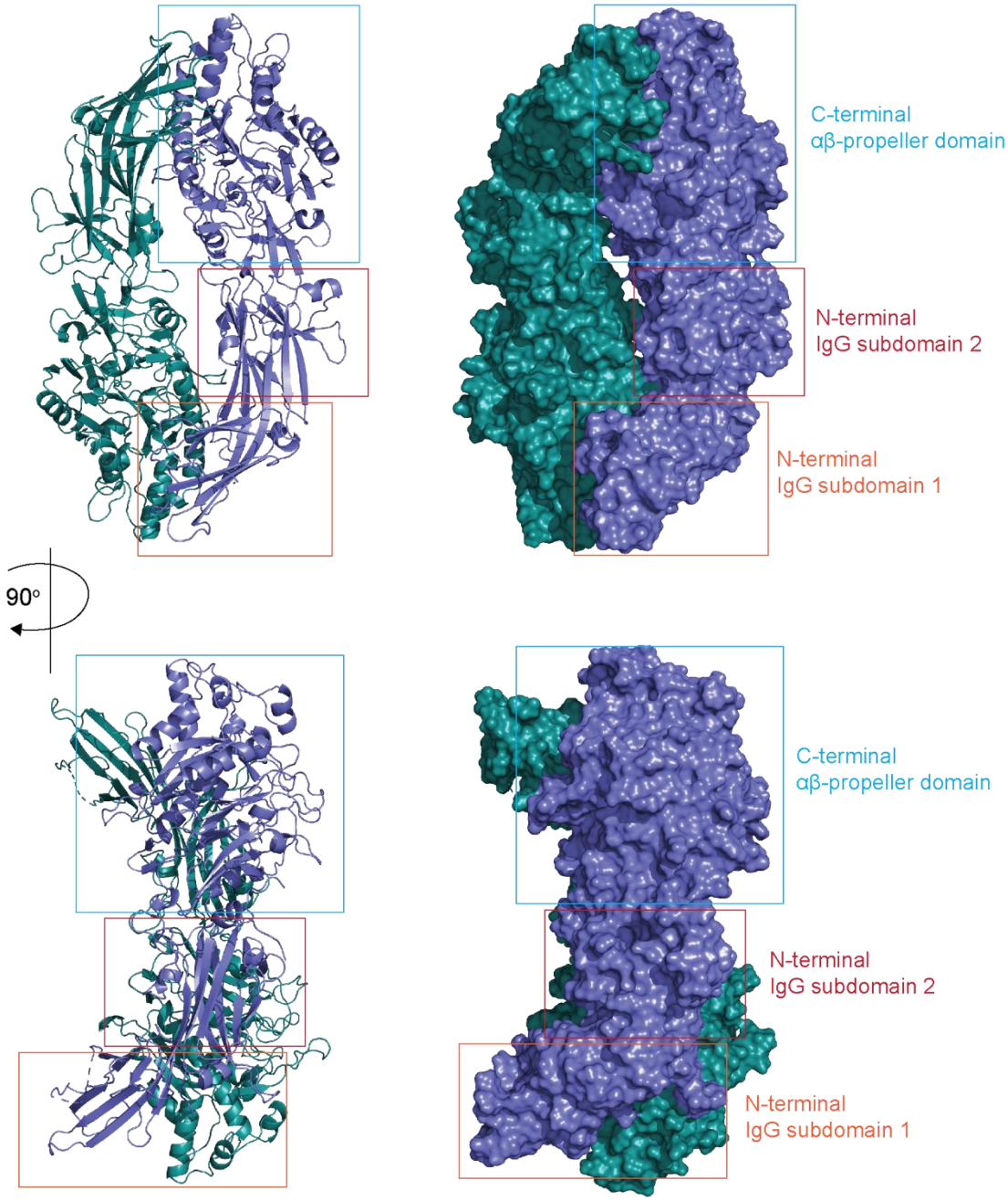


Figure 3.13 Overall hPADI6 X-Ray crystal structure
Ribbon and surface representation of the hPADI6 dimer. Top panel: Top view, Bottom panel: 90° rotation side view. Unmodelled loops from amino acids 64 to 71 and 167 to 179 represented by dashed lines. Structure analysed and images produced using PyMol. IgG = immunoglobulin.

3.2.12.1 Comparison of hPADI6 with hPADI4 structure

Calcium binding causes significant structural changes in the hPADI4 structure, resulting in the activation of the protein, as discussed in **Chapter 1.1.5.** Data presented above suggest that hPADI6 does not bind calcium ions. Indeed, no calcium ions were observed in the crystal structure. hPADI4 has been crystallised both in the absence (apo-hPADI4, PDB: 1WD9) and presence (Ca-hPADI4, PDB: 1WD8) of calcium (28). A calcium and BAEE complexed hPADI4 crystal structure has also been published, with hPADI4 possessing a single amino acid substitution, where the key catalytic cysteine (C645) has been substituted to alanine to prevent turnover of BAEE (Ca-hPADI4^{C645A}-BAEE) (28). To determine whether hPADI6 structurally aligned more with the calcium free or calcium bound hPADI4 structure, root-mean-square-distance (RMSD) values were calculated comparing to our hPADI6 structure. Root-mean-square-distance is the standard method for representing structural similarity, by calculating the average distance between backbone atoms, with a value of 0 indicating the structures align perfectly. Using PyMol, the RMSD was calculated for hPADI6 to the published apo-hPADI4 and Ca-hPADI4^{C645A}-BAEE structures, generating RMSD scores of 1.205 and 1.204 respectively. This suggests that hPADI6 is equally different from the hPADI4 calcium free and calcium bound structures. Visualising this alignment shows that the majority of alpha helices and beta sheets in the C-terminal αβ-propeller domain and N-terminal IgG subdomain 2 are fairly conserved structurally (**Figure 3.14a**). The C-terminal αβ-propeller domain of hPADI4 is made up of five ββαβ modules, all off which appear to be conserved in the hPADI6 C-terminal domain. The largest diversion with hPADI6 is in N-terminal IgG subdomain 2, most visible in the side-on view, which is angled more in hPADI6 relative to IgG subdomain 2 compared with hPADI4.

Interestingly, the structure reveals that hPADI6 has three loop insertions that are not conserved in the other hPADIs (**Figure 3.14b**). Each insertion is situated within the C-terminal domain and range from four to eleven amino acids. All of the inserted loops are on the surface of hPADI6, with loops 535 to 543, and 596 to 609 being situated around the catalytic pocket region of hPADI4 (**Figure 3.14c**). The evolutionary and functional significance of these insertions is unknown. It is possible they are involved in PPIs or, in the case of loops 535 to 543, and 596 to 609, they could conceivably be involved in substrate recognition or the regulation of catalytic activity.

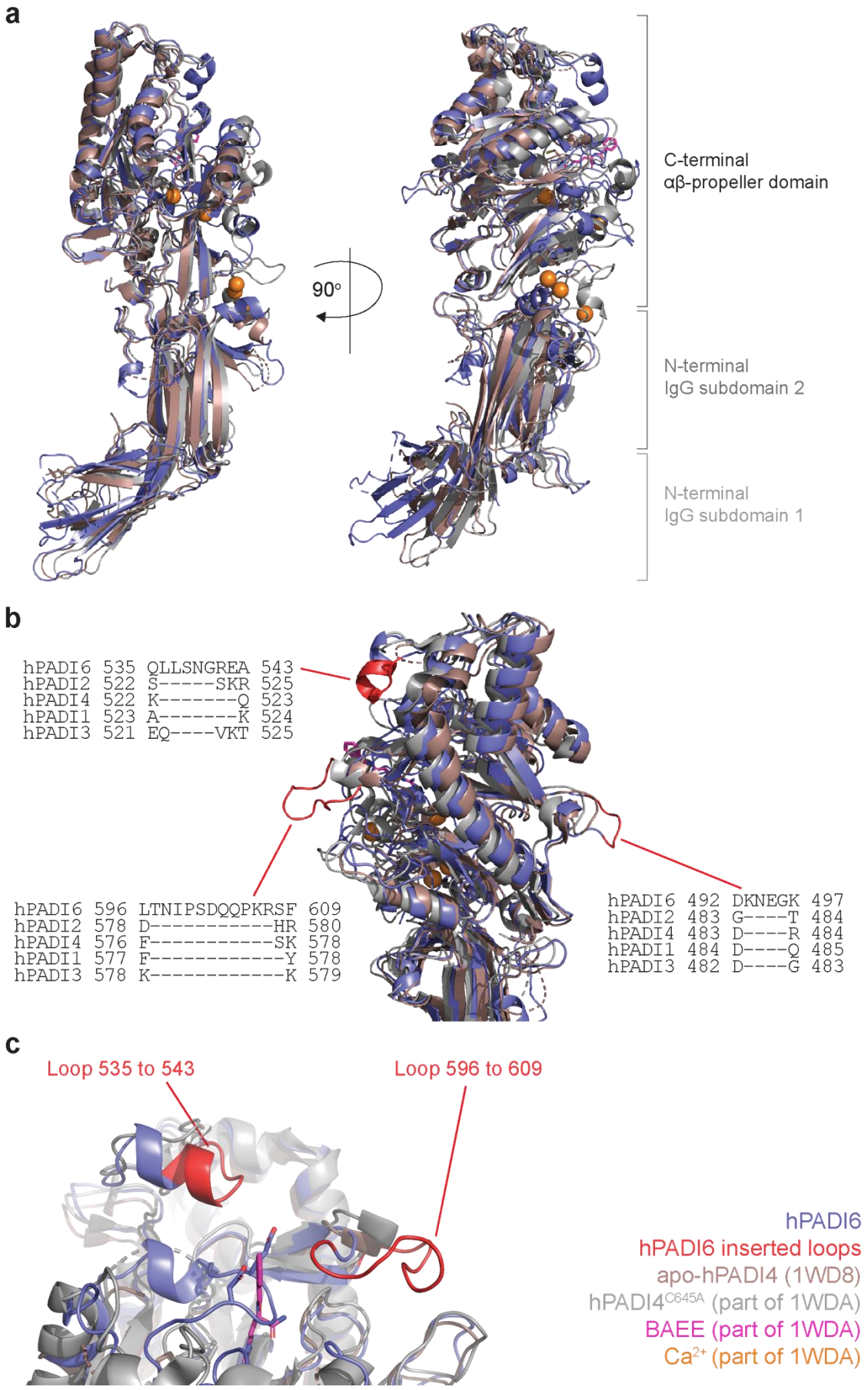

a
90°
C-terminal
αβ-propeller domain
N-terminal
IgG subdomain 2
N-terminal
IgG subdomain 1
b
hPADI6 535 QLLSNGREA 543
hPADI2 522 S-----SKR 525
hPADI4 522 K-------Q 523
hPADI1 523 A-------K 524
hPADI3 521 EQ----VKT 525
hPADI6 596 LTNIPSDQQPKRSF 609
hPADI2 578 D----------HR 580
hPADI4 576 F----------SK 578
hPADI1 577 F----------Y 578
hPADI3 578 K----------K 579
hPADI6 492 DKNEGK 497
hPADI2 483 G----T 484
hPADI4 483 D----R 484
hPADI1 484 D----Q 485
hPADI3 482 D----G 483
c
Loop 535 to 543
Loop 596 to 609
hPADI6
hPADI6 inserted loops
apo-hPADI4 (1WD8)
hPADI4C645A (part of 1WDA)
BAEE (part of 1WDA)
Ca2+ (part of 1WDA)

Figure 3.14 hPADI6 structural conservation with hPADI4
a) Ribbon representation of the top and side view superimposed structure of hPADI6 (blue), apo-hPADI4, PDB: 1WD8 (bronze), and Ca-hPADI4^{C645A}-BAEE, PDB: 1WDA (grey). Ca^{2+} shown as orange spheres. Substrate BAEE as part of 1WDA is BAEE displayed in magenta. hPADI6 IgG subdomain 1 appears more angled compared to hPADI4. **b)** Close up view of C-terminal domain of superimposed structure displayed in **a)**. hPADI6 loop insertions are coloured red and labelled with corresponding sequence alignments produced using Clustal Omega (31). **c)** Close up view of hPADI4 active site pocket of superimposed structure displayed in **a)**. hPADI6 loops 535 to 543, and 596 to 609 coloured red and sit above the pocket.

3.2.12.2 Catalytic pocket structure

Looking at the catalytic pocket of the Ca-hPADI4^{C645A}-BAEE structure, a clear pocket is formed in the surface of hPADI4 where the substrate BAEE sits (**Figure 3.15a**). Many of the loops in the apo-hPADI4 structure on the other hand are flexible and as such couldn't be modelled, suggesting that the structural changes induced by calcium binding stabilise the structure. The structure of hPADI6 in this region could be modelled, suggesting that even in the absence of calcium the structure of the hPADI6 active site region is stable. Interestingly, a surface loop in hPADI6 (670 to 673) appears to be sitting over the active site pocket, shown in **Figure 3.15a**. As confirmed by the surface view, this loop closes off the pocket (**Figure 3.15b**). Loop 670-673 is not an insertion in hPADI6, however, the sequence and structure of the loop is not conserved between hPADIs (**Figure 3.15c**). Whilst in hPADI6 the loop is sitting on top of the pocket apparently closing it off, in the Ca-hPADI4^{C645A}-BAEE structure the loop folds downwards, leaving the pocket open. The reason for this loop being folded upwards and blocking the active site in hPADI6 could be explained by one of the previously described loop insertions, specifically loop 596 to 609. Specifically, asparagine 598 (N598) points towards glutamic acid 670 (E670) and aspartic acid 673 (D673) of the blocking loop (**Figure 3.15d**). The proximity of these suggests they are forming hydrogen bonds and potentially holding the 670 to 673 loop upwards blocking the active site. This closed off active site explains the lack of catalytic activity observed in COLDER assays with this recombinant hPADI6.

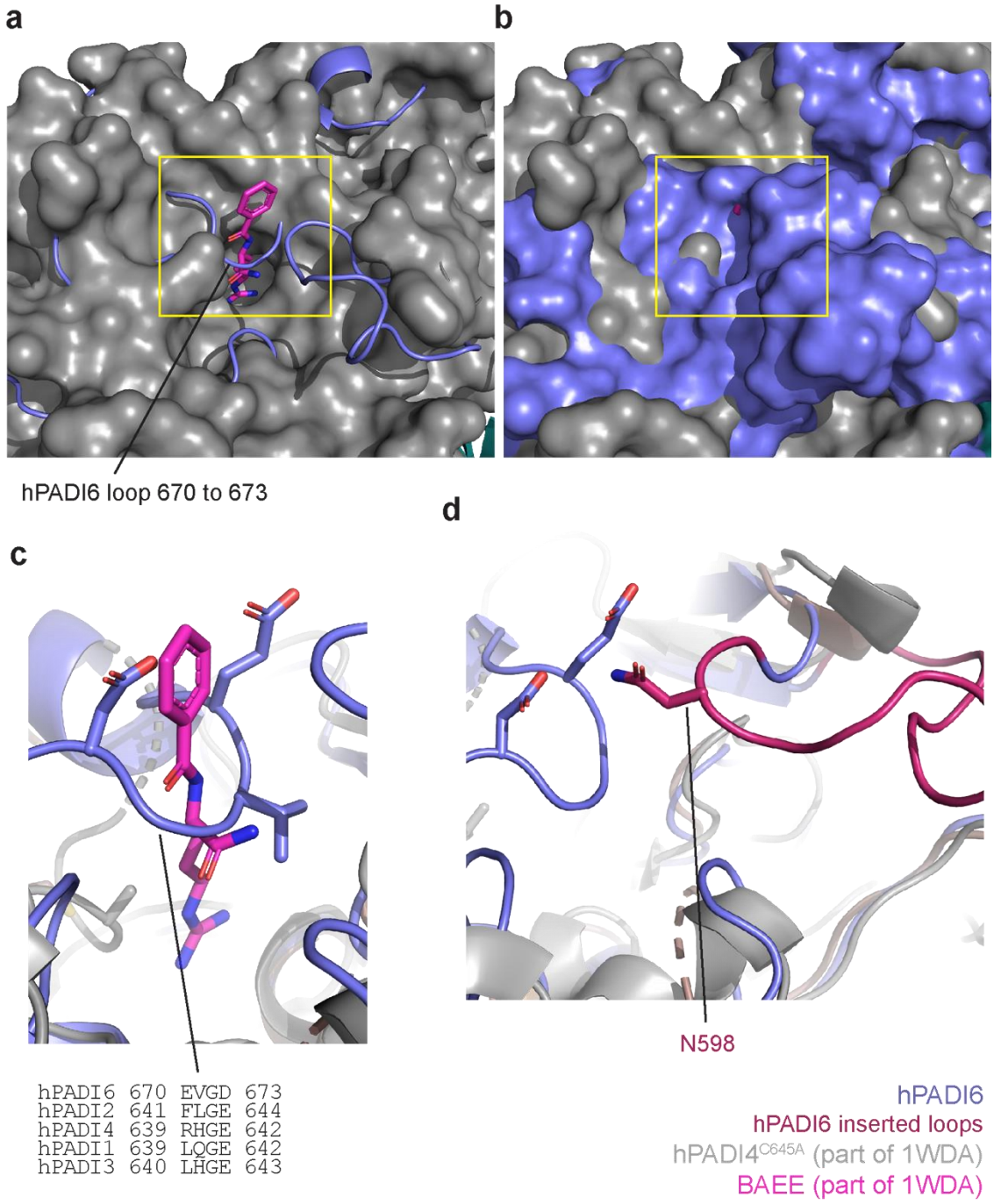


Figure 3.15 hPADI6 active site pocket structural conservation with hPADI4
a) Surface representation of the Ca-hPADI4^{C645A}-BAEE active site cleft, PDB: 1WDA (grey), superimposed with ribbon representation of hPADI6 chain A (blue) and its dimer partner chain B (blue). Substrate BAEE as part of 1WDA displayed in magenta. Active site cleft highlighted with yellow box and hPADI6 loop 670 to 673 indicated by an arrow. **b)** Same structure superposition as **a)** however surface view of hPADI6 instead of ribbon. Yellow box remains in the same position highlighting there is no active site cleft in the hPADI6 surface at that position. **c)** Close up view of hPADI6 loop 670 to 673 structure (sequence EVGD), superimposed with Ca-hPADI4^{C645A}-BAEE, PDB: 1WDA (grey). Substrate BAEE in magenta. Sequence and conservation of loop 670 to 673 highlighted, produced by Clustal Omega (31). **d)** Close up view of hPADI6 loop 670 to 673 and loop 596 to 609 (crimson). N598 labelled showing close proximity to E670 and D673 of loop blocking hPADI6 active site.

3.2.12.3 Active site amino acids

As discussed in **Chapter 1.1.4**, PADI catalytic activity relies on a key active site cysteine residue, as well as two aspartic acids and a histidine. The key cysteine conserved in hPADIs 1-4 is not conserved in hPADI6 and is instead replaced by an alanine flanked by two cysteines. This raises the question of whether one of these two cysteines may be a key residue for citrullination in hPADI6. Based on sequence conservation, only the first of these two cysteines (C675 in hPADI6) is conserved across multiple species, and thus is more likely to be the catalytic cysteine (**Figure 1.4**). However, examination of C675 in the hPADI6 structure reveals that it does not align with the hPADI4 catalytic cysteine C645 in either apo-hPADI4, and or the C645A substitution in Ca-hPADI4^{C645A}-BAEE structures (**Figure 3.16a**). In the Ca-hPADI4^{C645A}-BAEE structure the distance between the substituted C645A alanine β-carbon atom and the electrophilic carbon atom of BAEE is 3.2 Å. In the apo-hPADI4 structure C645 is situated 7.6 Å away. This indicates that calcium binding causes conformational rearrangements in hPADI4 to enable the catalytic residues to act upon the substrate. In the case of hPADI6, C675 is similarly further away to C645 in the apo-hPADI4 structure, at 7.6 Å.

Given the surface loop detailed in **Chapter 3.2.12.2** appears to be closing off the active site pocket hPADI6, if hPADI6 is catalytically active, then it is likely that significant structural rearrangements, would be required to open the active site pocket. Such structural rearrangements would likely also alter the positioning of C675. As hPADI6 does not bind calcium, this could conceivably be induced by an interaction with a co-activating protein. Without knowledge of whether this is the case, the structure informs us that the hPADI6 active site cleft is closed and the cysteine is displaced too far from where the hPADI4 substrate binds, and as such, in this conformation it is highly unlikely hPADI6 is catalytically active.

Contrastingly, the structural positioning of the histidine and aspartic acids that make up the rest of the catalytic tetrad is conserved between hPADI6 and hPADI4 (**Figure 3.17b**). In particular, hPADI6 D359 and D482 appear to align more with the Ca-hPADI4^{C645A}-BAEE than apo-hPADI4 structure.

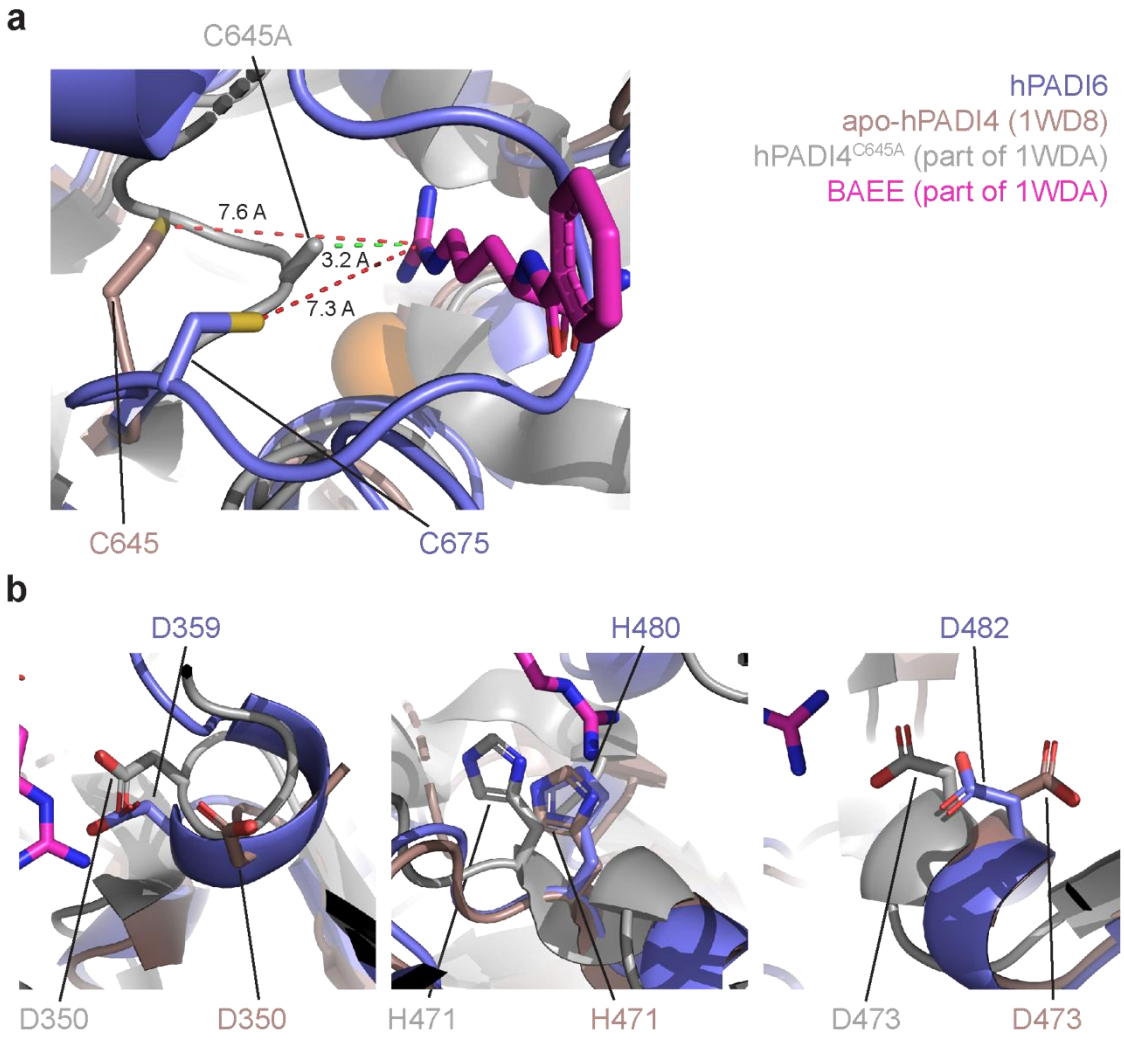


Figure 3.16 hPADI6 active site structure
a) Close up view of hPADI6_C675 (blue), superimposed with apo-hPADI4, PDB: 1WD8 (bronze), and Ca-hPADI4^{C645A}-BAEE, PDB: 1WDA (grey). Substrate BAEE in magenta. hPADI6 C675, apo-hPADI4 C645 and Ca-hPADI4^{C645A}-BAEE C645A are highlighted. Atomic distances between the thiol of hPADI6 C675, the thiol of apo-hPADI4 C645, and the β-alanine of Ca-hPADI4^{C645A}-BAEE C645A, and the electrophilic carbon of the substrate BAEE, calculated using PyMol are displayed on the figure by two red dashed lines and a green dashed line respectively. **b)** Close up views of hPADI4 catalytic site residues D350, H471, and D473, and their conserved equivalents in hPADI6, in the same superimposed structure as **a)**. hPADI6 D359 and D482 align more with the Ca-hPADI4^{C645A}-BAEE equivalent amino acids D350 and D473. hPADI6 H480 aligns more with the apo-hPADI4 equivalent H471.

We hypothesised that if hPADI6 were to be catalytically active, then it would be through C675 as it was the only one that was conserved across multiple species. Closer examination of the local structure around C675 showed there is another cysteine in close proximity, C666, with the sulphur atoms of each separated by 4.9 Å (**Figure 3.17a**). C666 is within an alpha helix that precedes C675, and hPADI4 also

has an alpha helix preceding catalytic cysteine C645 that similarly aligns with the helix of hPADI6 (**Figure 3.17b**). Interestingly, hPADIs 1-4 do not have a cysteine in this helix, and so this cysteine is unique to hPADI6. The biological function and significance of C666 is not clear, however. Whilst the electron density map shows C666 and C675 are not disulphide bonded, the protein was prepared in the presence of TCEP, a reducing agent. It is therefore conceivable that in nature these two cysteine residues could form a disulphide bond. Furthermore, if hPADI6 is a citrullinating enzyme through C675, then this could conceivably be under some form of redox control with C666. Before this can be investigated, the question of whether hPADI6 is capable of citrullination must be definitively answered.

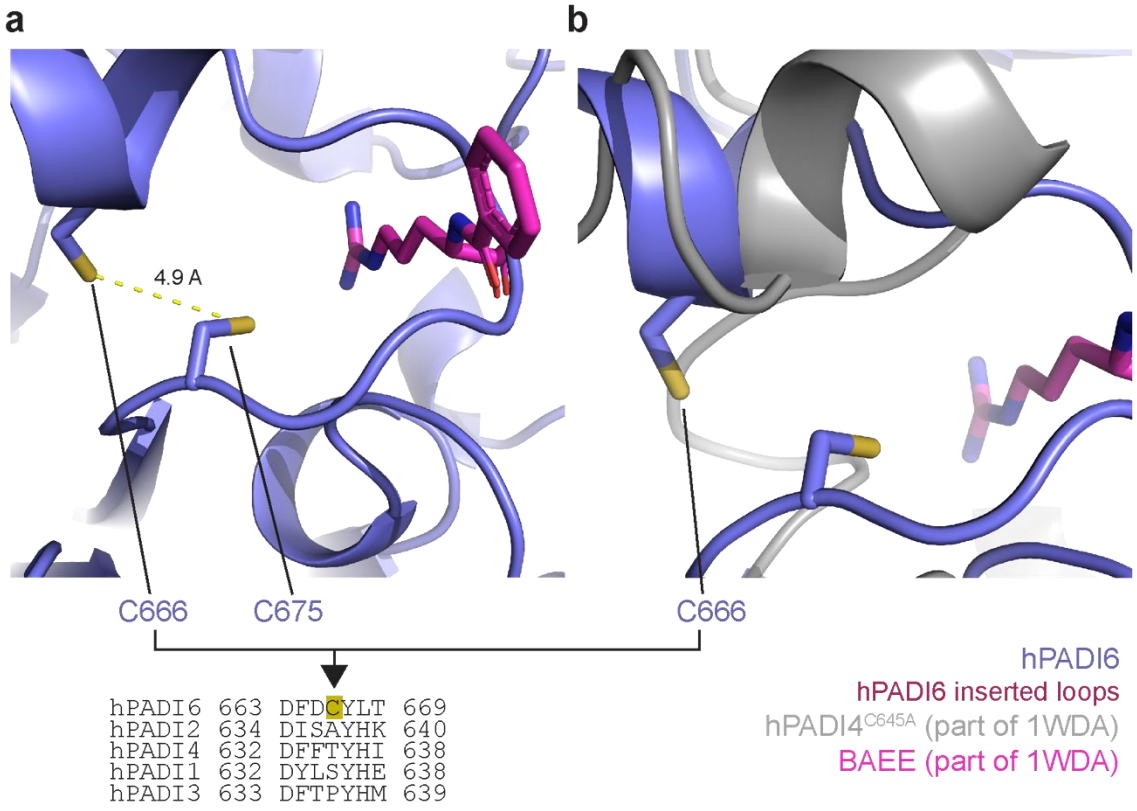


Figure 3.17 Cysteines around the hPADI6 active site
a) Close up view of hPADI6 (blue) C675 and C666. Substrate BAEE as part of 1WDA displayed in magenta. **b)** Close up view of the hPADI6 (blue) C666 α-helix, superimposed with Ca-hPADI4^{C645A}-BAEE (grey), PDB: 1WDA. Insert: conservation of the helix across hPADIs. Atomic distance calculated in PyMol. Substrate BAEE as part of 1WDA displayed in magenta.

3.2.12.4 Embryonically important variants

As discussed in **Chapter 1.2** and **Chapter 1.3**, a total of 44 pathogenic *PADI6* variants have been reported in developmental phenotypes. 29 of these result in single amino acid changes in the resulting hPADI6 protein. Having obtained a hPADI6 crystal

structure, the location and potential detrimental effects of these substitutions on the structure of hPADI6 could be interrogated. Understanding how these substitutions impact hPADI6 could be useful for elucidating its biological functions. Variants that result in frameshifts or premature stop codons would likely result in no folded protein being produced and so these variants were not investigated.

Initially, all variants that result in an amino acid change on the surface of hPADI6, or at the dimer interface were selected as these could conceivably still produce a folded protein but may disrupt key PPIs or dimerization respectively. Following analysis of the positioning and local structure of each substituted amino acid, as well as possible interactions with nearby amino acids, six variants were selected for further study. Three produced variants with a solvent exposed surface substitution, and three produced variants with substitutions at the dimer interface (**Figure 3.18a-c**). The predicted structural significance of these amino acid substitutions is discussed below:

- *H211Q*: Histidine 211 is situated at the dimer interface, and in close proximity to proline 445 (P445) on the other hPADI6 molecule (**Figure 3.18c**). Substitution of H211 to glutamine may create a steric clash between the two chains and potentially weakening the interactions between the two molecules.
- *P289L*: Proline 289 is situated in a loop between two beta sheets in the N-terminal IgG subdomain 2 at the interface between the two hPADI6 molecules. Whilst there does not appear to be an interaction between P289 and the other hPADI6 molecule, substitution to leucine could clash with arginine 607 (R607) of the other molecule (**Figure 3.18c**).
- *G540R*: Glycine 540 is situated on surface of the C-terminal αβ-propeller domain of hPADI6 (**Figure 3.18a**). Whilst the arginine conformation of the G540R substitution is not known, substitution to arginine could potentially result in the arginine side chain pointing outwards and solvent exposed. If this surface is the location of a PPI, then it is conceivable that switching the glycine for an arginine could have detrimental effects on the PPI.
- *D547N*: Aspartic acid 547 is also situated on the surface, solvent exposed and pointing out (**Figure 3.18a**). There does not appear to be any likely folding interactions on the neighbouring alpha helix. Substitution to asparagine would result in a loss of charge which could potentially reduce the strength of an interaction with another protein.

- *E586K*: Glutamic acid 586 is situated at the dimer interface pointing towards and in close proximity to histidine 16 (H16) on the other hPADI6 molecule, possibly forming salt bridge interaction between the two (**Figure 3.18b**). Substitution to the positively charged lysine therefore could lead to electrostatic repulsion and a weakening of the interaction.
- *N598S*: Situated on the surface, asparagine 598 looks to be solvent exposed (**Figure 3.18c**). As previously discussed, however, N598 also appeared to be involved in H-bond interactions with E670 and D673 on loop 670 to 673, holding the loop upwards closing the active site cleft. Whether it is involved maintaining the structure of the loop blocking the active site cleft or an interaction is unclear.

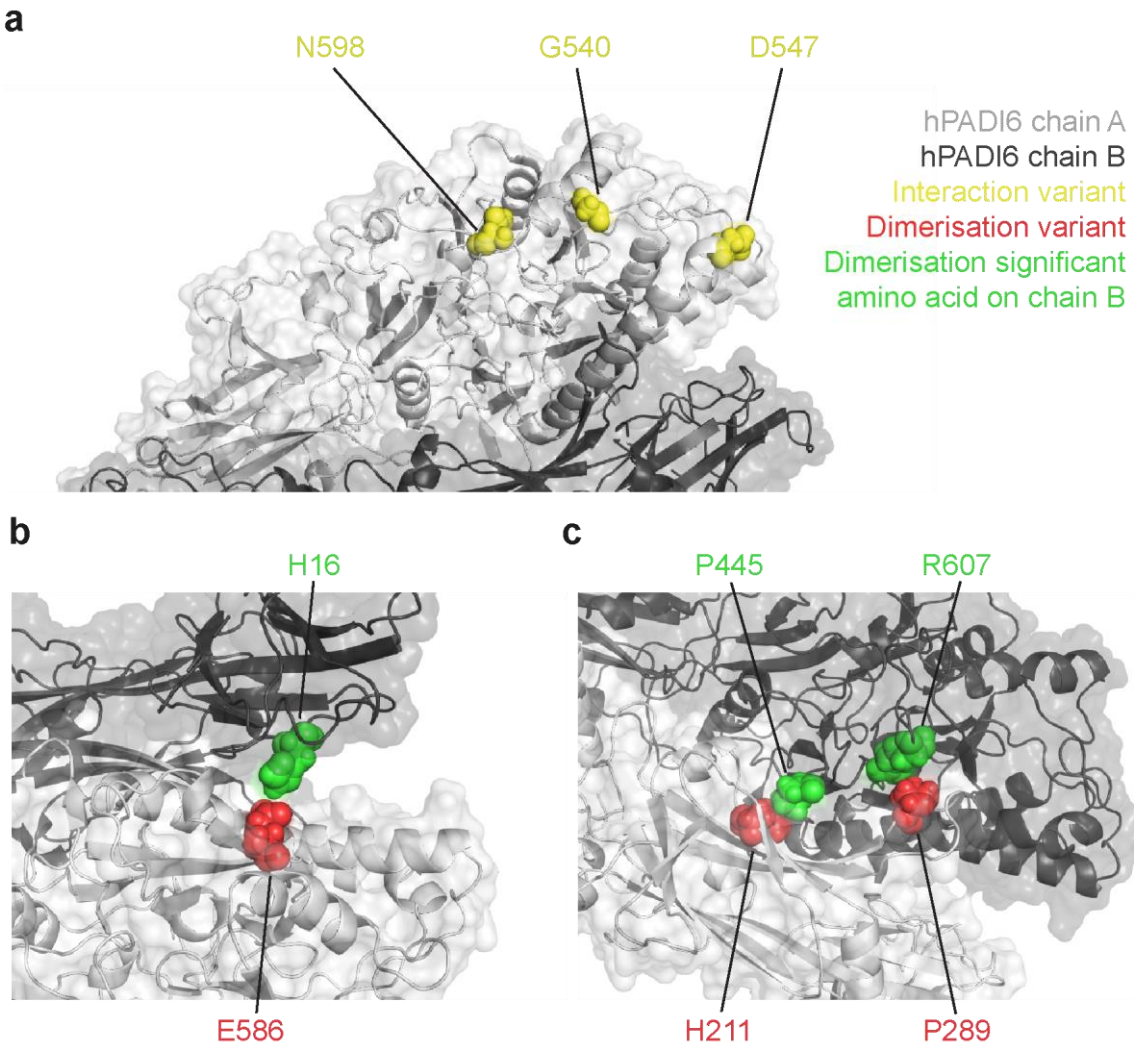


Figure 3.18 Local structure of selected hPADI6 dimerisation and interaction variants
a) hPADI6 structure with predicted interaction variant amino acids G540, D547 and N598 highlighted as spheres in yellow. **b)-c)** hPADI6 structure with predicted dimerisation variant amino acids H211, P289 and E586 highlighted as spheres in red. Dimerisation significant amino acids on chain B, predicted to be important for dimerisation, or blocking variant dimerisation (H16, P445, R607) highlighted green.

In addition to of these six structural variants, a further five variants were chosen for comparison substituted amino acids that were buried in the structure and thus likely to be important for protein stability. Additionally, a premature stop codon variant, W356*, was chosen. The premature stop codon after W356 falls in the region between the second N-terminal IgG subdomain and the C-terminal domain of hPADI6, and it is therefore conceivable that the N-terminal IgG subdomains could still be produced. In total, 11 substitutions and one truncation were selected for further characterisation (**Table 3.1**).

Table 3.1 hPADI6 variants selected for further characterisation
Selected hPADI6 variants, phenotypes and predicted significance based on the hPADI6 structure. Structural variants predicted to be involved in PADI6 dimerisation are coloured red, or potential PPIs in yellow.

PADI6 variant	Phenotype	Predicted significance
H211Q	Early embryonic arrest	Dimerisation
P289L	Early embryonic arrest	Dimerisation
W356*	MLID	Protein stability
L375S	MLID	Protein stability
M477V	MLID	Protein stability
G540R	Early embryonic arrest	Interaction
D547N	MLID	Interaction
E586K	Early embryonic arrest	Dimerisation
N598S	Miscarriages and HM	Interaction
P632A	MLID	Protein stability
P632L	Early embryonic arrest	Protein stability
P632T	Early embryonic arrest	Protein stability

Site directed mutagenesis was employed to produce FLAG-tagged mammalian expression plasmids for each of the selected hPADI6 variants (hPADI6^{variant}). Each variant plasmid was transfected into HEK-293T cells, and their expression level probed via western blot and compared to WT hPADI6 (**Figure 3.19**).

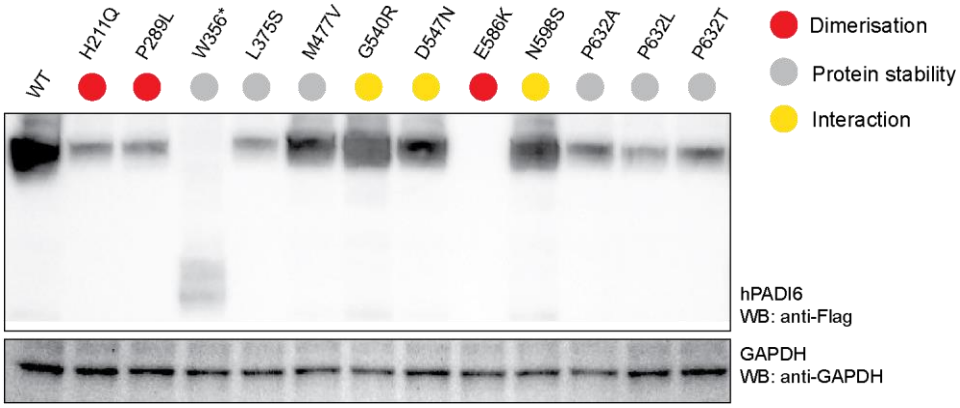


Figure 3.19 hPADI6 variant expression in HEK-293T cells
Western blot analysis of HEK293T lysates (5 μg total protein) transiently transfected with FLAG-hPADI6 and selected FLAG-hPADI6^{variant}. GAPDH was used as a loading control. Western blots were probed with antibodies indicated in figure. Experiment performed in duplicate.

All of the control variants predicted to be important for protein stability with the exception of hPADI6^{M477V}, appeared to express to a significantly lower level than WT hPADI6. This validates the hypothesis that each residue is important for the stability and correct folding of hPADI6. hPADI6^{M477V} however, expressed to a more comparable level to WT, suggesting the phenotypes associated with this variant are more nuanced. Interestingly, the truncated variant hPADI6^{W356*} was able to produce a soluble truncated protein, although to a much lower level than the WT protein, suggesting that the location of the truncation does allow the N-terminal IgG subdomains 1 and 2 to fold to some extent.

Pleasingly, the interaction variants that are predicted to disrupt PPIs, all expressed to a comparable level to wild-type hPADI6, although slightly lower. Whilst this suggests that these amino acid substitutions are not detrimental to the stability of hPADI6, full biophysical characterisation of recombinantly produced variants is required to determine their exact impact. If they are indeed similarly stable to WT hPADI6, it would be of interest to determine whether these substitutions disrupt known PPIs. As the interactome of hPADI6 has not been fully characterised, when novel interacting partners are identified, interaction assays should also be performed with these variants. This could provide valuable information on not only where the interaction surfaces are on hPADI6, but also biologically explain the developmental phenotypes associated with each variant.

Of the three variants predicted to be involved in hPADI6 dimerisation, hPADI6^{H211Q} and hPADI6^{P289L} were expressed to a lower level than WT hPADI6, and interestingly hPADI6^{E586K} did not express at all. Based on the X-ray crystal structure, none of these amino acids looked to be involved in stabilisation of the monomeric hPADI6 structure, and so the lower expression levels raise the question of whether hPADI6 dimerisation is key for its stability. Whilst it is highly unlikely that hPADI6^{E586K} could be recombinantly expressed and purified, in the future it would be of interest to determine whether recombinant hPADI6^{H211Q} and hPADI6^{P289L} are able to dimerise. To investigate the significance of E586 for dimerisation, a variant could be produced with both the E586K substitution, as well as substituting its predicted interacting amino acid on the other monomer H16 to glutamic acid or aspartic acid to determine whether this rescues dimerisation.

Further characterisation of the above variants is required to fully understand their importance to the function of hPADI6, and the varying phenotypes associated with each. A thorough investigation was not possible in this work due to time constraints. First, each variant should be recombinantly expressed and purified. CD and NanoDSF of purified protein could then provide conclusive information on the stability and folded state of each variant. Mass photometry would allow the characterisation of the dimerisation state of those variants predicted to be involved in dimerisation. Finally, whilst it appears that the variants predicted to disrupted PPIs can be stably expressed without confirmed interacting partners, we cannot know whether they do actually disrupt interactions. Towards this, the interactome of hPADI6 must be elucidated, and interactions with each variant tested.

3.3 Mouse PADI6

Alongside expressing hPADI6, it was also of interest to express and purify mPADI6. This would allow direct comparison of the biophysical properties of the two, and to interrogate whether the structural features of hPADI6 caused by a lack of conservation, are conserved in mPADI6. Furthermore, given mouse models are commonly used to investigate early embryo development, it was important to identify any conserved and non-conserved structural features between hPADI6 and mPADI6, to better correlate with any work performed in the mouse embryo.

3.3.1 Initial expression and purification of mPADI6 in *E. coli*

mPADI6 was initially expressed with an N-terminal His tag and 3C-protease cleavage site (pMCSG_His-3C-mPADI6) in the NiCo21(DE3) *E. coli* strain, using the same conditions as for pMCSG_His-3C-hPADI6. This yielded 0.17 mg from 4 L of culture (0.043 mg/L), over 14-fold lower than for hPADI6 in the same expression system. Representative SDS-PAGE gels and SEC chromatogram are displayed in **Figure 3.20a-e**. Band-ID mass spectrometry analysis of SEC fraction B4 confirmed mPADI6 was present, however it was the third most abundant protein, behind ArnA and DnaK. This protein was therefore not suitable for biophysical characterisation.

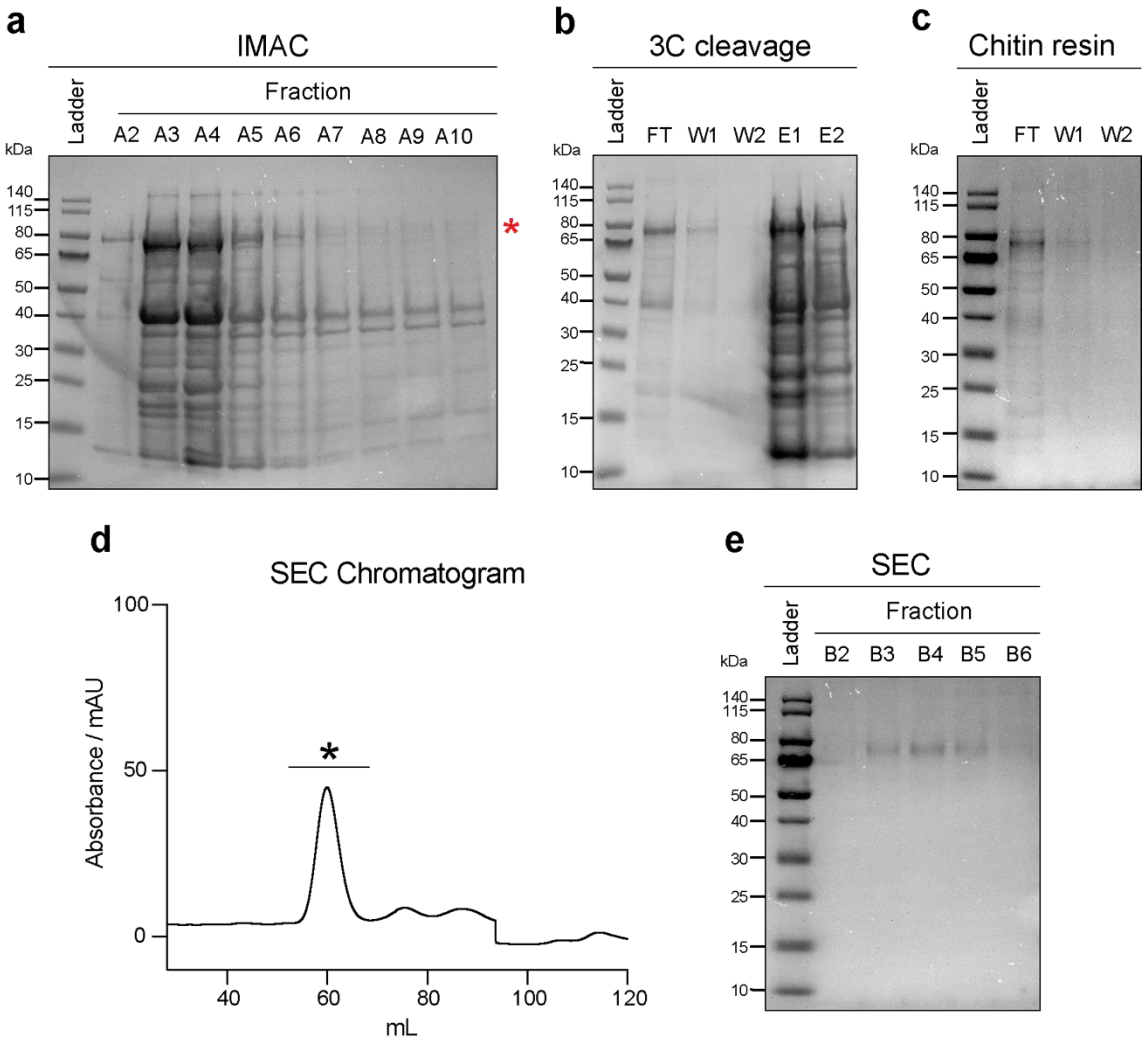


Figure 3.20 Expression and purification of pMCSG_His-3C-mPADI6 from *E. coli*
a) SDS-PAGE gel of IMAC purification of pMCSG_His-3C-mPADI6. Red star indicates band at approximate mPADI6 M_w. **b)** SDS-PAGE gel of His tag cleavage by 3C protease from mPADI6. FT = flow through of cleaved protein, W1 W2 = washes, E1 E2 = elution of un-cleaved protein. **c)** SDS-PAGE gel of sample after application to chitin resin. FT = flow through, W1 W2 = washes. **d)** Chromatogram of SEC purification of mPADI6. Fractions taken for SDS-PAGE characterisation are labelled with a black star. **e)** SDS-PAGE gel of SEC purified fractions. Fractions B9 to C7 were combined and taken forward. SEC = size-exclusion chromatography.

3.3.2 Expression and purification of mPADI6 in Expi293 cells

Due to the low quantity and purity of mPADI6 obtained from *E. coli*, alternate expression systems were required. After the success of using the Expi293 mammalian cell line to express hPADI6, mPADI6 was trialled in this system. An expression and purification following the same protocol as for hPADI6 produced 0.5 mg from 400 mL of culture (1.25 mg/L), a lower yield than hPADI6, but a sufficient quantity for biophysical characterisation. Representative SDS-PAGE gels tracking the purification and the SEC

chromatogram of the purified mPADI6 are displayed in **Figure 3.21a-d**. The purified protein was confirmed to be mPADI6 by Intact-MS (**Appendix 10**).

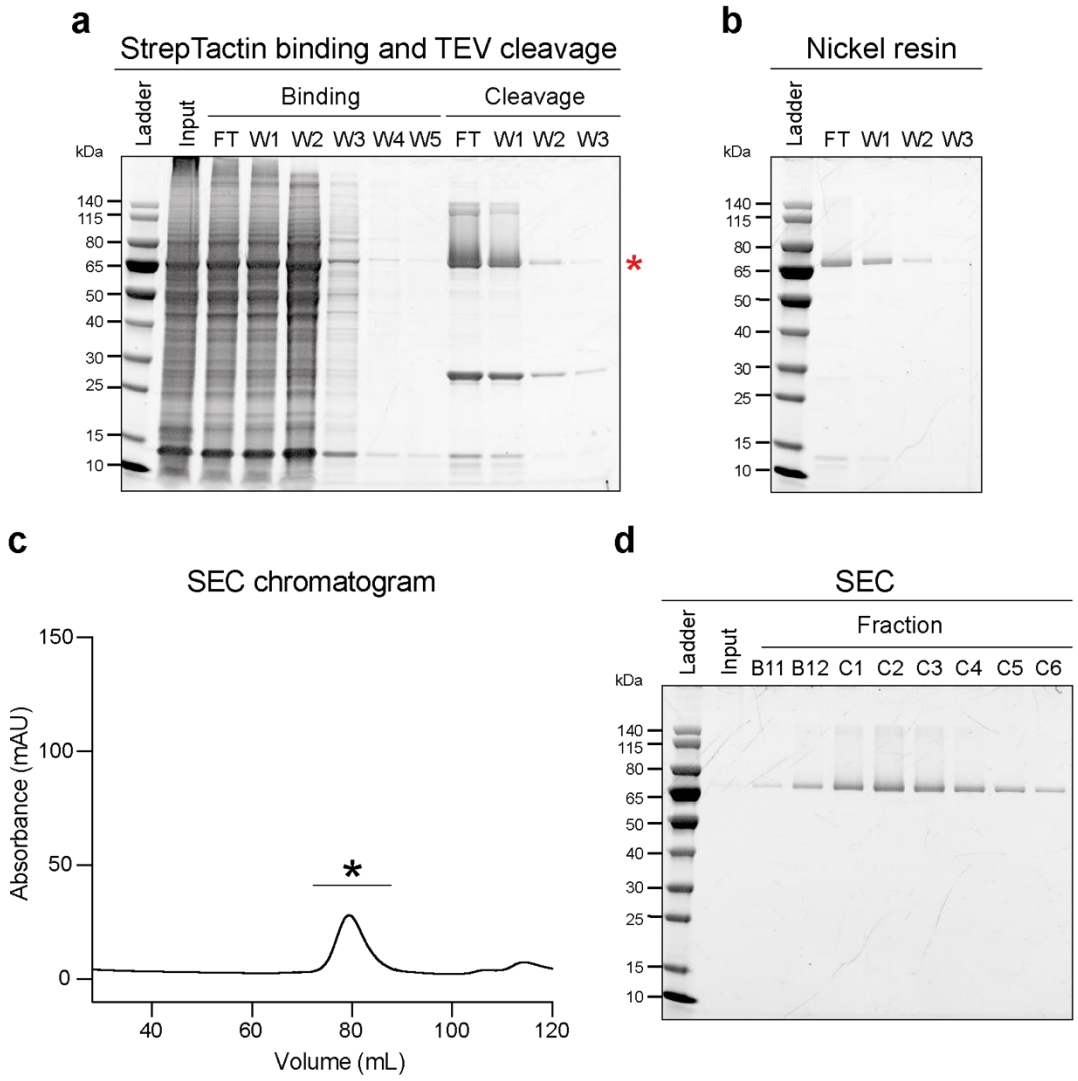


Figure 3.21 Expression and purification of pcDNA3.1_SS-TEV-mPADI6 from Expi293 cells **a)** SDS-PAGE gel after binding of Expi293 lysate to StrepTactin resin, followed by on-resin cleavage of mPADI6 with TEV protease. FT = Flow through, W1-W3/4 = washes. Red star indicates band at approximate mPADI6 M_w. **b)** SDS-PAGE gel of sample before and after application to Nickel resin to remove TEV protease. FT = Flow through, W1-W3 = washes. **c)** Chromatogram of SEC purification of mPADI6. Fractions taken for SDS-PAGE characterisation are labelled with a black star. **d)** SDS-PAGE gel of SEC purified fractions. Fractions C1 to C7 were combined and taken forward. SEC = size-exclusion chromatography.

3.3.3 Biophysical characterisation of recombinant mPADI6

The recombinant mPADI6 was then biophysically characterised to compare its stability and oligomerisation state to that of hPADI6. CD confirmed that the purified

protein possessed a similar secondary structure profile to the human protein, however, like the human protein the signal intensity was weak and limited structural information could be obtained (**Figure 3.22a**). NanoDSF determined the melting temperature to be 47.7 °C, slightly higher than but with a similar profile to the human protein (**Figure 3.22b**). Mass photometry confirmed that mPADI6 was predominantly present as a dimer, as for hPADI6, with a small amount of monomer present, likely due to the low concentrations used in the experiment (**Figure 3.22c**). Similarly, mPADI6 was inactive in a COLDER assay against BAEE, replicating previous observations in the literature (**Figure 3.22d**) (254).

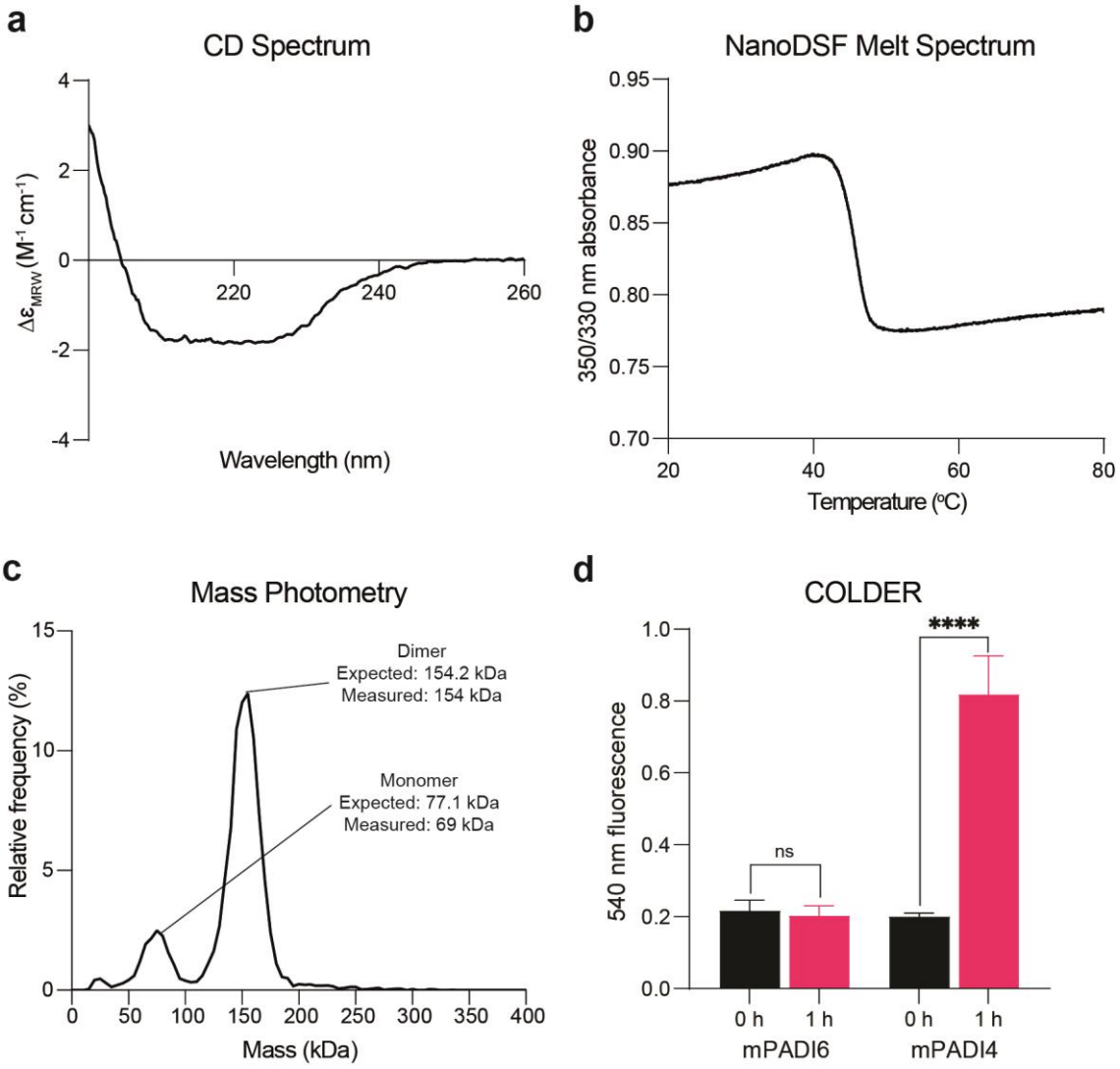


Figure 3.22 Biophysical characterisation of mPADI6
a) CD spectrum of mPADI6 purified from Expi293 cells. Observed CD Signal in millidegrees was converted to mean residue CD extinction coefficient, as per (232). **b)** NanoDSF melting spectrum of mPADI6 purified from Expi293 cells. Absorbance was measured at 350 and 330 nm and the ratio plotted. T_m = 47.7 °C. Experiment was performed in triplicate. **c)** Mass photometry histogram of mPADI6 expressed from Expi293 cells. Bin size = 5 kDa. **d)** COLDER assay plot of mammalian expressed mPADI6 and *E. coli* expressed mPADI4 (donated by Dr.

Tessa Bertran of the Walport Group) against BAEE. Reactions were performed in 10 mM $CaCl_2$ and quenched with EDTA at 0 and 1 h. mPADI6 concentration = 500 nM, mPADI4 concentration = 50 nM, BAEE concentration = 10 mM. Unpaired parametric t-test, **** < 0.0001.

3.3.4 Cryo-EM structure determination of recombinant mPADI6

Due to the difficulty in crystallising hPADI6 and the limited quantities of mPADI6 available, Cryo-EM was instead employed to produce a high-resolution structure. Cryo-EM had previously been used to generate structures of hPADI4 for the Walport group, indicating that it may also be possible for PADI6. As mPADI6 also dimerises, its larger size improves the chances of obtaining a high-resolution structure by Cryo-EM. Donald Benton of the Crick Structural Biology STP prepared Cryo-EM grids using Expi293-produced mPADI6 and acquired a dataset of 15,001 movies on a Thermo Scientific Titan Krios microscope.

Motion in acquired movies was corrected and the Contrast Transfer Function (CTF) estimated using RELION. Particles were picked in RELION using Topaz for automated picking, and then 2D classified using cryoSparc. Particles from classes with clear secondary structure were selected and an initial 3D model calculated using RELION. Selected particles were refined with RELION 3D auto-refine followed by Bayesian polishing and 3D classification of the polished particles. Particles from the top three 3D classes were extracted and transferred to cryoSPARC where they were refined using non-uniform refinement followed by local and global CTF refinement, and a final non-uniform retirement resulting in a map of 3 Å final resolution. A workflow of this processing is displayed in **Figure 3.23**. Finally, local resolution was estimated, maps filtered by local resolution and global sharpening implemented, all in cryoSPARC. Due to time constraints, the mPADI6 sequence could not be fitted to the produced electron density map, and so modelling the structure is a future aim of the work.

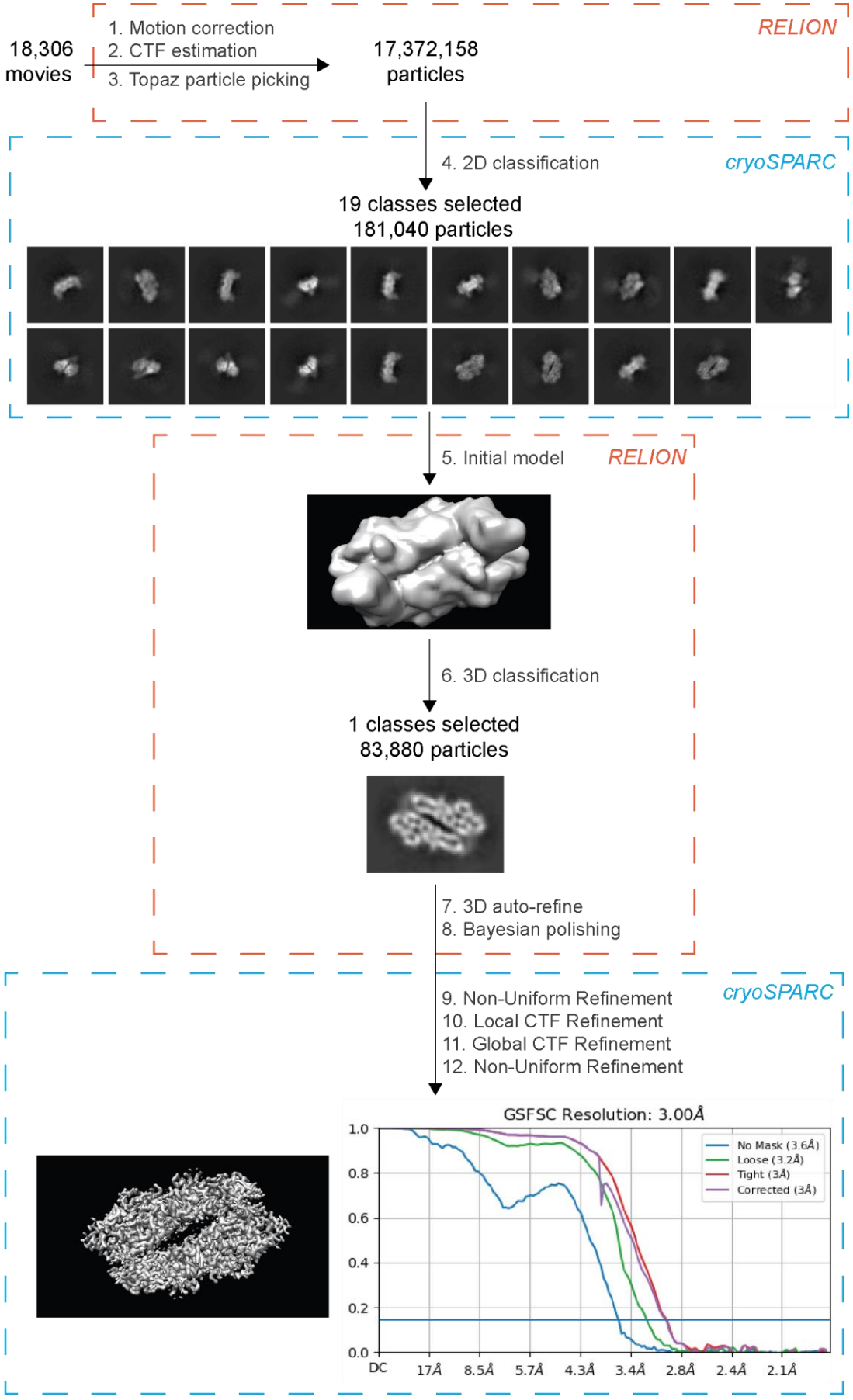

RELION
18,306 movies
1. Motion correction
2. CTF estimation
3. Topaz particle picking
17,372,158 particles
4. 2D classification
cryoSPARC
19 classes selected
181,040 particles
5. Initial model
RELION
6. 3D classification
1 classes selected
83,880 particles
7. 3D auto-refine
8. Bayesian polishing
cryoSPARC
9. Non-Uniform Refinement
10. Local CTF Refinement
11. Global CTF Refinement
12. Non-Uniform Refinement
GSFSC Resolution: 3.00Å
No Mask (3.6Å)
Loose (3.2Å)
Tight (3Å)
Corrected (3Å)
1.0
0.8
0.6
0.4
0.2
0.0
DC
17Å
8.5Å
5.7Å
4.3Å
3.4Å
2.8Å
2.4Å
2.1Å

Figure 3.23 mPADI6 Cryo-EM electron density map calculation workflow
Cryo-EM data processing pipeline for the structure of mPADI6. Final resolution 3.0 Å. Processing performed using a combination of RELION (237) and cryoSPARC (238).

3.4 Summary

Towards structurally characterising PADI6, I optimised workflows for both hPADI6 and mPADI6 production, and was able to express and purify both using mammalian cell culture to a high purity. Using biophysical techniques, I showed both human and mouse PADI6 form dimers in solution. I confirmed that hPADI6 is not stabilised by Ca^{2+} like hPADI4, suggesting that as predicted, hPADI6 does not bind calcium. A similar assay needs to be conducted with mPADI6. Next, I confirmed that both human and mouse PADI6 have no *in vitro* activity against standard PADI substrate BAEE in the COLDER assay. Finally, with the aid of the Crick Structural Biology STP, we generated a 2.44 Å X-Ray crystal structure of my recombinantly purified hPADI6, and I have generated a 3 Å Cryo-EM electron density map of mPADI6, into which the mPADI6 sequence can be modelled. To my knowledge, these are the first high resolution structures of both human and mouse PADI6. Unfortunately, due to time constraints, the mPADI6 sequence couldn't be modelled into this electron density map, and this remains a top priority moving forward. With high resolution structures of both human and mouse PADI6 in hand, it will be of interest to determine whether the structural features discussed in hPADI6 caused by lack of conservation, are also present the mouse protein.

Using our hPADI6 X-ray crystal structure, I investigated key questions we had regarding the active site structure, sequence conservation, and embryonically significant variants. The active site of hPADI6 appeared to be closed when aligned with the calcium and substrate bound human PADI4 structure, with a loop sitting over the opening of the substrate binding pocket. As such, it appears that in this form, it is highly unlikely hPADI6 is catalytically active, and it would require significant structural rearrangements to open the active site pocket for substrate binding. Given the structural rearrangements induced by calcium binding in the other PADIs, this remains a possibility. The two aspartic acids, and the histidine of the hPADI catalytic tetrad are conserved in hPADI6, and their structural positioning aligns with that of hPADI4. The predicted key catalytic cysteine residue of hPADI6 however appears displaced away from the substrate BAEE when aligned with the Ca-hPADI4^{C645A}-BAEE structure.

Interestingly, this predicted key catalytic cysteine residue is in close proximity to another cysteine that is present in hPADI6 but not conserved in the other hPADIs. The proximity of the two cysteines raises the question of whether they form a disulphide bond, and furthermore, if hPADI6 is catalytically active, whether it's regulated by redox chemistry.

Interrogation of the hPADI6 structure showed some other key differences. Three loop insertions in the hPADI6 sequence result in novel loops protruding from the surface of the protein, creating interaction faces not present in hPADI4 and likely the other hPADIs. If PADI6 indeed functions through interactions with other proteins, then it is likely these loops are of evolutionary importance. Identifying PADI6 interacting proteins, and the sites of interaction could provide us with further information on why PADI6 is less conserved than the other PADIs.

Finally, I used our structure to investigate how the different reported embryonically significant hPADI6 variants could be affecting the protein structure. I identified three variants located on the surface of hPADI6 that I predict would result in no disruption to protein folding and stability, and as such likely result in the disruption of key PPIs. I characterised the impact of these variants on protein folding by transient transfection and probing the expression level compared to WT, confirming that these three interaction variants do not result in a reduction in protein expression. Now, it is of interest to further characterise the interactome of PADI6, and for every novel interacting protein, confirm whether these amino acid variants disrupt the interaction. This will not only provide us with information on the location of the interaction but provide answers to why those variants result in such serious developmental phenotypes.

As well as the three variants predicted to disrupt PPIs, I identified three variants that I predicted would disrupt the dimerisation of hPADI6. Upon transient transfection into HEK-293T cells however, two of these variants resulted in reduced expression of PADI6, and the third did not express at all. This result is interesting as based on the structural analysis, none of these variants appeared to be important in protein folding. One theory is that dimerisation is crucial for the stability of PADI6. To test this, it would be interesting to recombinantly express and purify not only these three PADI6 variants, but also other variants that could disrupt dimerisation without impacting folding.

Biophysical experiments to characterise the stability and dimerisation capacity of such variants would then provide further information on whether dimerisation is important for the stability of hPADI6.

In this work, both human and mouse PADI6 were successfully recombinantly expressed and purified. Both proteins were biophysically characterised and the first high resolution structure of hPADI6 determined, with a high-resolution electron density map of mPADI6 ready to be modelled. The structural insights gained from this work suggest that in the state that it was purified, hPADI6 would be incapable of catalytic activity. Given the structural rearrangements induced by calcium binding in PADIs 1 to 4 however, the possibility of similar structural rearrangements in PADI6 being induced by a PPI cannot be ruled out. The possibility of PADI6 catalytic activity is discussed further and speculated upon in **Chapter 7.2**.

Chapter 4. hPADI6 and the subcortical maternal complex

4.1 Introduction

As discussed in the **Chapter 1.2.1**, the SCMC is a complex of 3 core proteins (OOEP, TLE6, NALP5) with a possible 3 further non-core proteins (KHDC3L, ZBED3, NALP2, NALP7) (79,80,82,83). Like PADI6, the SCMC is vital for early embryonic development, with the majority of its constituent proteins also being maternal-effect genes, with expression limited to the oocyte and early embryo. The function of the SCMC appears broad and is still not fully understood, but it is believed to be involved in multiple cellular processes in the early embryo, including symmetric division, organelle localisation, and mRNA storage and translation (80,87). Variants of *PADI6* or SCMC genes in women results in very similar phenotypes, namely early embryonic developmental arrest, or imprinting disorders in offspring (33,84,89,159,161,230). As such, it has been suggested that PADI6 is part of the SCMC, although there is no reported direct experimental evidence confirming an interaction or functional relationship. Characterising the functional relationship between PADI6 and the SCMC could be valuable towards understanding the function of PADI6 in the early embryo.

Towards this, first an interaction between PADI6 and the SCMC must be experimentally confirmed. If PADI6 does indeed interact with the SCMC, determining which member of the complex PADI6 interacts with, as well as the stoichiometry and strength of any interactions will be of interest. Regarding the purpose of the interaction, it is conceivable that SCMC proteins are substrates of PADI6, or potentially act as co-activators to enable citrullination activity. On the contrary, the interaction could have a purely structural function. Investigating these questions will provide insight into the overall function of PADI6 in early embryo development. Confirming an interaction with the SCMC would also be useful for further investigating the effects of *PADI6* gene variants and their impact on protein structure and function.

This chapter describes work done to characterise the functional relationship between hPADI6 and the SCMC. First, using co-immunoprecipitation experiments, an interaction between hPADI6 and the SCMC was confirmed. Individual interactions between hPADI6 and different members of the SCMC were characterised, with hPADI6

forming strong interactions with core members NALP5, TLE6 and OOEP, and weaker interactions with ZBED3 and KHDC3L. The effect of selected embryonically significant hPADI6 variants on the SCMC interactions was then investigated. Expression of the SCMC in *E. coli* and mammalian Expi293 cells was attempted to structurally characterise the complex and its interaction with hPADI6. Finally, towards a possible citrullination function, a chemical probe was employed to determine if the members of the SCMC were able to act as activators or substrates of hPADI6 in a cellular context.

4.2 Investigation into the interaction between the SCMC and PADI6

4.2.1 hPADI6 interaction with the SCMC

To investigate the possible interaction between PADI6 and the SCMC, co-immunoprecipitations with the different members of the complex were carried out. Protein co-immunoprecipitation (co-IP) is a common method for investigating an interaction between two or more proteins. One limitation of co-IP experiments, however, is they require a large amount of input material, as such, performing a co-IP from embryos would require such a large number of embryos it is experimentally unfeasible. Due to restriction of the expression profiles of PADI6 and the SCMC to the oocyte and early embryo, they are not present in commonly used immortalised cell lines. Transient transfection of proteins into a cell line of choice, such as HEK-293T, therefore poses an alternative method to interrogate interactions in a cellular context. As neither hPADI6 or any of the SCMC proteins are present endogenously in the cell line, interactions between hPADI6 and different members of the complex can be probed in isolation from other SCMC proteins. To avoid the need for antibodies specific to each protein, epitope tags, such as FLAG-, HA- and Myc- can be incorporated into the protein expression constructs for transient transfection.

Initially, the coding sequences of *PADI6* and human core SCMC proteins NALP5, TLE6 and OOEP were sub-cloned into mammalian expression vectors using different expression tags. hPADI6 will serve as the bait protein with an N-terminal FLAG tag (FLAG-hPADI6), whilst NALP5 with a Myc tag (Myc-NALP5), and both TLE6 and OOEP with HA tags (HA-TLE6, HA-OOEP) act as prey. Myc-NALP5, HA-TLE6 and HA-OOEP were co-transfected with either FLAG-hPADI6 or FLAG-GFP, with GFP

serving as a negative control bait protein as it would not be expected to interact with the SCMC. Transfected cell lysates were immunoprecipitated with anti-FLAG DynaBeads and the immunoprecipitated fraction probed by western blot (**Figure 4.1**). Pleasingly, FLAG-hPADI6 was able to immunoprecipitate the complex shown by the presence of bands in the FLAG-hPADI6 IP fraction, whereas FLAG-GFP was not. This is the first direct experimental evidence to confirm that hPADI6 does indeed interact with the SCMC.

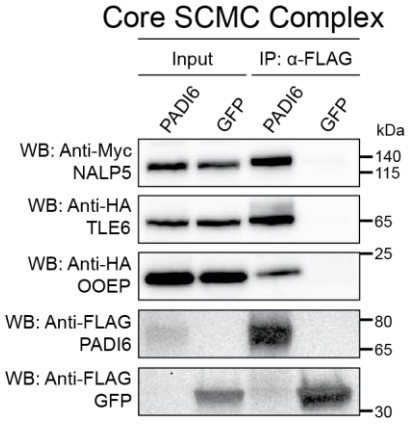


Figure 4.1 Co-IP of full SCMC complex with hPADI6
Western blot analysis of co-immunoprecipitation experiments of HEK-293T lysates (Input) using FLAG(M2)-beads. HEK-293T cells were transiently co-transfected with Myc-NALP5, HA-TLE6, HA-OOEP and either FLAG-hPADI6 or FLAG-GFP. Western blots were probed with antibodies indicated in figure. A representative blot is shown. Experiment performed in duplicate.

4.2.2 Individual SCMC protein interactions

Co-IP experiments with SCMC protein pairs have previously been used to identify specific interactions between member proteins (83). To identify which members of the complex are directly interacting with hPADI6, an equivalent co-IP approach was employed. FLAG-hPADI6 or FLAG-GFP were co-transfected with either Myc-NALP5, HA-TLE6 or HA-OOEP and interactions probed by co-IP (**Figure 4.2a-c**). Interestingly, hPADI6 was able to immunoprecipitate each core member, suggesting it is forming direct interactions with all three of them. Interactions with other reported SCMC proteins KHDC3L (V5 tagged) and ZBED3 (FLAG tagged) were also investigated. In the case of the latter, ZBED3 acted as the bait protein for the co-IP, and an untagged PADI6 or Myc tagged GFP was used as the prey. In both cases, the bait protein was

able to immunoprecipitate the target protein, however, the strength of the interaction appears weaker than the core members (**Figure 4.2d-e**).

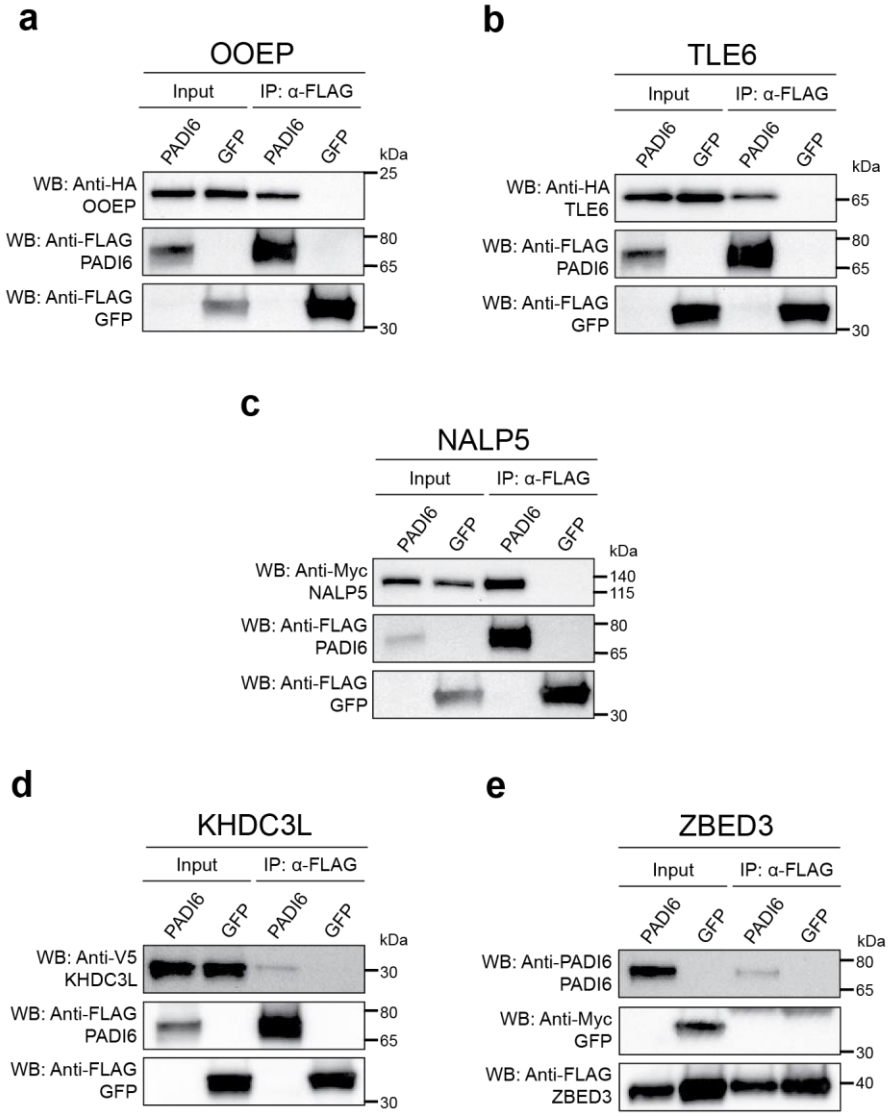


Figure 4.2 Co-IP of individual SCMC proteins with hPADI6
a-d) Western blot analysis of co-immunoprecipitation experiments of HEK-293T lysates (Input) using FLAG(M2)-beads. Cells were transiently co-transfected with either FLAG-hPADI6 or FLAG-GFP and a different HA-tagged SCMC protein. Western blots probed with antibodies indicated in figure. Experiment performed in biological duplicate. **e)** Western blot analysis of co-immunoprecipitation experiments of HEK-293T lysates (Input) using FLAG(M2)-beads of cells transiently co-transfected with FLAG-ZBED3, and either un-tagged hPADI6 or Myc-GFP. Western blots probed with antibodies incubated in figure. Representative blots are shown. Experiment performed in biological duplicate.

4.2.3 Core SCMC protein interactions with hPADI6 variants

As discussed in **Chapter 3.2.12.4**, three of the reported *PADI6* variants that produce embryonic phenotypes result in an amino acid substitution on the surface of the protein (G540R, D547N and N598S). It was further confirmed via western blot that all three variants can be expressed at equivalent levels to the wild-type protein (**Figure 3.19**). Given the serious phenotypes associated with these variants, and their location on the surface of the protein, it is conceivable that they are disrupting key PPIs of hPADI6. To test this hypothesis, it was of interest to determine whether they have any disruptive effect on the interaction between hPADI6 and the SCMC. This will not only shed light on the functional significance of those variants but also allow identification of the surface through which the interactions occur. Co-IP experiments equivalent to those conducted in **Chapter 4.2.2** were employed initially to test for changes in the ability of the three hPADI6 variants (mammalian hPADI6^{variant} expression constructs produced in **Chapter 3.2.12.4**) to immunoprecipitate NALP5, TLE6 and OOEP (**Figure 4.3a-c**). Disappointingly, none of the variants decreased the capacity of hPADI6 to pull-down any of the core SCMC proteins.

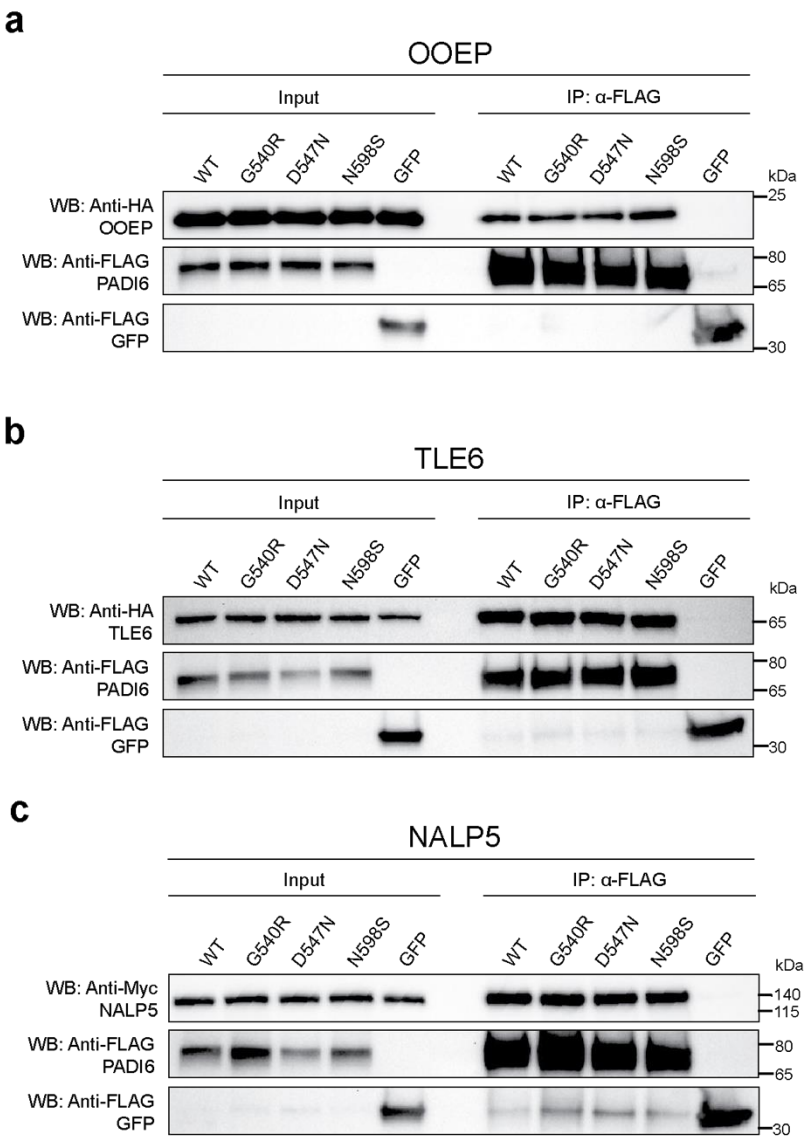


Figure 4.3 Co-IP of individual core SCMC proteins with hPADI6 variants
a)-c) Western blot analysis of co-immunoprecipitation experiments of HEK-293T lysates (Input) using FLAG(M2)-beads. HEK-293T cells were transiently co-transfected with HA-OOEP, HA-TLE6, or Myc-NALP5, and either FLAG-hPADI6, FLAG-hPADI6^{G540R}, FLAG-hPADI6^{D547N}, FLAG-hPADI6^{N598S}, or FLAG-GFP. Western blots probed with antibodies indicated in figure. Representative blots are shown. Experiment performed in biological duplicate.

4.3 SCMC recombinant expression

Whilst co-IP experiments can give a qualitative assessment of a PPI, it gives no information on the affinity, stoichiometry or interaction surface. Recent unpublished work in the Walport Lab has identified a cyclic peptide binder of hPADI4 that binds in the calcium switch region. Peptide binding reduces the calcium concentration requirement for hPADI4 activity and has an activating effect both *in vitro* and in cellular

assays. As endogenous calcium concentrations at their highest (10 μM) rarely if ever reach the concentrations required for activity *in vitro* (10 mM), it is highly likely that PADIs are regulated and activated *in vivo* through unknown PPIs, likely at this calcium switch region (16). It is conceivable that if hPADI6 were to have a catalytic function then it could be activated through a PPI at this location. Deducing the interaction surfaces between hPADI6 and the SCMC, and any structural changes these interactions induce in hPADI6 could give an indication of the function of the interaction and shed light on the catalytic capabilities of hPADI6. Acquisition of structural information about the SCMC would also provide valuable insight into its function, and its relationship with the CPLs.

Towards answering the above questions, a high-resolution structure of the complex, produced through Cryo-EM for example, would be the invaluable. Biophysical techniques such as mass photometry could then provide further information on the size and stoichiometry of the complex in solution, and surface plasmon resonance (SPR) to determine interaction affinity. All of these techniques require large quantities of stable pure protein however and so first, the SCMC proteins had to be recombinantly expressed and purified.

4.3.1 Expression and purification of SCMC proteins from *E. coli*

None of the SCMC proteins have previously been expressed. To start out with, coding sequences of core SCMC proteins NALP5, TLE6 and OOEP were sub-cloned into the pET28a_His bacterial expression vector used for hPADI6 in **Chapter 3.2.6**. Trial expressions equivalent to those in **Chapter 3.2.7** were conducted (**Figure 4.4a-c**).

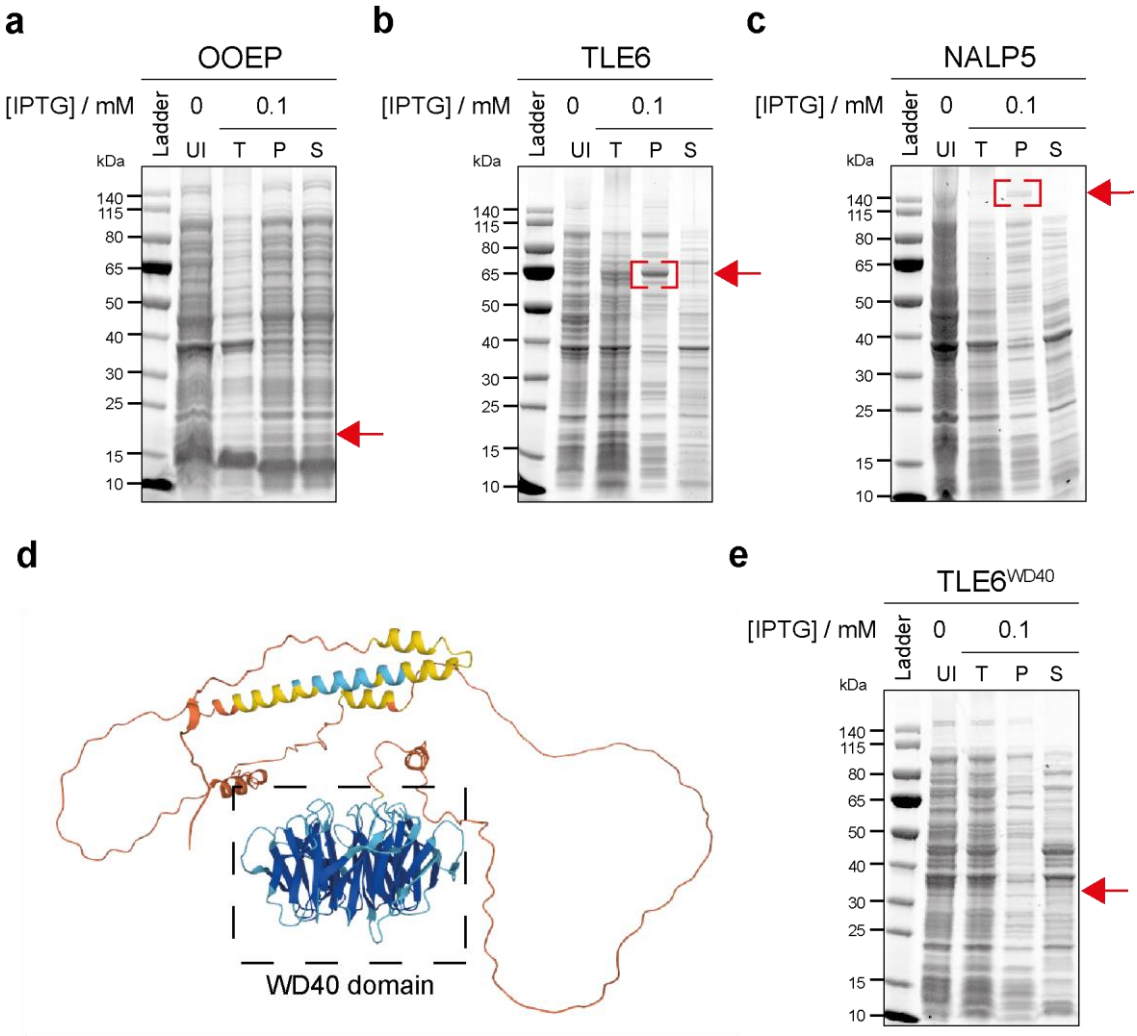


Figure 4.4 Trial expressions of SCMC proteins in *E. coli*
a) Coomassie stained SDS-PAGE gel of pET28a_His-OOEP test expression. A transformed *E. coli* colony was grown in TB to OD600 ~ 1.5 and either not induced or induced with 0.1 mM IPTG overnight at 20 °C. A total sample was taken of both the uninduced and induced cultures (UI, T respectively) before lysis and centrifugation to yield pellet (P) and soluble supernatant (S) fractions. Red arrow indicates expected M_W of expressed protein. **b)** Same as **a)** except pET28a_His-TLE6 instead of pET28a_His-OOEP. Predicted insoluble TLE6 band indicated by red brackets. **c)** Same as **a)** except pET28a_His-NALP5 instead of pET28a_His-OOEP. Predicted insoluble NALP5 band indicated by red brackets. **d)** AlphaFold predicted structure of TLE6, coloured by predicted structure confidence from high (blue) to low (red). WD40 domain indicated by black box. **e)** Same as **a)** except pET28a_His-TLE6^{WD40} instead of pET28a_His-OOEP.

No visible overexpression bands in the soluble fraction were seen for any of the three proteins. To complicate matters, in each of the gels, highly expressed bacterial proteins are present around the expected M_w of each SCMC protein. Therefore, only bands unique to one expression were considered as an SCMC protein. For TLE6 however, a band at approximately 65 kDa, matching well the M_w of TLE6 (64.8 kDa)

was present in the pellet fraction. Similarly, for NALP5, a band at approximately 140 kDa was visible in the pellet fraction, again matching well the Mw of NALP5 (134.3 kDa). This suggests that while both TLE6 and NALP5 had been expressed, both proteins were insoluble when purified from *E. coli*.

Looking at the AlphaFold 2.0 (268) predicted structure of TLE6, there is a conserved WD40 domain predicted with very high confidence surrounded by low confidence unstructured chains with a few low confidence alpha helices (**Figure 4.4d**). As the globular WD40 domain should be more soluble, a truncated construct possessing only the WD40 domain was sub-cloned into the pET28a expression vector (pET28a_His-TLE6^{WD40}). A trial expression of this construct however resulted in no new bands when analysed by SDS-PAGE (**Figure 4.4e**). As no clear over expression was observed for OOEP, and TLE6 and NALP5 were expressed but insoluble, it was decided that expression in Expi293 cells, which had been successful for hPADI6, would be attempted.

4.3.2 Expression and purification of SCMC proteins from Expi293 cells

As before, the coding sequence of each of the three core SCMC proteins were sub-cloned into a pcDNA3.1_Strep-Strep-TEV expression vector (SS-TEV-OOEP, SS-TEV-TLE6, SS-TEV-NALP5). Small scale test expressions of 50 mL cell culture were conducted where lysate was incubated with StrepTactin resin, and bound protein was cleaved off before analysis by SDS-PAGE to evaluate expression (**Figure 4.5a-c**).

Two unique bands were visible in the TLE6 cleavage fraction, both around the expected M_W. The specific band that corresponds to TLE6 could not be identified. A unique but weak band was present in the cleavage fractions at ~140 kDa for NALP5, matching well the expected M_w. No unique bands were present in the OOEP test expression, however. Unfortunately, subsequent 400 mL expressions of TLE6 and NALP5 were unable to replicate the observed bands (**Figure 4.6a-b**). A band was present in the large scale TLE6 expression at ~70 kDa, however this band was present in all expression gels, and is too large to be TLE6.

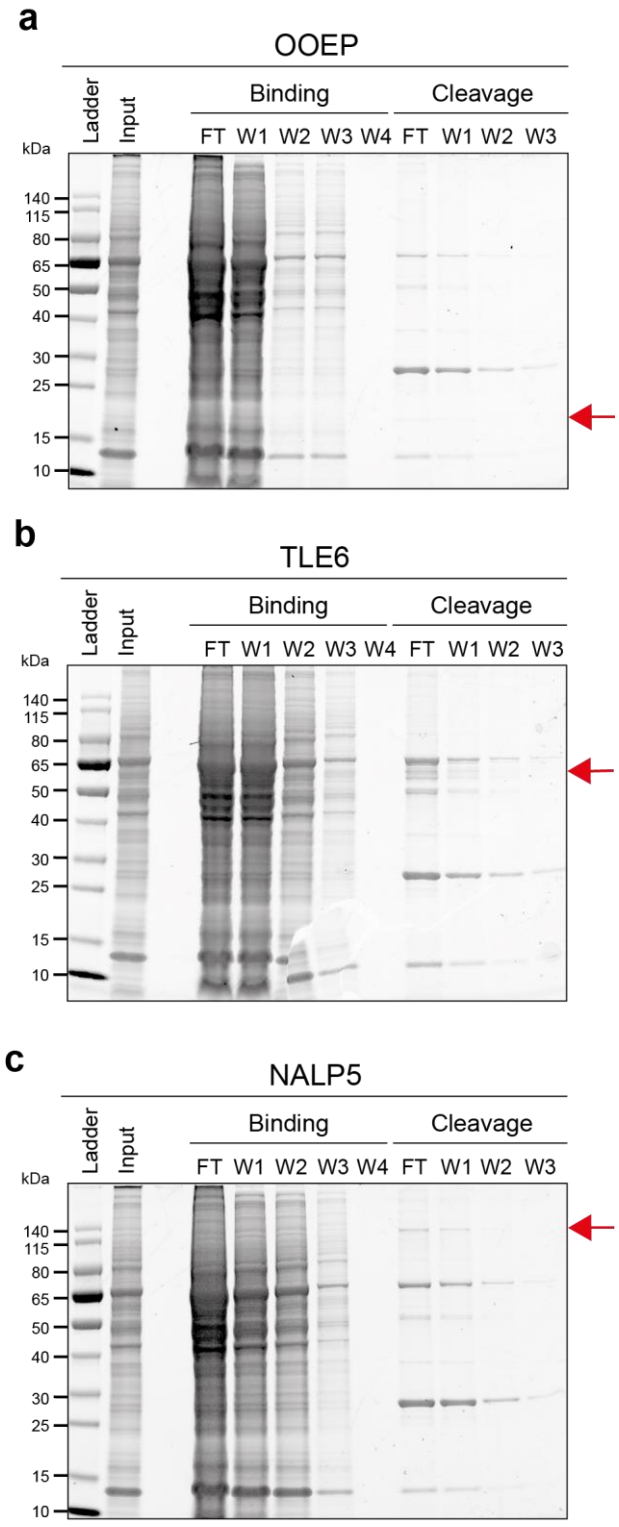


Figure 4.5 Trial expressions of SCMC proteins in Expi293 cells
a) Coomassie stained SDS-PAGE gel of lysates from 50 mL cultures of Expi293 cells transfected with SS-TEV-OOEP (Input) pulled down with StrepTactin resin, followed by cleavage of protein off the resin with TEV protease. FT = Flow through, W1-W3/4 = washes. Red arrow indicates expected M_W of expressed protein. **b)** Same as **a)** except SS-TEV-TLE6 instead of SS-TEV-OOEP. Possible TLE6 unique bands indicated by red arrow. **c)** Same as **a)** except SS-TEV-NALP5 instead of SS-TEV-OOEP.

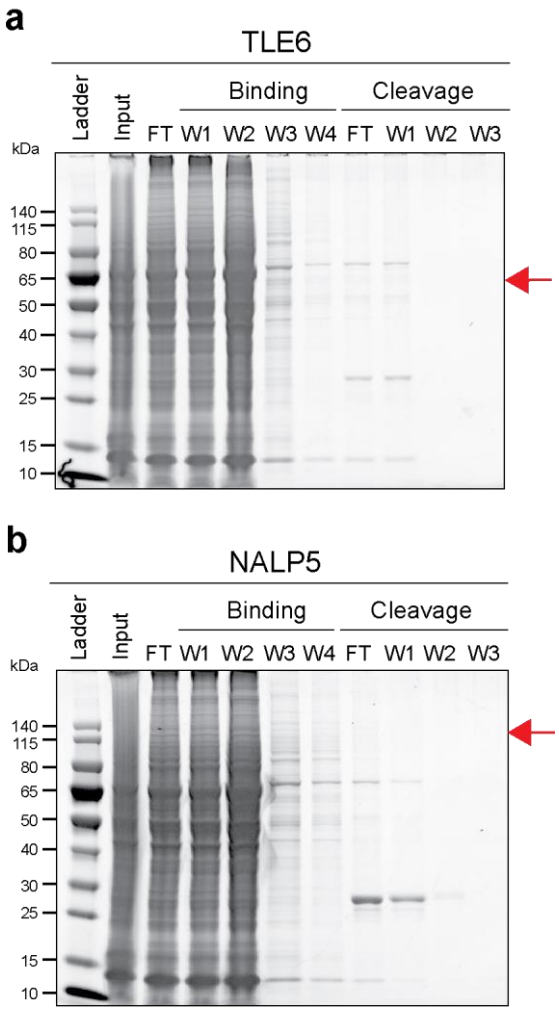


Figure 4.6 Large scale expressions of TLE6 and NALP5 in Expi293 cells
a) Coomassie stained SDS-PAGE gel of lysate from 400 mL Expi293 cells transfected with SS-TEV-TLE6 (Input) pulled down with StrepTactin resin, followed by Cleavage off resin with TEV protease. FT = Flow through, W1-W3/4 = washes. Red arrow indicates expected M_W of expressed protein. **b)** Same as **a)** except SS-TEV-NALP5 instead of SS-TEV-TLE6.

As expression of each member in isolation was proving difficult, it was decided that further attempts would instead focus on co-expressing the entire core SCMC complex. An initial test expression of the full complex, with and without hPADI6 was not successful in expressing the complex to a level sufficient for further purification (**Figure 4.7a-b**). A new unique band at approximately 70 kDa was present in the SCMC + hPADI6 expression, suggesting hPADI6 was being expressed however, indicating the transfection of the cells had worked. As solubility had previously been an issue, buffers with the addition of 5% Glycerol, 0.1% Triton X-100 or both were trialled in an attempt to improve solubility. Expressions were conducted on small scale and expression of the SCMC proteins in the total un-lysed fraction, and soluble fraction was probed via western blot (**Figure 4.7c**). With no additives, TLE6 was completely absent from the

soluble fraction, consistent with previous observations concerning its low solubility. OOEP and NALP5 were present however their bands were faint. With the addition of glycerol, solubility was slightly improved for each protein, however addition of both glycerol and Triton X-100 resulted in the largest improvement in solubility. Following this, a large-scale expression was conducted in a new buffer containing both 5% Glycerol and 0.1% Triton X-100 (**Figure 4.7d**). Unfortunately, no clear overexpression bands were visible by Coomassie staining for any member of the complex. As the previous solubility optimisation experiment had confirmed the presence of each protein by western blot, the only explanation is the SCMC proteins are expressed at very low level precluding clear visualisation by Coomassie staining. It is therefore highly unlikely the SCMC proteins could be expressed using this current system and further optimisation routes or alternate expression strategies need to be explored.

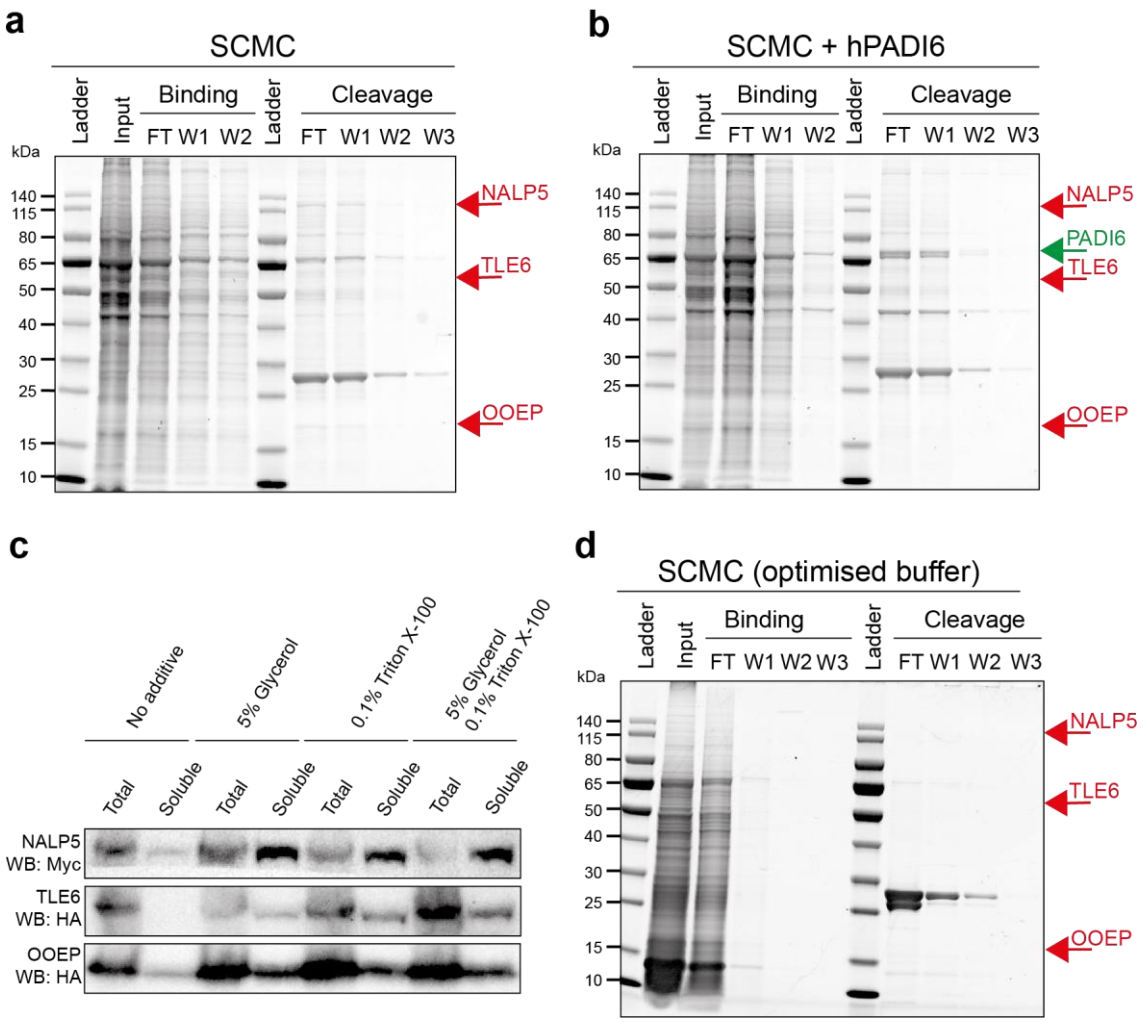


Figure 4.7 Optimisation of SCMC expression in Expi293 cells
a) Coomassie stained SDS-PAGE gel of lysate from 100 mL Expi293 cells transfected with SS-TEV-OOEP, HA-TLE6 and Myc-NALP5 (Input) pulled down with StrepTactin resin, followed by Cleavage off resin with TEV protease. FT = Flow through, W1-W2/3 = washes. Red arrow indicates expected M_W of expressed proteins. **b)** Same as **a)** except SS-TEV-hPADI6, HA-OOEP, HA-TLE6, and Myc-NALP5 instead of SS-TEV-OOEP, HA-TLE6, and Myc-NALP5. Green arrow indicates possible hPADI6 band. **c)** Western blot analysis of cell suspension before lysis (Total) and the soluble fraction after lysis (Soluble) from Expi293 cells transiently co-transfected with HA-OOEP, HA-TLE6, and Myc-NALP5. Lysis buffer contains 50 mM Tris-HCl (pH 8), 250 mM NaCl, 1 mM TCEP and either 5% Glycerol, 0.1% Triton X-100, both, or neither. Western blots probed with antibodies indicated in figure. **d)** Same as **a)** except buffer containing 5% Glycerol and 0.1% Triton X-100 used for lysis.

4.4 Investigation into possible citrullination by, or activation of hPADI6 by the SCMC

So far, an interaction between hPADI6 and the SCMC has been confirmed, but the function of the interaction remains unknown. In **Chapter 3**, hPADI6 was recombinantly expressed and displayed no citrullination activity in an *in vitro* assay with BAEE as a substrate. Whilst the current view is that hPADI6 is not an active citrullinating enzyme, it is possible that the lack of activity seen in previous experiments is due to incorrect substrates, or activating conditions being used. Given calcium binding activates hPADI4 by inducing structural changes in the protein, it is conceivable that similar structural changes could be induced in hPADI6 through a PPI. Performing COLDER assays with SCMC proteins to determine if they can be citrullinated by, or have an activating effect on hPADI6, would require recombinant protein, however. As recombinant protein was proving difficult to produce, it was decided these hypotheses would be investigated in a cellular context using a chemical probe.

Selectively probing citrulline over its precursor arginine is difficult. Towards remedying this, Paul Thompson and colleagues developed a phenyl glyoxal (PG) probe to label citrulline over arginine in milder, lysate compatible conditions (**Figure 4.8**) (240,241). The PG probe relies on the capacity and chemo-selectivity of glyoxal groups to react with citrulline under acidic conditions, and arginine under basic conditions (269). Crucially, the PG group can selectively react with citrulline under milder acidic conditions (5% TCA) and at lower temperatures (37 ^{o}C) than other methods (240,241). Conjugation of the reactive PG group, for example through Click

chemistry to fluorescent moieties can allow the visualisation of citrullinated proteins in an SDS-PAGE gel by in-gel fluorescence. Alternatively, biotin can be conjugated to allow the pull-down of citrullinated proteins and their subsequent identification by proteomics (241). It was decided that the PG-TAMRA probe displayed in **Figure 4.8** be synthesised to determine whether co-expressing hPADI6, with and without the SCMC in HEK-293T cells resulted in any changes in citrullination. The probe was synthesised, purified and characterised by another member of the Walport group, Dr Ewen Calder.

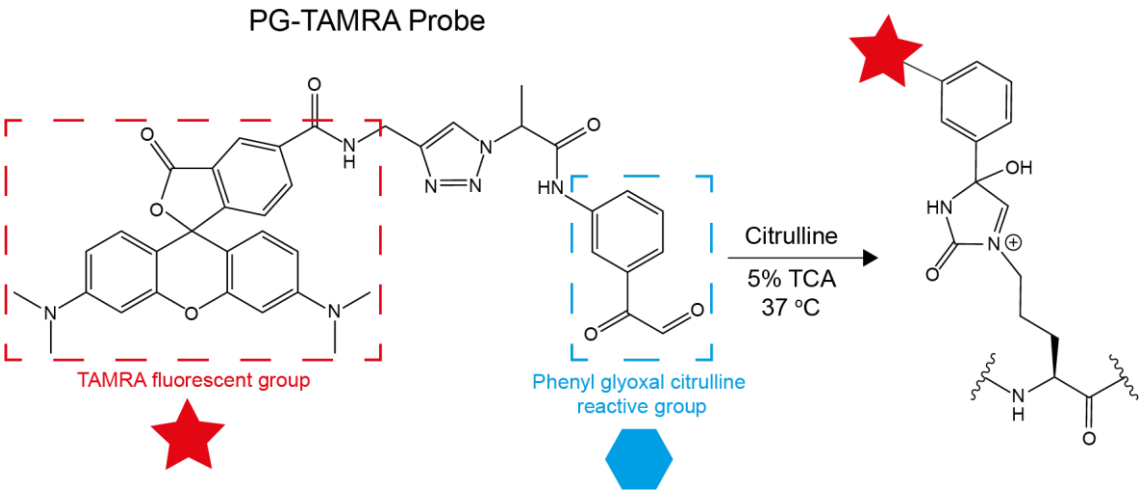


Figure 4.8 PG-TAMRA citrulline probe
PG-TAMRA probe chemical structure and citrulline labelling scheme. TAMRA fluorescent group indicated by red box and star. PG reactive group indicated by blue box and hexagon. Under acidic conditions (5% TCA) the PG probe reacts specifically with citrulline over arginine.

4.4.1 Probe validation

As an initial test, a time-course experiment with hPADI4 and purified histones was conducted to validate the PG-TAMRA functioned as desired. Recombinant hPADI4 was incubated at 10 mM $CaCl_2$ with purified histones for a period of between 0 and 60 mins. Citrullinated proteins were labelled *in vitro* by the probe, and the sample was resolved by SDS-PAGE before in-gel fluorescent imaging at 532 nm (**Figure 4.9a-b**). As shown in **Figure 4.9b** a time-dependent increase in in-gel fluorescence signal was observed at a mass of approximately 15-20 kDa, matching histone M_W, suggesting the probe was successfully labelling citrulline over arginine. A faint band of consistent size was also present around 75 kDa, likely hPADI4 auto-citrullination which has been reported to happen (263).

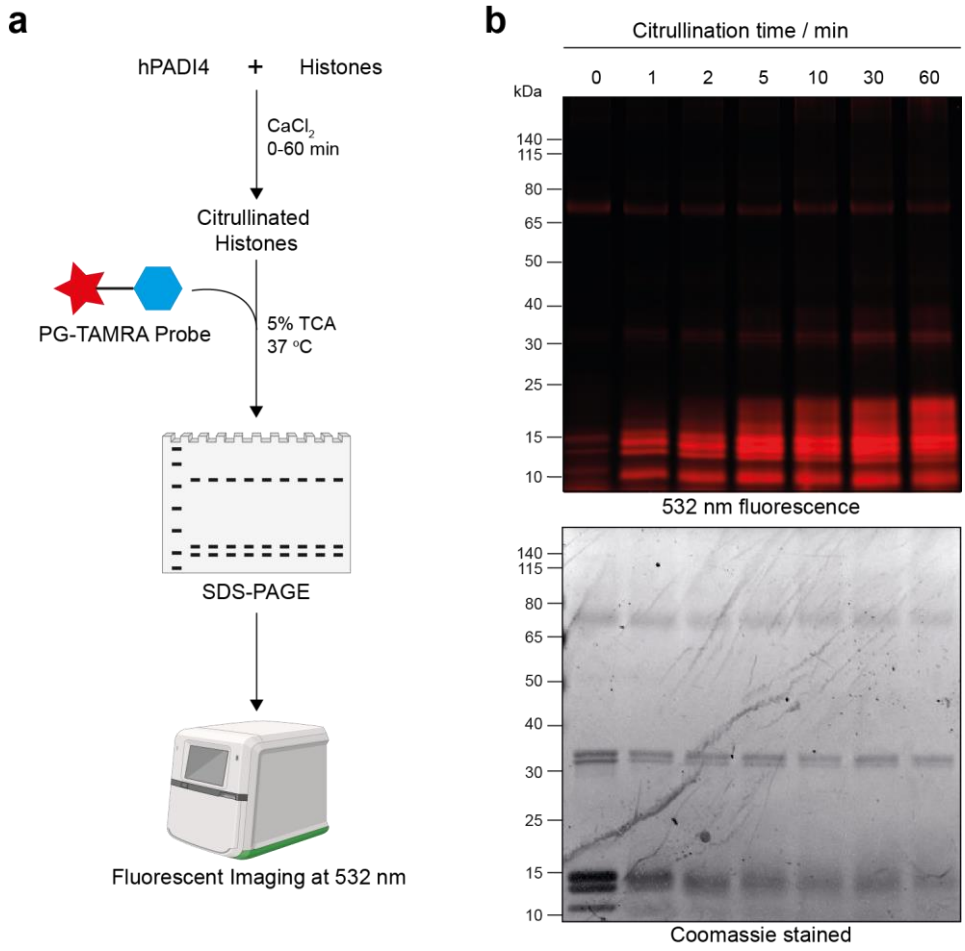


Figure 4.9 PG-TAMRA probe validation with hPADI4 and histone citrullination
a) Schematic of probe validation workflow. 0.5 μM recombinant hPADI4 and 1.5 μg purified histones were incubated in 10 mM $CaCl_2$ for between 0 and 60 mins, quenched with 50 mM EDTA and labelled with 0.1 mM PG-TAMRA in 5% TCA at 37 °C for 30 mins before quenching with 100 mM citrulline. Labelled proteins were precipitated, washed with acetone, dried and resuspended before being run on an SDS-PAGE 4-20% gel. The gel was fluorescence imaged at 532 nm using a Typhoon FLA 9500 imager, before being stained with Coomassie and imaged using a Bio-Rad ChemiDoc system to characterise gel loading. **b)** Top Panel: 532 nm fluorescence gel image of citrullination. Bottom Panel: Coomassie stained gel image of protein loading.

4.4.2 Investigation into possible citrullination of hPADI6 in HEK-293T cells

To probe possible citrullination of or co-activation by the SCMC, hPADI6 alone or with core SCMC proteins NALP5, TLE6 and OOEP were co-transfected into HEK-293T cells. As positive and negative controls, parallel co-transfections were conducted with GFP or hPADI4 replacing hPADI6. Cells were lysed in a high calcium (10 mM) lysis buffer and subjected to labelling by the PG-TAMRA probe as in **Chapter 4.4.1**. Labelled lysates were then resolved by SDS-PAGE and imaged (**Figure 4.10a-b**). No

difference in fluorescence signal was observed between the GFP and hPADI6 transfected samples, with or without the SCMC. A significant increase in fluorescence signal across the gel was however seen with hPADI4 transfected samples, indicating the labelling was successful. Together, this result suggests that no additional citrullination occurred in the cells transfected with hPADI6 compared to GFP. This observation is in line with previous *in vitro* citrullination assays suggesting that hPADI6 is incapable of citrullination, and that the SCMC and HEK-293T proteome are not sufficient to activate hPADI6 under the conditions tested.

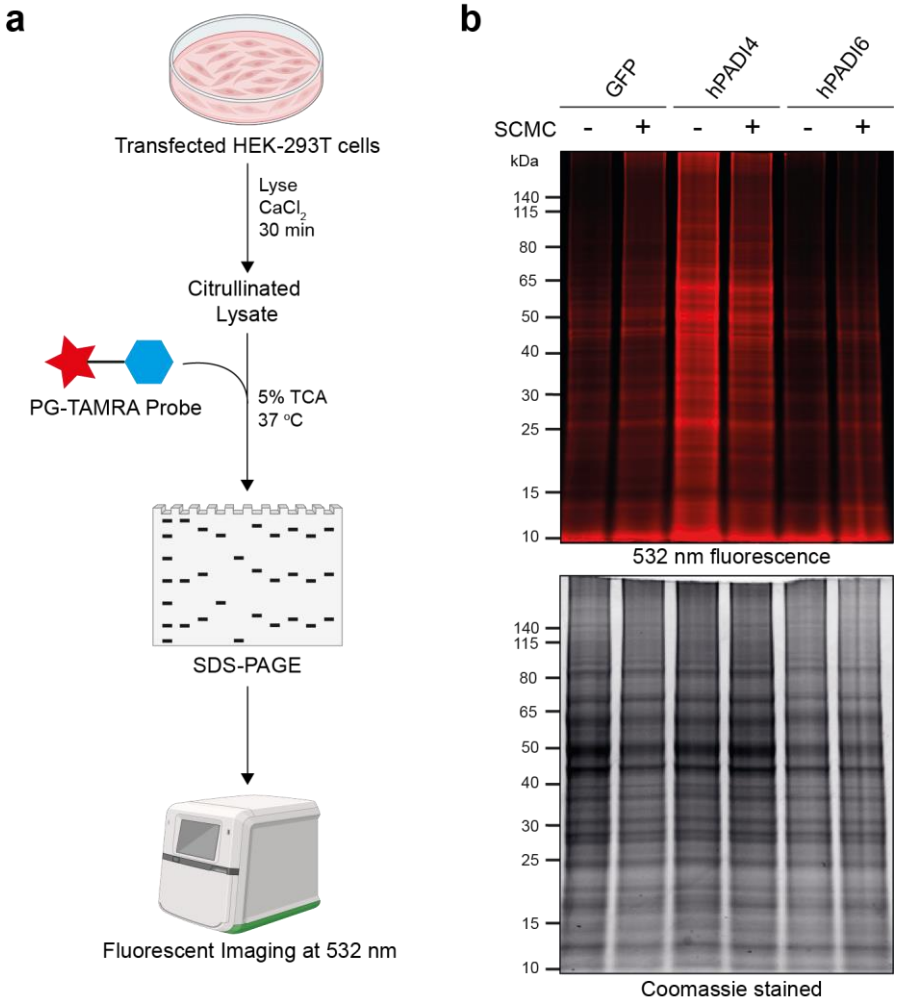


Figure 4.10 PG-TAMRA probe investigation into possible citrullination by hPADI6 in transfected HEK-293T cells
a) Schematic of HEK-293T citrulline characterisation workflow. HEK-293T cells were transfected with either FLAG-GFP, FLAG-hPADI4, or FLAG-hPADI6, with and without Myc-NALP5, HA-TLE6 and HA-OOEP (SCMC). Cells were lysed in a buffer with 10 mM $CaCl_2$ and labelled, processed and imaged as described in **Figure 4.9a**. **b)** Top Panel: 532 nm fluorescence gel image of citrullination. Bottom Panel: Coomassie stained gel image of protein loading.

4.5 Summary

The aim of this work was to characterise the functional relationship between PADI6 and the SCMC. Whilst reported data pointed strongly towards an interaction between PADI6 and the SCMC, no direct experimental evidence for an interaction had been reported. In this work, I confirmed that hPADI6 does indeed interact with the SCMC and is able to co-immunoprecipitate the core complex of OOEP, TLE6 and NALP5. Interestingly, human PADI6 was able to co-immunoprecipitate each core member individually, as well as showing a weaker interaction for non-core members KHDC3L and ZBED3. I then tested if three phenotypically significant PADI6 variants resulted in a loss of interaction with OOEP, TLE6 and NALP5. All were capable of co-immunoprecipitating each core SCMC protein, suggesting the tested residues are not important for these interactions. Moving forward, I need to test whether these variants disrupt interactions with non-core SCMC proteins KHDC3L and ZBED3, as well as any other reported PADI6 interacting protein. Whilst these experiments suggest hPADI6 is indeed interacting with the SCMC, no quantitative information could be obtained on the strength of the interaction, or stoichiometry.

Towards structurally characterising the SCMC, and the interaction with hPADI6, I attempted to recombinantly purify the SCMC proteins from both bacterial and mammalian cell culture. These attempts were unsuccessful, and moving forward, alternate expression systems will be tested. As it appears solubility is an issue, one strategy could be to express globular truncated constructs of the SCMC proteins, based off the folded domains in AlphaFold 2.0 predicted structures (268). In-line with this, purification tags that can aid in solubility such as GST, could be trialled. Alternatively, another option is baculovirus-insect cell expression, which often succeeds in expressing proteins where others fail (270). In the case that the SCMC proteins cannot be expressed in sufficient quantities in any system, mass spectrometry methods could be employed for further characterisation. Various protocols have been reported that can use mass spectrometry to determine protein interaction stoichiometry and quantitative information on binding affinities (271). Alternatively, chemical cross-linking experiments combined with mass spectrometry could provide information on where the interaction surfaces are in the proteins (272). This could allow generation of a low-resolution model of the SCMC.

Given the interaction with the SCMC, I hypothesised that these proteins could either act as co-activators, or substrates of hPADI6. To answer these questions, I employed a chemical probe that specifically reacts with citrulline over arginine under acidic conditions and possesses a TAMRA fluorescent group for characterisation. In-gel fluorescence signal was no different in lysate from HEK-293T cells transfected with hPADI6, with and without the SCMC core complex, compared to from cells transfected with GFP. This result is consistent with the previous findings that hPADI6 is not catalytically active, although the experiment cannot rule out other possibilities. For example, it is possible that the SCMC core proteins are not substrates of hPADI6, they cannot activate hPADI6, or, either is true but no natural substrates or the correct co-activating factors are present in the HEK-293T cell line used. It would be interesting to repeat a similar experiment with non-core SCMC proteins, such as KHDC3L or ZBED3, and moreover with any newly identified PADI6 interacting proteins.

It is intriguing that hPADI6 appears to interact with each core member of the SCMC, as well as non-core members KHDC3L and ZBED3. It is most likely that PADI6 interacts with these proteins as a dimer, however, several questions come to mind when thinking of a model for these interactions. Firstly, is PADI6 forming a specific interaction with each protein? Are these interactions at different faces on hPADI6 (**Figure 4.11a**)? Or does hPADI6 form interactions with each SCMC protein on the same face, potentially at a point of the complex where all three core proteins are in close proximity to one another (**Figure 4.11b**)? As the CPLs appear to be composed of long fibres of proteins, one possibility is that PADI6 interacts with different SCMC proteins at opposite interfaces, potentially forming a link between different SCMC subunits (**Figure 4.11c**). If hPADI6 is forming specific and unique interactions with each SCMC protein, then a more likely model for the hPADI6 and the SCMC forms the CPLs involves larger repeat units with hPADI6 and the SCMC proteins in different orientations (**Figure 4.11d**). Elucidating the interaction faces and stoichiometry of the complex, and hPADI6 will be imperative for answering these questions. Furthermore, answering the above questions will provide invaluable information in the characterisation of the CPLs. As such, the recombinant expression and purification of the core members of the complex, as well as the non-core members is of upmost importance and is a key future objective from this work.

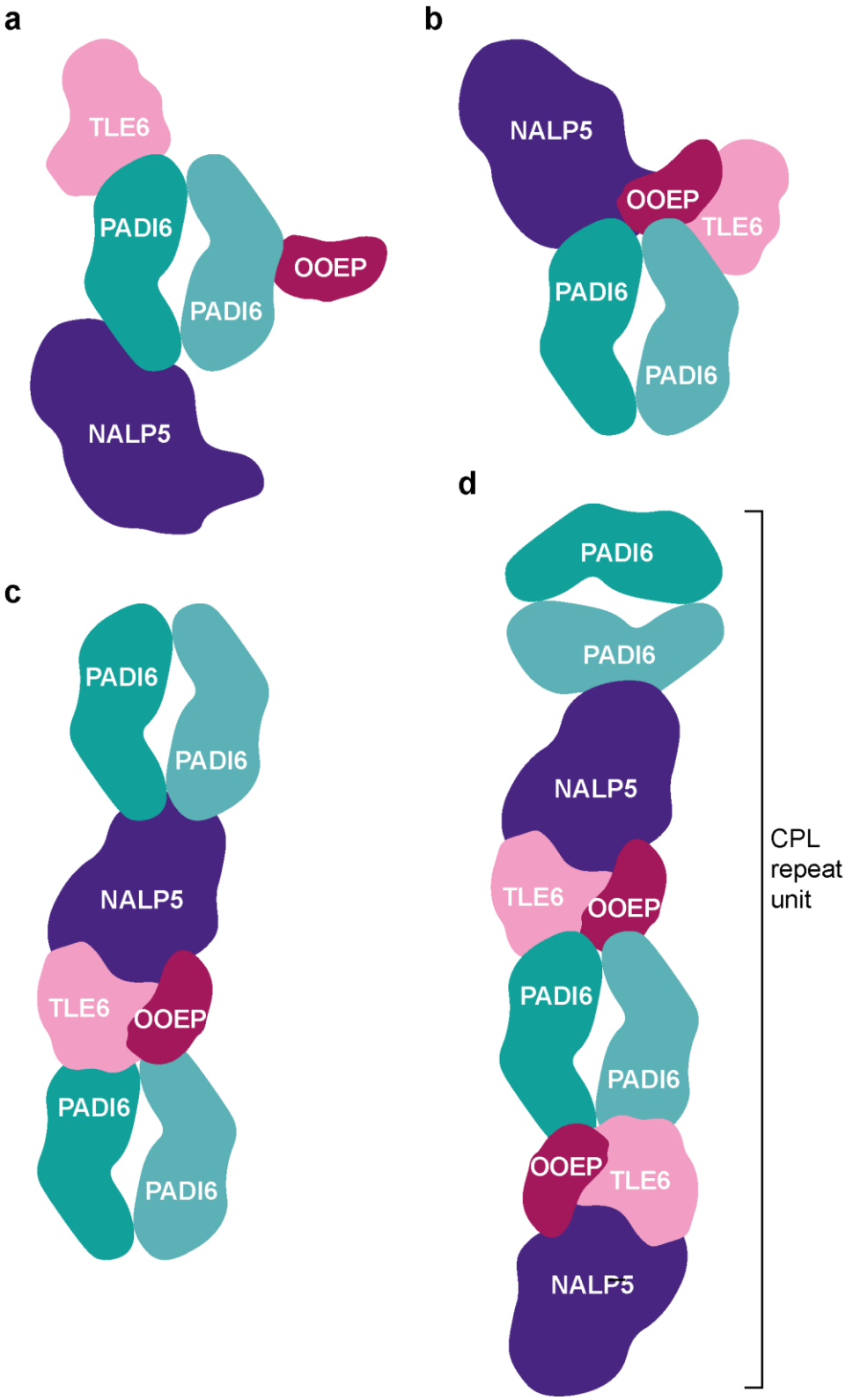


Figure 4.11 PADI6 and SCMC suggested interactions
a) SCMC proteins each interact with a different face of PADI6 dimer. **b)** PADI6 dimer interacts with the SCMC on a face where each core member is in close proximity to one another. **c)** PADI6 dimer interacts with different SCMC proteins on opposite surfaces to form long protein complexes of repeat units. **d)** More realistic view of SCMC and PADI6 interaction structure to form CPLs.

An interaction between PADI6 and the newly identified CPL protein UHRF1 was not investigated in this work and must be characterised. If PADI6 does interact with UHRF1 by co-immunoprecipitation, then UHRF1 will be recombinantly expressed and purified for further structural studies, such as Cryo-EM. In contrast to the SCMC proteins, it has been reported that UHRF1 can be recombinantly expressed and purified from *E. coli* (273). Characterisation of a potential relationship between PADI6 and UHRF1 is a key future aim arising from this work.

Overall, in this work I have confirmed an interaction between hPADI6 and the SCMC. As well as identifying that hPADI6 forms interactions with each core SCMC protein, as well as two non-core proteins in isolation of the rest of the complex. Further work now is required to understand the structural arrangement of these interactions, and the complex, and to determine the function of these interactions.

Chapter 5. PADI6 in the early embryo

5.1 Introduction

As discussed in **Chapter 1.2**, PADI6 is indispensable for early embryo development in humans and mice. *PADI6* variants in women often result in their infertility, with the most common phenotype being early embryonic developmental arrest around the 4- to 8-cell stage. Knocking-out *Padi6* (*Padi6*$^{-/-}$) in mice renders the females infertile with their embryos arresting at the 2-cell stage. The coincidence of embryo arrest stage and timing of genome activation in humans and mice has led to speculation as to whether PADI6 is involved in EGA (68). Other theories include the storage, localisation and regulation of maternal mRNA and ribosomes, organelle localisation, and symmetric division (70). All together however, there is currently no clear understanding of the function of PADI6 in early embryo development.

Oocytes from *Padi6*$^{-/-}$ mice appear to mature and can be fertilised without any consequences (32). The only observed difference between null oocytes and WT oocytes is that they lack CPLs. Like PADI6, the exact nature and function of the CPLs is unknown, however current leading theories suggest that they are made up of PADI6 along with the SCMC and are involved in maternal protein storage in the oocyte (67).

One of the most important questions that must be answered regarding the function of PADI6 is, does PADI6 have a catalytic citrullination function? Currently, there is no reported PADI6 catalytic activity. This does not necessarily mean that there isn't any, just that we may not have found the right activating conditions and/or substrates to demonstrate citrullination activity *in vitro*. As previously discussed, there is conflicting evidence regarding citrullination in the early embryo, as discussed in **Chapter 1.2.5**. On the other hand, given the lack of PADI6 conservation in the catalytic cysteine and calcium binding residues, potentially PADI6 evolved to not have a catalytic function. It is therefore conceivable that PADI6 functions not through citrullination but through PPIs or scaffolding in the oocyte and early embryo. This chapter describes the work aiming to elucidate the function of PADI6 in the early embryo, and crucially whether PADI6 has a catalytic function or not by using two mouse models.

5.2 Catalytic mutant mouse model

5.2.1 Rationale

Answering the question of whether PADI6 does or does not have a catalytic function is of vital importance to understanding its overall function. Knock-out models are commonly used to untangle protein function and can prove valuable for determining cellular mechanisms, for example by identifying which cellular pathways, mRNA, or proteins are perturbed. In the case of PADI6 however, whilst a knock-out model could allow the identification of the cellular perturbations caused by the absence of PADI6, it would not give any clarification to whether these are due to a lack of citrullination or the absence of key interactions in the early embryo. An alternate model, where any potential catalytic activity is extinguished, but PADI6 retains its natural folded structure, however, would be invaluable in the pursuit of these answers.

As discussed in **Chapter 1.1**, PADI catalytic activity is dependent on a key nucleophilic cysteine residue in the active site. This cysteine, along with the other key amino acids in the catalytic pocket are conserved in PADIs 1 to 4 (**Figure 1.4a-c**). In PADI6 however, the key cysteine is replaced by an alanine, flanked by two cysteines in hPADI6, and by a cysteine and a serine in mPADI6. As only the first of these two hPADI6 cysteines conserved in PADI6 across multiple species, it is probable that if PADI6 were to have a catalytic function, it would be through this first cysteine (C675 in human PADI6, C663 in mouse PADI6) (**Figure 1.4d**). Substitution of the key cysteine in hPADI4 to alanine has been shown to eliminate any catalytic activity whilst retaining the structure of the protein (28). We hypothesise that if embryos from female mice possessing an mPADI6 C663 to A substitution (*Padi6*^{C663A}) display any developmental phenotypes, then this would be due to a lack of PADI6 catalytic activity. Such a model would therefore allow the direct interrogation of possible PADI6 catalytic activity in the early embryo, and through comparison with a knock-out model (*Padi6*$^{-/-}$), would be invaluable for understanding the function of PADI6 in embryo development (**Figure 5.1a**). As *Padi6* is a maternal-effect gene, only expressed from the maternal genome, embryos from homozygous mutant females mated with WT males display the mutant phenotype (**Figure 5.1b**). Such embryos are thus termed mat-*Padi6*$^{C663A/C663A}$ or mat-*Padi6*$^{-/-}$, for the catalytic mutant and knock-out lines respectively.

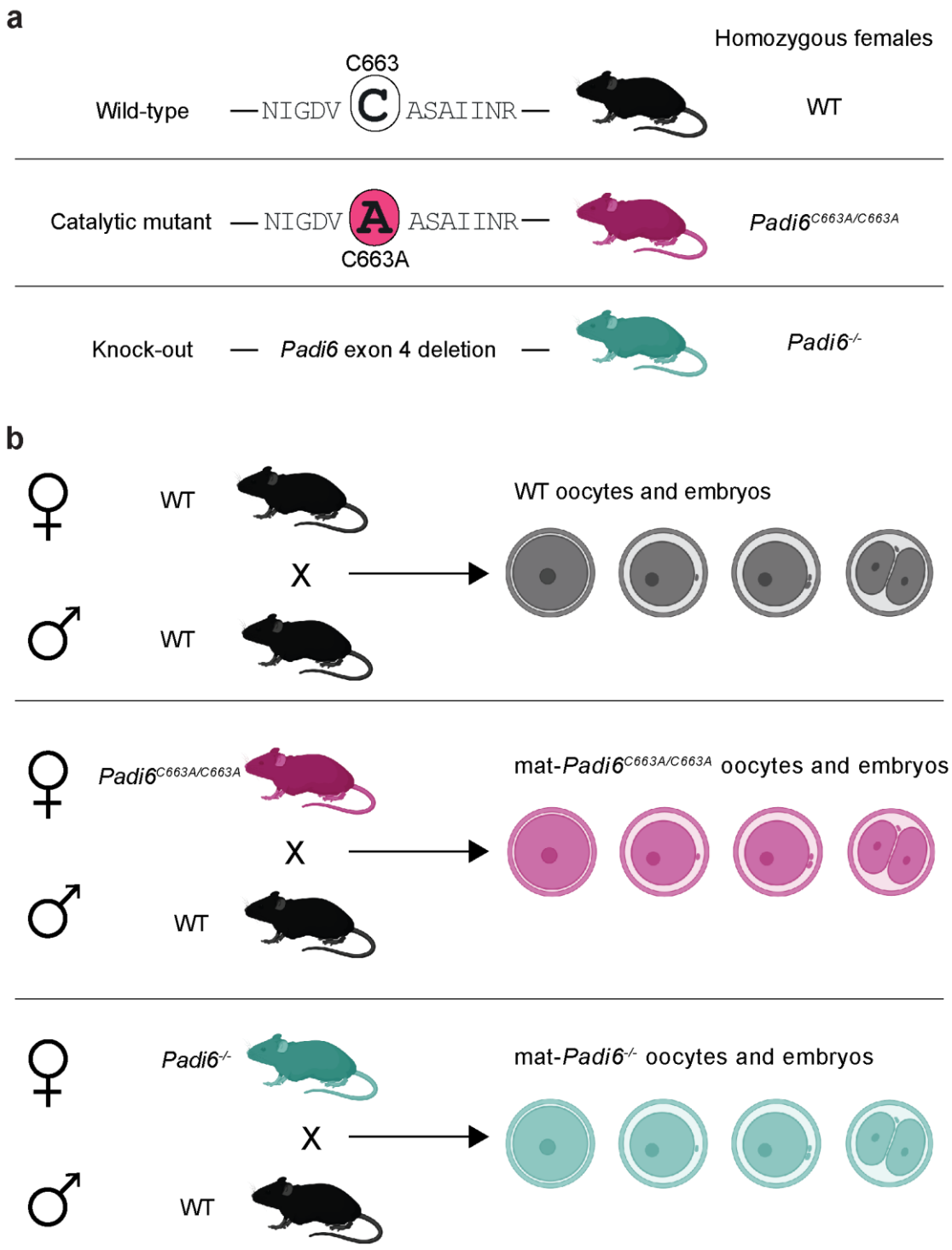


Figure 5.1 *Padi6* transgenic mouse lines
a) Mouse line definitions. Wild-type (WT) C57BL/6N female mice. Catalytic mutant (*Padi6*$^{C663A/C663A}$) homozygous females where cysteine 663 of PADI6 is substituted to alanine. Knock-out (*Padi6*$^{-/-}$) homozygous females where exon 4 of the *Padi6* gene is deleted resulting in a knock-out phenotype. **b)** Natural mating overview of WT and homozygous *Padi6*$^{C663A/C663A}$, and *Padi6*$^{-/-}$ females with WT males to produce WT, mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ oocytes and embryos respectively.

5.2.2 Characterisation of recombinant mPADI6^{C663A}

Before generating a *Padi6*^{C663A} transgenic mouse line however, we needed to ensure that the folded structure and biophysical properties of mPADI6 were unaffected by the substitution. Using the same protein expression workflow as for wild-type mPADI6 in **Chapter 3.3.2**, mPADI6^{C663A} was expressed and purified (**Figure 5.2a-d**). Pleasingly, similar yields of mPADI6^{C663A} to mPADI6 were obtained and Intact-MS confirmed mPADI6^{C663A} had been purified with a mean M_w of 77125 Da, matching the M_w of the mPADI6^{C663A} variant (**Appendix 11**).

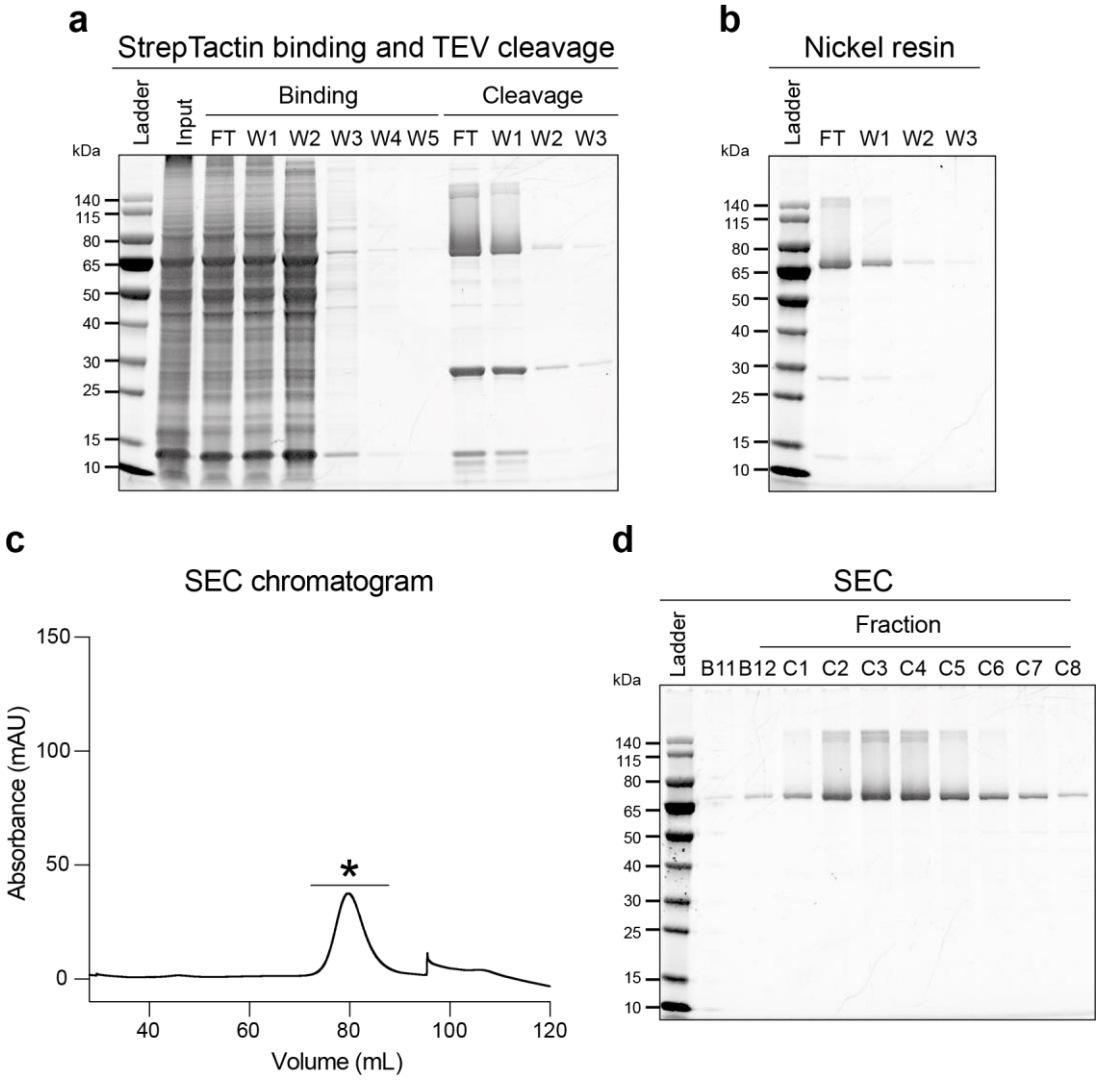


Figure 5.2 Expression and purification of mPADI6^{C663A} from Expi293 cells
a) SDS-PAGE gel after Binding of Expi293 lysate to StrepTactin resin, followed by on-resin Cleavage of mPADI6^{C663A} off resin with TEV protease. FT = Flow through, W1-W3/5 = washes. **b)** SDS-PAGE gel of sample before and after application to Nickel resin to remove TEV protease. FT = Flow through, W1-W3 = washes. **c)** Chromatogram of SEC purification of mPADI6^{C663A}. Fractions taken for SDS-PAGE characterisation are labelled with a black star. **d)** SDS-PAGE gel of SEC purified fractions. Fractions B12 to C8 combined and taken forward.

The purified protein was then subjected to CD, NanoDSF and mass photometry to confirm it behaved the same as WT protein. CD of mPADI6 and mPADI6^{C663A} produced almost identical spectra (**Figure 5.3a**). NanoDSF showed the proteins melted at similar temperatures (**Figure 5.3b**, mPADI6 = 47.8 °C, mPADI6^{C663A} = 47.5 °C). The general profile of the melt was the same for each protein however a lower overall absorbance was observed for mPADI6^{C663A}, likely due to protein concentration differences in the experiment. Finally, mass photometry confirmed that mPADI6^{C663A} can dimerise with a similar ratio of monomer to dimer compared with mPADI6 (**Figure 5.3c**).

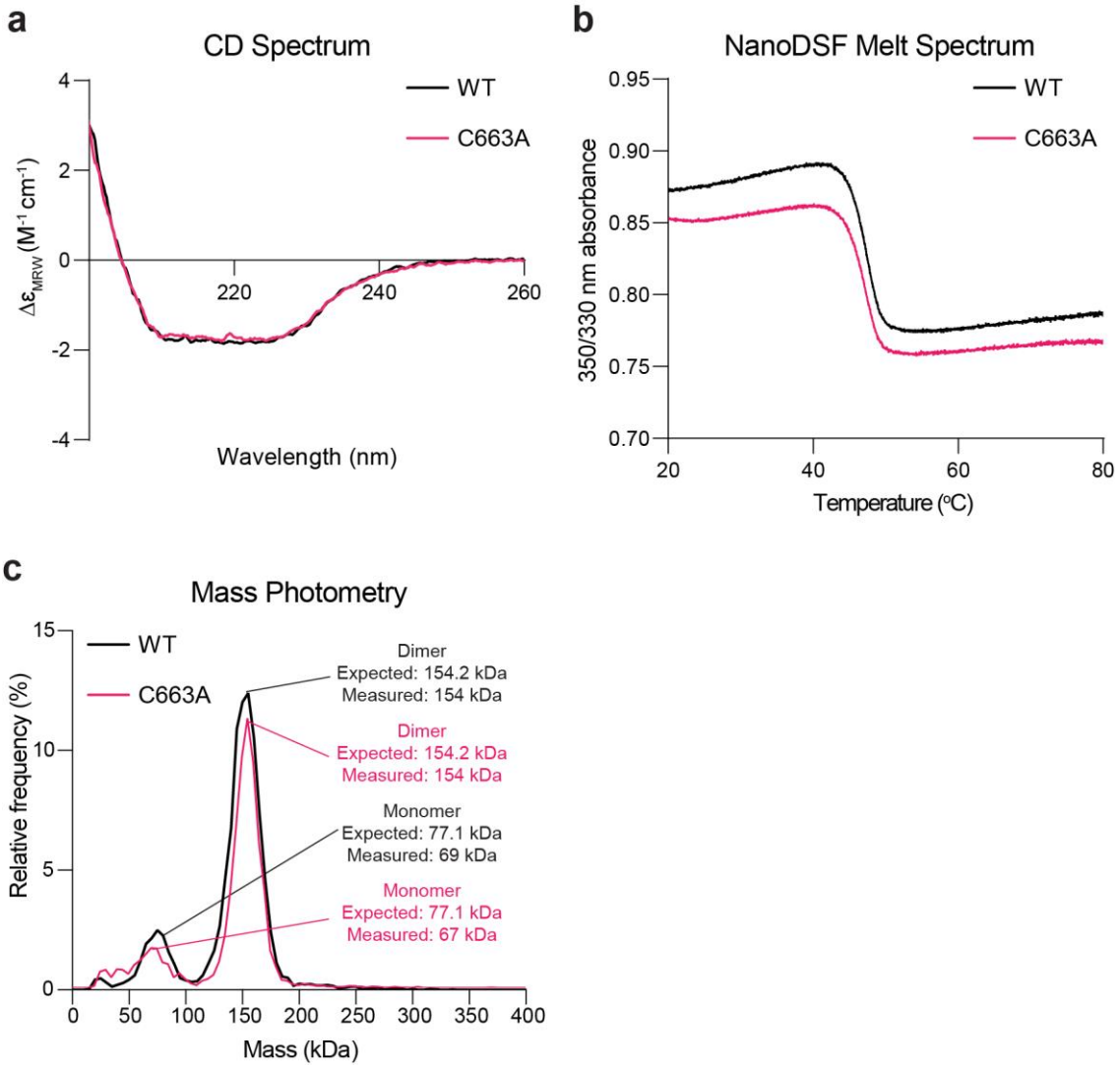


Figure 5.3 Biophysical characterisation of mPADI6^{C663A}
a) CD spectrum of mPADI6 (black) and mPADI6^{C663A} (pink) purified from Expi293 cells. Observed CD Signal in millidegrees was converted to mean residue CD extinction coefficient, as per (232). **b)** NanoDSF melting spectrum of mPADI6 (black) and mPADI6^{C663A} (pink). Absorbance was measured at 350 and 330 nm and the ratio plotted. Experiment was performed in triplicate. **c)** Mass photometry histogram of mPADI6 (black) and mPADI6^{C663A} (pink) expressed from Expi293 cells. Bin size = 5 kDa.

5.2.3 Generation of *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ female mice

CRISPR-Cas9 technology was used to generate the *Padi6*^{C663A} mouse line in collaboration with the Turner lab. This was performed by the Genetic Modification Service (GeMS) at the Crick. In parallel, the C57BL/6N *Padi6*$^{tm1a(KOMP)Wtsl}$/MbpMmucd line (ID:48953) was obtained from Mutant Mouse Resource & Research Centres (MMRC). Full knock-out *Padi6*$^{-/-}$ mice were then generated by crossing homozygous *Padi6*^{tm1a} mice with PGK-Cre mice, resulting in the deletion of a region spanning exon 4 of the *Padi6* gene. Homozygous females from both lines were then continually produced by mating homozygous males (shown to have no phenotype from knocking out *Padi6*) with heterozygous females (32). As a control in this work, wild-type C57BL/6N females (WT) were used. Throughout this work, all mouse animal husbandry was performed by the Crick Biological Research Facility, matings were set up by Andrew Ojarikre of the Turner lab, and mice were culled for harvesting by myself or members of the BRF. With the exception of Embryoscope+ work, all oocytes and embryos were harvested by myself.

To confirm *Padi6* mRNA levels were not affected in the *Padi6*$^{C663A/C663A}$ line relative to WT, quantitative real time-polymerase chain reaction (qRT-PCR) of *Padi6* transcripts was performed on mRNA extracted from the ovaries of 8-week-old females from each transgenic line. Double delta C_t values were generated for *Padi6* transcript levels in *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ ovaries compared to WT ovaries normalised to beta-2-microglobulin (*B2m)* as the housekeeping gene (**Figure 5.4a**). No significant difference was observed between WT and *Padi6*$^{C663A/C663A}$ ovaries, and as expected no *Padi6* mRNA was detected from *Padi6*$^{-/-}$ ovaries. To confirm *Padi6*$^{C663A/C663A}$ females possessed the desired mutation, a fragment of approx. 1000 bp in the *Padi6* mRNA spanning C663 was amplified by PCR and sent for Sanger sequencing (**Appendix 12**). The *Padi6*$^{C663A/C663A}$ fragment possessed the desired TGT (cysteine) to GCC (alanine) codon switch, along with other synonymous mutations included in the guide RNA cassette. Finally, immunofluorescence imaging of PADI6 in zygotes was used to confirm that mat-*Padi6*$^{C663A/C663A}$ embryos possessed comparable signal intensity and distribution to WT (**Figure 5.4b**).

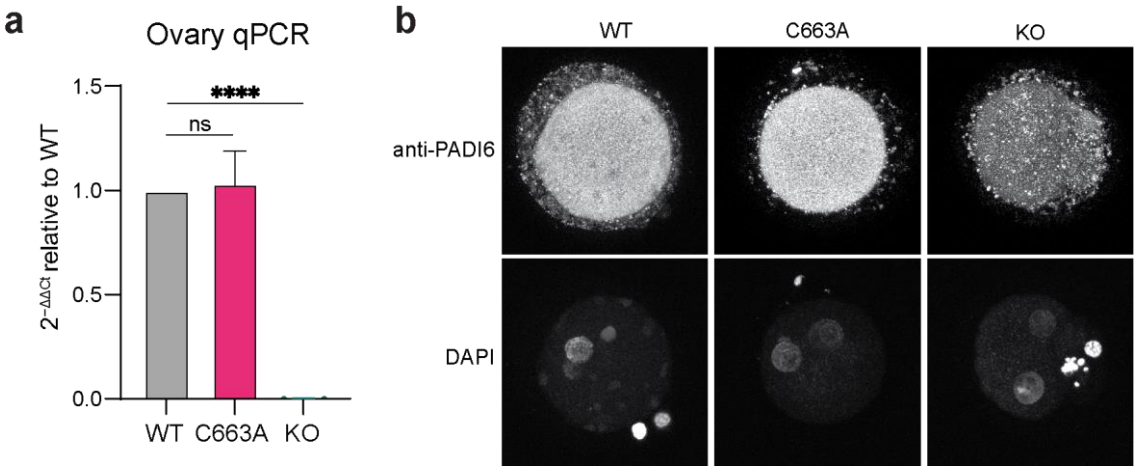


Figure 5.4 Transgenic mouse line validation
a) Relative *Padi6* expression levels in ovaries from 8-week-old WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females determined by qRT-PCR. The data were normalised to *B2m* and the relative expression of *Padi6* determined compared to WT type ovaries. Un-paired parametric t-test performed to determine statistical significance. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. **b)** Representative max intensity Z-projection images of immunostaining for PADI6 (anti-PADI6) and DNA (DAPI) in zygotes from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

Together these data demonstrate that the mutation to substitute C663 to A has been successfully incorporated, and it has no effect on *Padi6* mRNA levels, and protein level or distribution in the ovary and zygote respectively. Furthermore, as expected no *Padi6* mRNA is present in *Padi6*$^{-/-}$ ovaries, and a significantly reduced signal is observed in mat-*Padi6*$^{-/-}$ zygotes, with the remaining background signal likely the result of unspecific antibody interactions.

5.3 Fertility assessment of *Padi6* transgenic female mice

Padi6$^{-/-}$ females have been reported to be infertile. The first step in identifying any developmental phenotypes of *Padi6*$^{C663A/C663A}$ females was to assess their fertility and any developmental phenotypes in their embryos. WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females were set up in natural matings with WT males and the size of litters produced was recorded (**Figure 5.5a**). As expected, *Padi6*$^{-/-}$ females were incapable of producing any pups. Interestingly however, *Padi6*$^{C663A/C663A}$ females were fertile and produced litters of comparable size to WT females. This result immediately informs us that whilst there may still be a catalytic function of PADI6 in mice, it is not essential for embryo development.

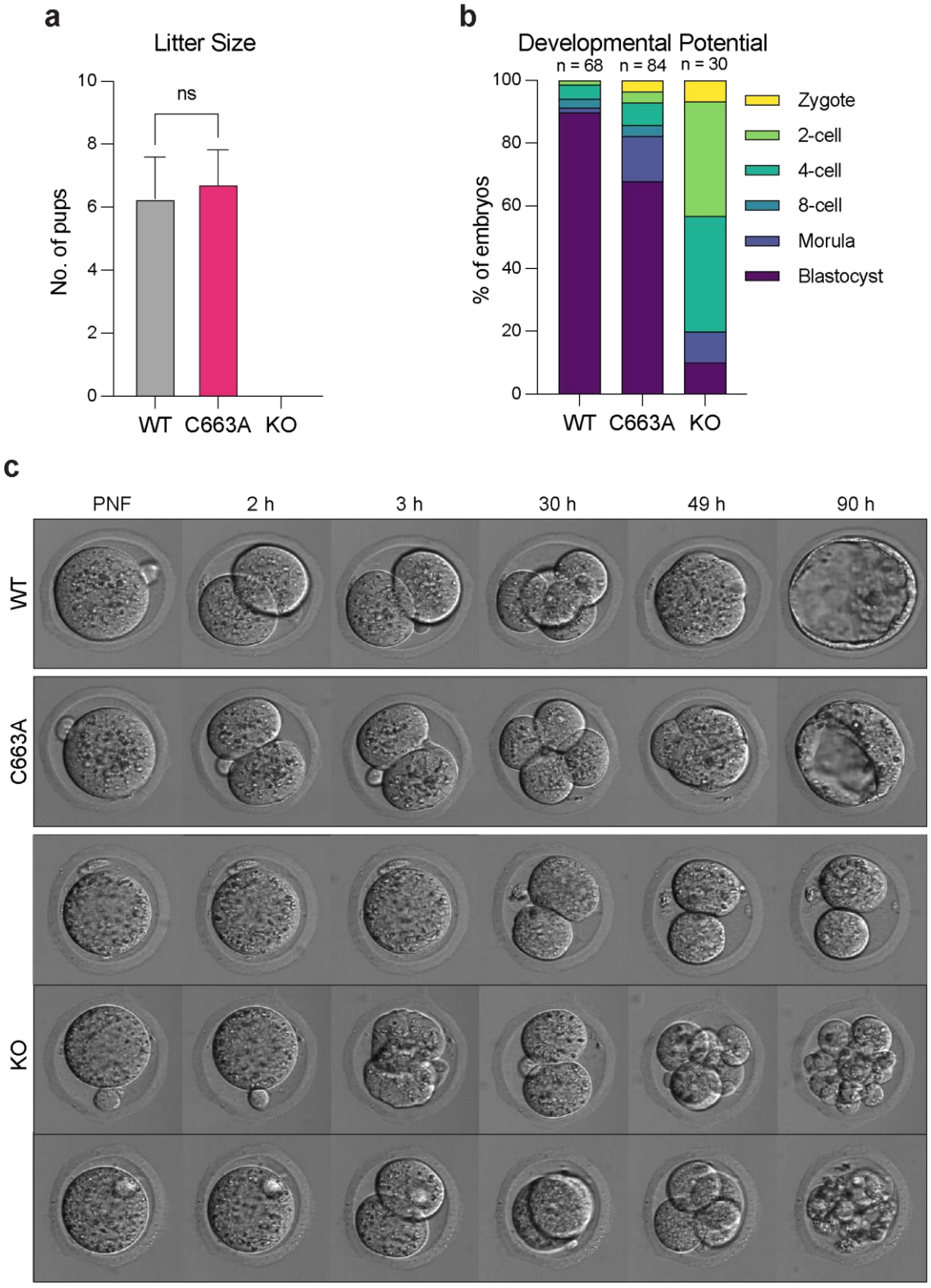


Figure 5.5 Mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ embryo developmental potential
a) Litter size of WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females when naturally mated with WT males. Unpaired t-test performed on WT and *Padi6*$^{C663A/C663A}$ litter sizes. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. **b)** Percentage of embryos that either arrested or degenerated at each developmental stage, or successfully formed a blastocyst. Regarding embryos from *Padi6*$^{-/-}$ females, blastocysts appeared abnormal and are classed as blastocyst like and not full

blastocyst. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. **c)** Representative time-lapse snapshots of the development of embryos from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females at PNF, and 2, 3, 30, 49, and 90 h post PNF. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

To gain a better understanding of how embryos from each line were developing, WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females were naturally mated with WT males and zygotes were extracted by Afshan McCarthy of the Niakan lab at the Crick and imaged for a period of 5 days using an EmbryoScope+ incubator. I then classified embryos by what developmental stage they successfully reached before either arresting or degenerating, as well as the time taken to reach each developmental stage (**Figure 5.5b**).

Interestingly, fewer embryos from *Padi6*$^{C663A/C663A}$ females (67.9%) successfully reached the blastocyst stage compared to WT (82.4%) although the blastocysts that were formed appeared healthy (**Figure 5.5c**). Mat-*Padi6*$^{C663A/C663A}$ embryos that did not successfully reach the blastocyst stage either arrested or degenerated primarily at the 4-cell and morula stages, however some embryos did fail at the zygote, 2-cell and 8-cell stages. These data suggests whilst *Padi6*$^{C663A/C663A}$ females were fertile, embryos from these females are partially inhibited in their development.

The observed phenotype of knocking-out *Padi6* was significantly more pronounced with most mat-*Padi6*$^{-/-}$ embryos arresting at the 2-cell stage or reaching the 4-cell stage and then fragmenting and degenerating rapidly (**Figure 5.5c**). Embryos that did continue to develop beyond the 4-cell stage experienced high levels of fragmentation, un-synchronised and un-symmetric divisions, and eventual degeneration and apoptosis.

Surprisingly, three out of the 30 zygotes from *Padi6*$^{-/-}$ females did form blastocyst like embryos, however, they looked abnormal compared to WT blastocysts. Given no litters were ever obtained from *Padi6*$^{-/-}$ females, it is likely whilst these embryos resemble blastocysts, they are too unhealthy to develop further. This contrasts with the reported effects of knocking out *Padi6* in female mice where all embryos arrest at the 2-cell stage, however (32). One explanation could be that the mice used in the reported work were super-ovulated females. In a 2021 report by Taher *et al.*, the effect of superovulation on mouse oocyte and embryos was investigated from a proteomic perspective (274). It was found that superovulation is associated with proteome

changes which result in phenotypic traits in the embryos, including a thinner zona pellucida, smaller oocyte diameter, defective blastocyst formation and increased frequency of cleavage arrest. The difference in the developmental arrest timings of embryos from superovulated females in the previously reported work, and the embryos from naturally mated females in this work, could be therefore explained by superovulation sensitising embryos to molecular changes and accentuating the developmental phenotype. It is conceivable that this hard 2-cell arrest in the reported work was therefore an artifact of superovulation. We propose that knocking-out *Padi6* in mice does not result in a strict developmental arrest at the 2-cell stage, but a severely inhibited development where the majority arrest between the 2- and 4-cell stage. Furthermore, the few embryos that do go on to develop blastocyst-like embryos are very abnormal and fail to develop further, thus resulting in female infertility.

The time at which embryos reached each developmental stage relative to pronuclear fading (PNF) was determined, a summary of these developmental timings is displayed in **Figure 5.6**. The developmental timings of embryos from *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females mirrors their developmental capacity. Mat-*Padi6*$^{C663A/C663A}$ embryos divided into 2-cells on average at the same time as WT embryos, with median timings of 1.7 h post PNF for both. The range of timings for *Padi6*$^{C663A/C663A}$ embryos was larger however, with some embryos dividing at up to 3.4 h post PNF, compared to a latest time of 2.1 h in WT embryos. Interestingly, mat-*Padi6*$^{C663A/C663A}$ embryos were slightly delayed in reaching the 4-cell, 8-cell and morula stages compared to the WT, with Mann-Whitney p-values equalling 0.0147, 0.001 and 0.0001 respectively. This delay aligns with the observation that embryos that do not successfully reach the blastocyst stage primarily arrest or degenerate between the 4-cell and morula stages (**Figure 5.5c**). The coincidence of the start of this delay and partial developmental inhibition aligns with the timing of minor and major EGA in mice, suggesting there may be an additional non-essential catalytic function of PADI6 during EGA. The developmental delay was mostly rescued beyond the morula stage with no significant difference between WT and mat-*Padi6*$^{C663A/C663A}$ embryo development timings for cavity and blastocyst formation.

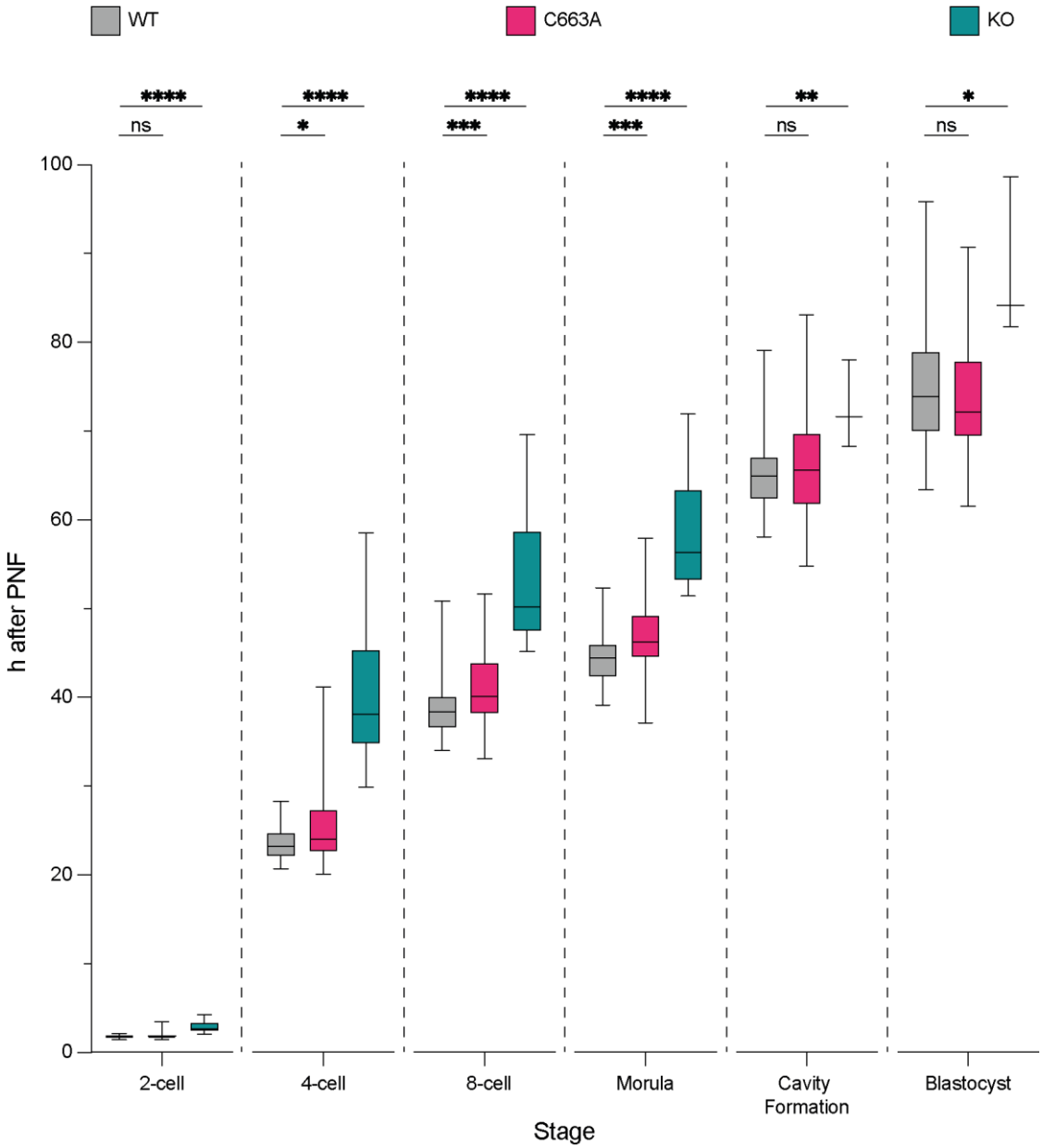


Figure 5.6 Mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ embryo developmental timings
PNF normalised times at which embryos from WT (grey), *Padi6*$^{C663A/C663A}$ (magenta) and *Padi6*$^{-/-}$ (green) females reach each developmental stage. Mann-Witney test performed to determine statistical significance. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

Mat-*Padi6*$^{-/-}$ embryos on the other hand were significantly delayed at reaching every developmental stage compared to WT embryos (**Figure 5.5c and 5.6**). Mat-*Padi6*$^{-/-}$ embryos reach the 2-cell stage, on average 1.2 h later than WT embryos. Most notably however, the knock-out embryos reach the 4-cell stage at 40.1 h post PNF on average, compared to 23 h for WT embryos. Interestingly, mat-*Padi6*$^{-/-}$ embryos that do develop beyond the 4-cell stage then develop at a pace comparable to or quicker than that of WT embryos (**Table 5.1**). Together, this suggests that PADI6 is most important

during the zygote to 2-cell and 2-cell to 4-cell transitions. A lack of PADI6 results in significant delays in transitioning to these stages, and in most cases the arrest or degeneration of the embryo. These stages align with the less pronounced developmental delay and higher incidence of arrest and degeneration observed in mat-*Padi6*$^{C663A/C663A}$ embryos. Together, these data substantiates the theory that PADI6 has an important function in EGA. For the few embryos that do successfully develop beyond the 4-cell stage, they transition to each stage in a comparable time to WT embryos. Although as discussed previously, these embryos develop with high levels of fragmentation, un-synchronised and un-symmetric divisions, and eventual degeneration and apoptosis.

Table 5.1 WT and mat-*Padi6*$^{-/-}$ embryo transition times
Average transition time between developmental stages for embryos from WT and *Padi6*$^{-/-}$ females.

Transition time	WT / h	Mat-*Padi6*$^{-/-}$ / h
Zygote to 2-cell	1.7	2.9
2-cell to 4-cell	21.7	37.2
4-cell to 8-cell	15.4	13.1
8-cell to Morula	5.5	5.1
Morula to Cavity Formation	21.1	14.5
Cavity Formation to Blastocyst	10.2	16.0

5.4 Cytoplasmic lattices in *Padi6* transgenic oocytes

As discussed in **Chapter 1.2.1**, one of the main phenotypes reported in *Padi6*$^{-/-}$ oocytes is the absence of the CPLs. It is therefore of interest to determine whether they are also absent in *Padi6*$^{C663A/C663A}$ oocytes. Historically, the only clear images of the CPLs have been taken with Transmission Electron Microscopy (TEM). To form TEM images, electron beams are transmitted through a thin slice of sample (typically < 100 nm) and onto a detection screen or camera. The significantly shorter wavelength of electrons than of visible light allow much higher resolution images to be obtained. Samples must first be processed through a series of fixation and staining steps before being embedded in resin, sliced into <100 nm sections using a diamond knife and imaged.

Four WT, six *Padi6*$^{C663A/C663A}$ and four *Padi6*$^{-/-}$ GV oocytes were harvested and fixed with glutaraldehyde and formaldehyde before being handed off to Anne Weston of the Microscopy STP for processing for TEM. As expected, no CPLs were observed in any of the *Padi6*$^{-/-}$ oocytes (**Figure 5.7a-b**). Three out of the four, and five out of the six oocytes from WT and *Padi6*$^{C663A/C663A}$ females possessed CPLs. The CPLs appeared to be similarly distributed throughout the oocyte in each line, and similar in size and morphology (**Figure 5.7a**). The lack of CPLs in one of the WT and one of the *Padi6*$^{C663A/C663A}$ oocytes is possibly due to natural defects occurring in those oocytes. As such, from these results it is clear that PADI6 catalytic activity is not involved in the formation or maintenance of the CPLs in mice.

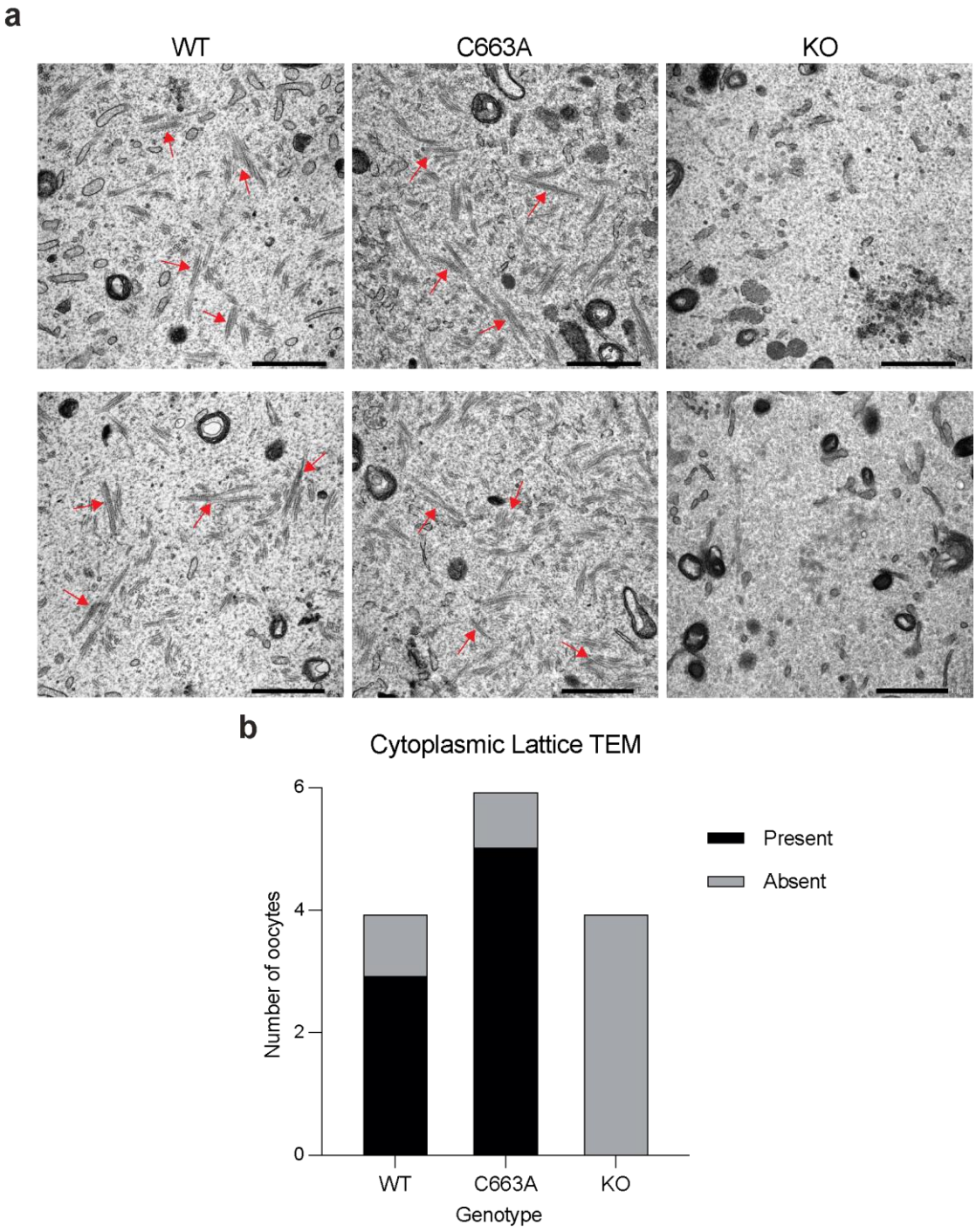


Figure 5.7 CPLs in GV oocytes from WT, *Padi6*^{C663A} and *Padi6*$^{-/-}$ females
a) Representative TEM images of two GV oocytes from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females. Scale bar = 1 μm. Representative CPLs indicated by red arrows. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.**b)** Absence or presence of CPLs in four, six, and four GV oocytes from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females respectively. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

5.5 Transcriptome characterisation of *Padi6* transgenic oocytes and embryos

So far, characterisation of possible phenotypes in *Padi6*$^{C663A/C663A}$ females has been inconclusive in determining whether PADI6 has a catalytic function in mice. Our current data suggests there may be an additional non-essential catalytic function of PADI6, however, more data is needed to confirm this. Furthermore, whilst the reported necessity of PADI6 in early embryo development has been validated, and the data points towards a function in EGA, we must gather further information to conclusively validate this.

RNA-sequencing is a commonly employed method to capture a snapshot of the transcriptome in cells, giving an indication of which proteins are being expressed and to what levels in the cells. By assessing transcriptomic differences between WT and genetically modified cells, information can be gathered on the molecular functions of target proteins by identifying genes that are up or downregulated. Bulk RNA-sequencing methods require large input quantities and will average out any differences between cells within a sample, and as such is not appropriate for this work. Single cell RNA sequencing (scRNA-seq) however, allows the transcriptome characterisation of single cells, massively reducing the amount of input material required and allowing for direct comparison between the transcriptome of individual cells. As such, scRNA-seq was the obvious first choice for characterising molecular differences between oocytes and embryos from *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females, and those from WT females.

5.5.1 Experimental design

The coincidence of developmental arrest timings of *Padi6*$^{-/-}$ and human *PADI6* variant embryos with EGA in mice and humans respectively suggests a possible function for PADI6 in EGA. Time-lapse imaging of embryos from both *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females in this work further points towards an inhibition in EGA with either a predicted catalytically defective, or complete absence of, PADI6. As discussed in the introduction, minor EGA occurs at the zygote stage in mice, and major EGA the 2-cell stage. As such, zygotes and 2-cell embryos are the clear choices for scRNA-seq to identify any transcriptomic differences during EGA in mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ embryos. To identify any differences before fertilisation and EGA that could potentially cause downstream EGA impairment, the transcriptome of MII oocytes from

transgenic embryos is also of interest. We therefore decided that MII oocytes, zygotes and 2-cell embryos from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females would be subjected to scRNA-seq. To account for any litter effects (variation within litters and between litters from different females), for each stage and genotype, three oocytes or embryos from mice from two separate litters were collected. To separate out any asymmetry in the transcriptomes of 2-cell embryos, they were dissociated into single blastomeres and processed as single cells. Altogether this resulted in 6 x MII oocytes and 6 x zygotes per genotype, and 12 x 2-cell blastomeres per genotype, with a total of 72 individual embryos or cells.

5.5.2 Sample harvesting and library preparation

To obtain MII oocytes, WT or *Padi6* transgenic females were naturally mated with a sterile male. Upon mating, non-fertilised MII oocytes released into the oviduct were extracted. Zygotes and 2-cell embryos were harvested after natural mating with a WT male. To extract zygotes, plugged females were culled at E0.5, zygotes were extracted and only those that possessed clear pronuclei and two polar bodies were taken forward. To extract 2-cell embryos, oviducts of plugged females were flushed at E1.5 to release the embryos. The zona pellucida of the 2-cell embryos was then removed using acidic Tyrode's solution, and embryos treated with accutase to facilitate subsequent dissociation of blastomeres. MII oocytes, zygotes and 2-cell blastomeres were then transferred into individual wells of 96-well plates containing SMART-seq lysis buffer, RNase inhibitor, and first strand synthesis primer, before being sealed, snap frozen and stored at -80 ^{o}C. Sample gathering was conducted over 18 sessions, resulting in 18 separate 96-well sample plates. Upon the collection of all required cells, samples in all plates were consolidated into a single submission plate and submitted to the Crick Advanced Sequencing Facility for library preparation. DNA libraries were prepared using the Takara SmartSeqHT kit for initial amplification, and the Nextera XT kit for library indexing. Libraries were pooled, sequenced, and mapped to the mouse genome and gene read counts extracted using the Nextflow RNA-seq pipeline.

5.5.3 Data quality analysis

Across the 72 cells, a total of 24411 genes were detected. Genes detected in fewer than six cells or with fewer than 1000 reads were removed. This resulted in 14,411 genes being removed leaving 10,000 genes. After this filtering, the total counts

for each cell was compared with the number of genes identified in each cell to identify any outlier cells (**Figure 5.8.a**). 71 out of the 72 cells clustered around 8500 to 10,000 genes identified, with between 2 and 8 million total counts. One cell (mat-*Padi6*$^{-/-}$ 2-cell embryo 1c blastomere 2) had fewer genes identified at only 7912 with 1.9 million total counts. Gene reads were then normalised to counts per million total reads (CPM) and converted to $\log_2$ CPM (logCPM). Principle component analysis (PCA) of the entire dataset and comparison with total read count of each cell showed principal components were not influenced by total read count (**Figure 5.8b**). Cells predominantly clustered by developmental stage, with the exception of *Padi6*$^{C663A/C663A}$ MII oocytes 1a, 1b and 1c, and *Padi6*$^{-/-}$ MII oocyte 2b which differed along PC2 (**Figure 5.8c**). MII oocytes and zygotes clustered closer together and away from the 2-cell blastomeres. This is expected as few new genes will be transcribed at the zygote stage during minor EGA compared to the increase in transcription during major EGA at the 2-cell stage.

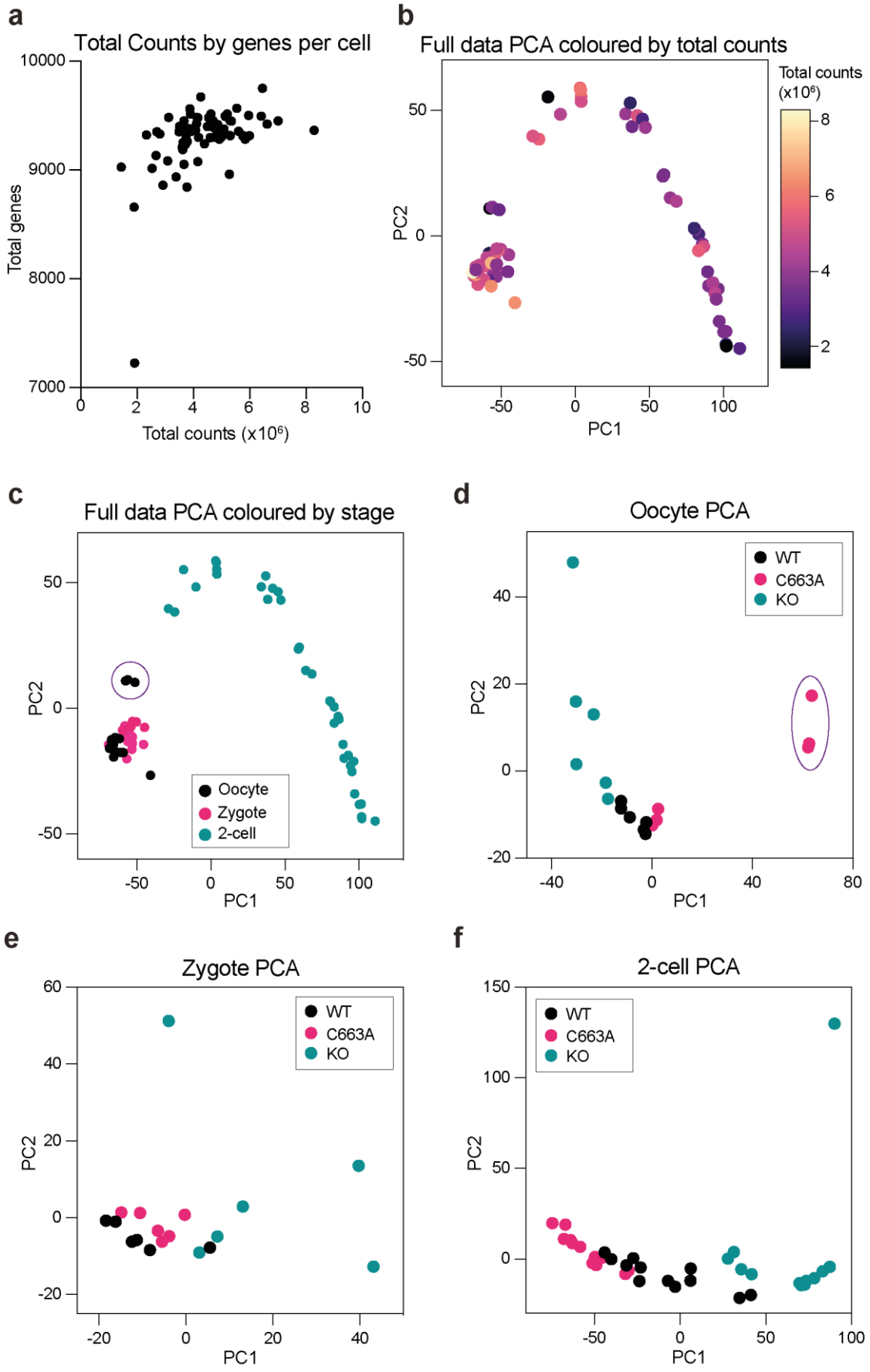

a
Total Counts by genes per cell
Total genes
Total counts (x10⁶)
b
Full data PCA coloured by total counts
Total counts (x10⁶)
PC1
PC2
c
Full data PCA coloured by stage
Oocyte
Zygote
2-cell
d
Oocyte PCA
WT
C663A
KO
e
Zygote PCA
f
2-cell PCA

Figure 5.8 scRNA-seq sample QC and clustering
a) Scatter plot of total genes identified and the number of total counts per cell. **b)** PCA plot for all cells calculated using the Python ScanPy package, coloured by total counts per cell. **c)** PCA plot for all cells calculated using the Python ScanPy package, coloured by developmental stage. *Padi6*$^{C663A/C663A}$ MII oocytes 1a, 1b, and 1c circled. **d)-f)** PCA plot for all MII oocytes, zygote and 2-cell blatomeres calculated using the Python ScanPy package, coloured by genotype. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. *Padi6*$^{C663A/C663A}$ MII oocytes 1a, 1b, and 1c circled.

Cells were divided by developmental stage and clustered. In the MII oocytes, WT oocytes clustered together, close to three of the *Padi6*$^{C663A/C663A}$ oocytes (**Figure 5.8d**). The *Padi6*$^{-/-}$ oocytes were more dispersed and did not cluster with the WT oocytes suggesting even at the MII stage there are large transcriptomic differences. As seen in the full data clustering, *Padi6*$^{C663A/C663A}$ MII oocytes 1a, 1b, and 1c clustered separately and distinctly from the others. These three *Padi6*$^{C663A/C663A}$ MII oocytes are from a separate litter to the three that cluster with the WT oocytes. This difference could therefore be due to a large litter variation between females, or other defects picked up during harvesting. At the zygote stage, embryos clustered similarly to the MII oocytes, with WT zygotes clustering fairly tightly together with mat-*Padi6*$^{C663A/C663A}$ zygotes (**Figure 5.8e**). As before, mat-*Padi6*$^{-/-}$ zygotes were more dispersed and did not cluster with the WT zygotes. At the 2-cell stage, blastomeres from each genotype clustered mostly together, predominantly separated by their PC1 contribution (**Figure 5.8f**). One mat-*Padi6*$^{-/-}$ blastomere did not appear to cluster with the others. An explanation for this could be unsymmetric 2-cell cleavage which has been observed in previous experiments. Whilst there are small overlaps between the clusters, mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ blastomeres did appear to cluster distinctly from the WT blastomeres suggesting at the 2-cell stage there are increased transcriptomic differences between each genotype. This is to be expected as major EGA occurs at the 2-cell stage. As PADI6 is predicted to be involved in EGA then it follows that any changes to PADI6 function may result in transcriptomic differences (20).

As expected, there are comparable CPM values for *Padi6* in WT and *Padi6*^{C663A} oocytes and embryos, with negligible values in *Padi6*$^{-/-}$ (**Figure 5.9**). It appears that expression of *Padi6* is slightly lower in the *Padi6*$^{C663A/C663A}$ MII oocyte compared to WT, and slightly higher in the zygote, however. As discussed in **Chapter 1.2.4**, data concerning the expression of PADIs 1-4 in the oocyte and early embryo is conflicting. RNA-seq and proteomics of mouse oocytes and embryos show no reads or peptides detected for any PADIs other than PADI6 (61). On the other hand, PADIs 1 to 4 have been observed in the early embryo by immunofluorescent staining (148). No reads for

any of the other PADIs were detected in any of the cells analysed in our dataset. As such, we propose that PADI6 is the dominant and likely only PADI expressed during these early stages of embryo development.

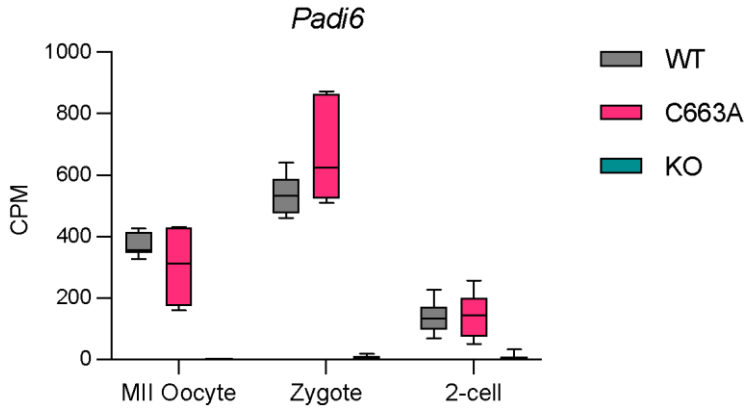


Figure 5.9 CPM of *Padi6* in MII oocytes, zygotes and 2-cell blastomeres
Average CPM of *Padi6* in MII oocytes, zygotes and 2-cell blastomeres from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

Regarding the expression of SCMC genes, varying differences are observed between mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ MII oocytes and embryos. *Ooep* expression appears comparable between WT, *Padi6*^{C663A} and knock-out oocytes and embryos, however slightly downregulated in *Padi6*$^{C663A/C663A}$ MII oocytes and upregulated in mat-*Padi6*$^{-/-}$ 2-cell embryos (**Figure 5.10a**). *Tle6* on the other hand, whilst also slightly down regulated *Padi6*$^{C663A/C663A}$ MII oocytes is dramatically upregulated in zygotes (**Figure 5.10b**). In 2-cell embryos, *Tle6* then appears slightly upregulated in both mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ compared to WT. In MII oocytes, expression of *Nlrp5* (gene of NALP5) are comparable between samples, however it appears slightly downregulated in mat-*Padi6*$^{-/-}$ zygotes and then upregulated in 2-cell embryos (**Figure 5.10c**). Expression levels for *Khdc3* are much lower than the *Ooep*, *Nlrp5* and *Tle6*, and appear comparable across all stages but are slightly upregulated in *Padi6*$^{-/-}$ MII oocytes, and mat-*Padi6*$^{C663A/C663A}$ zygotes (**Figure 5.10d**). *Zbed3* has significantly higher expression levels and is comparable at the MII oocyte and zygote stage, however, is then downregulated in mat-*Padi6*$^{C663A/C663A}$ but upregulated in mat-*Padi6*$^{-/-}$ 2-cell embryos (**Figure 5.10e**). Already at this point in the analysis it appeared that there were transcriptomic differences between mat-*Padi6*$^{C663A/C663A}$ or mat-*Padi6*$^{-/-}$ oocytes and embryos, compared to WT. Possibly the most interesting difference is that *Tle6* appeared upregulated in mat-*Padi6*$^{C663A/C663A}$ zygotes but not WT or mat-*Padi6*$^{-/-}$. A

mechanistic explanation for this is currently unclear, however, this result further highlights possible differences between mat-*Padi6*$^{C663A/C663A}$ and WT embryos.

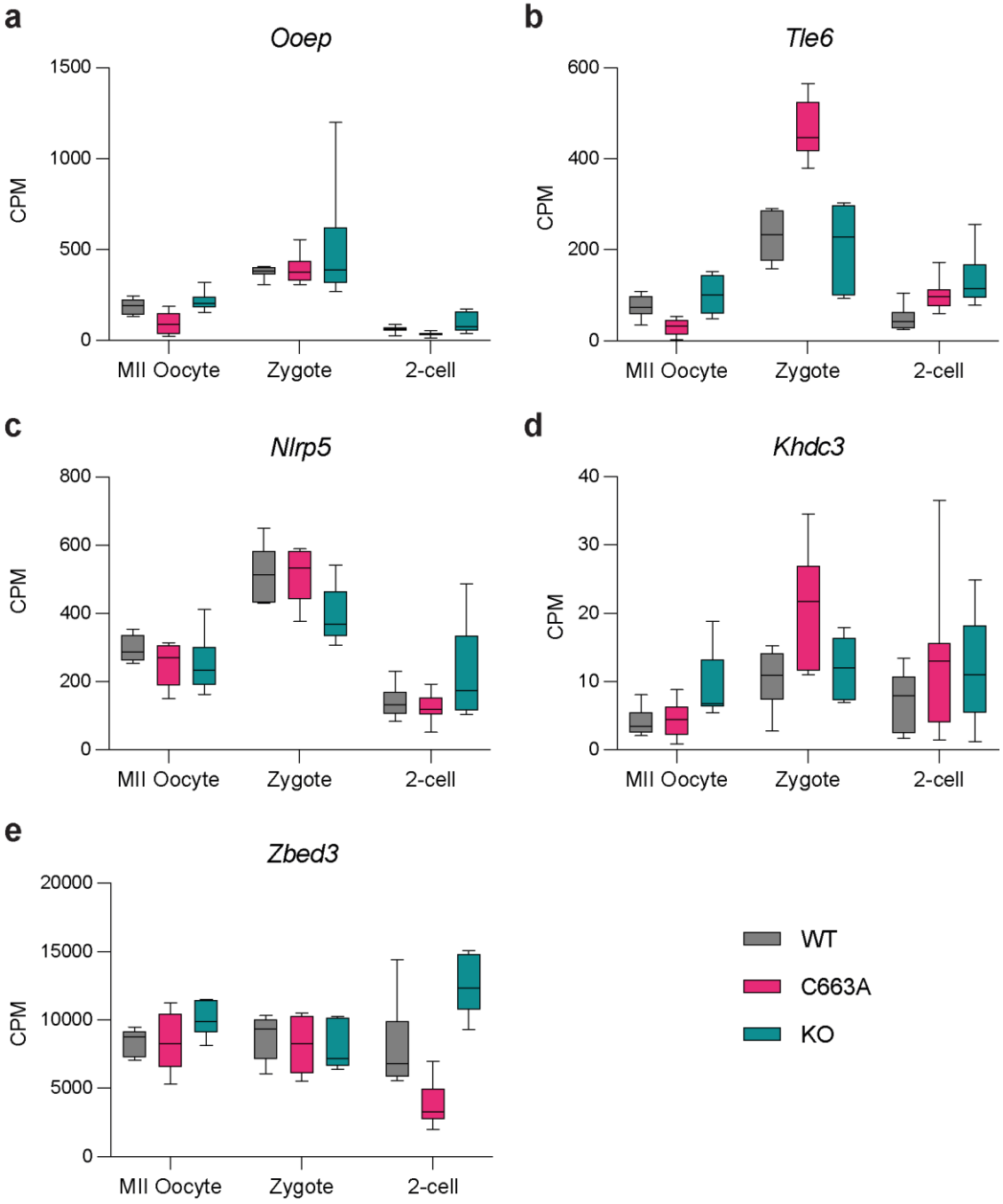


Figure 5.10 CPM of SCMC genes in MII oocytes, zygotes and 2-cell blastomeres
a)-e) Average CPM *Ooep, Tle6, Nlrp5, Khdc3, and Zbed3,* in MII oocytes, zygotes and 2-cell blastomeres from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

5.5.4 Transcriptome differences compared to WT oocytes and embryos

The transcriptomic differences between oocytes and embryos from *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females compared to WT were then interrogated further to identify up and downregulated genes at each stage. Genes that had a CPM fold-change [mat-*Padi6*$^{C663A/C663A}$/WT] or [mat-*Padi6*$^{-/-}$/WT] of greater than 2, with a false discovery rate (FDR) of 0.05 were classed as upregulated. Those with a fold-change of less than 0.5 and FDR of 0.05 were classed as downregulated. Differentially expressed genes were then visualised using an MA plot where the logCPM ratio is plotted against the average logCPM for each gene.

5.5.4.1 MII oocyte

In the MII oocyte, no genes were classed as significantly affected in *Padi6*$^{C663A/C663A}$ oocytes, whereas 60 and 17 were up and downregulated respectively in *Padi6*$^{-/-}$ oocytes (**Figure 5.11**). The lack of affected genes in *Padi6*$^{C663A/C663A}$ oocytes is likely due to the dispersion in the data (see **Chapter 5.5.3**) meaning no genes were accepted in the FDR check. Overall, whilst there are transcriptomic differences at this stage in the *Padi6*$^{-/-}$ oocytes compared to WT, they are minimal.

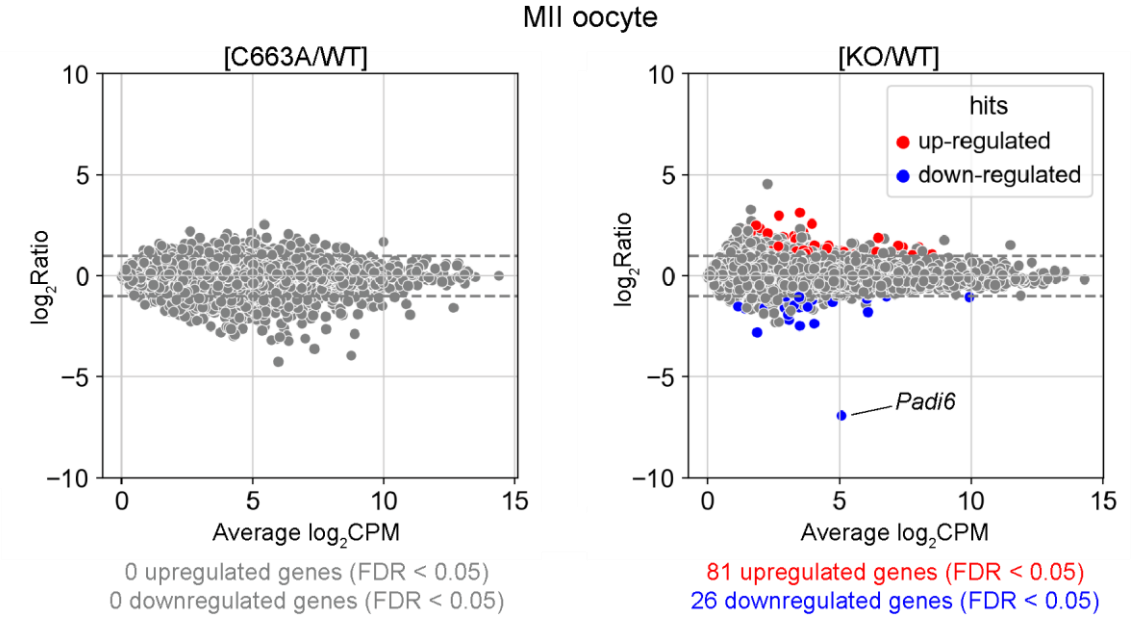


Figure 5.11 Up and downregulation of genes in the MII oocyte
Left Panel: MA plot of [*Padi6*$^{C663A/C663A}$/WT] log_2Ratio vs. average log_2CPM of genes in MII oocytes. Right panel: MA plot of [*Padi6*$^{-/-}$/WT] log_2Ratio versus average log_2CPM of genes in MII oocytes. Genes were selected that were more than 2-fold upregulated (red) or downregulated (blue) versus WT with an FDR of 0.05. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

5.5.4.2 Zygote

In the zygote, both mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ embryos had up and downregulation of genes (**Figure 5.12a**). 22 genes were upregulated in mat-*Padi6*$^{C663A/C663A}$ zygotes, with only one gene downregulated. 31 genes were upregulated in mat-*Padi6*$^{-/-}$ zygotes with 32 genes downregulated. Interestingly, in the mat-*Padi6*$^{-/-}$ zygotes, fewer genes were upregulated than at the MII oocyte stage. Gene ontology (GO) of each group of up or downregulated genes did not result in any biological pathways being enriched in any group.

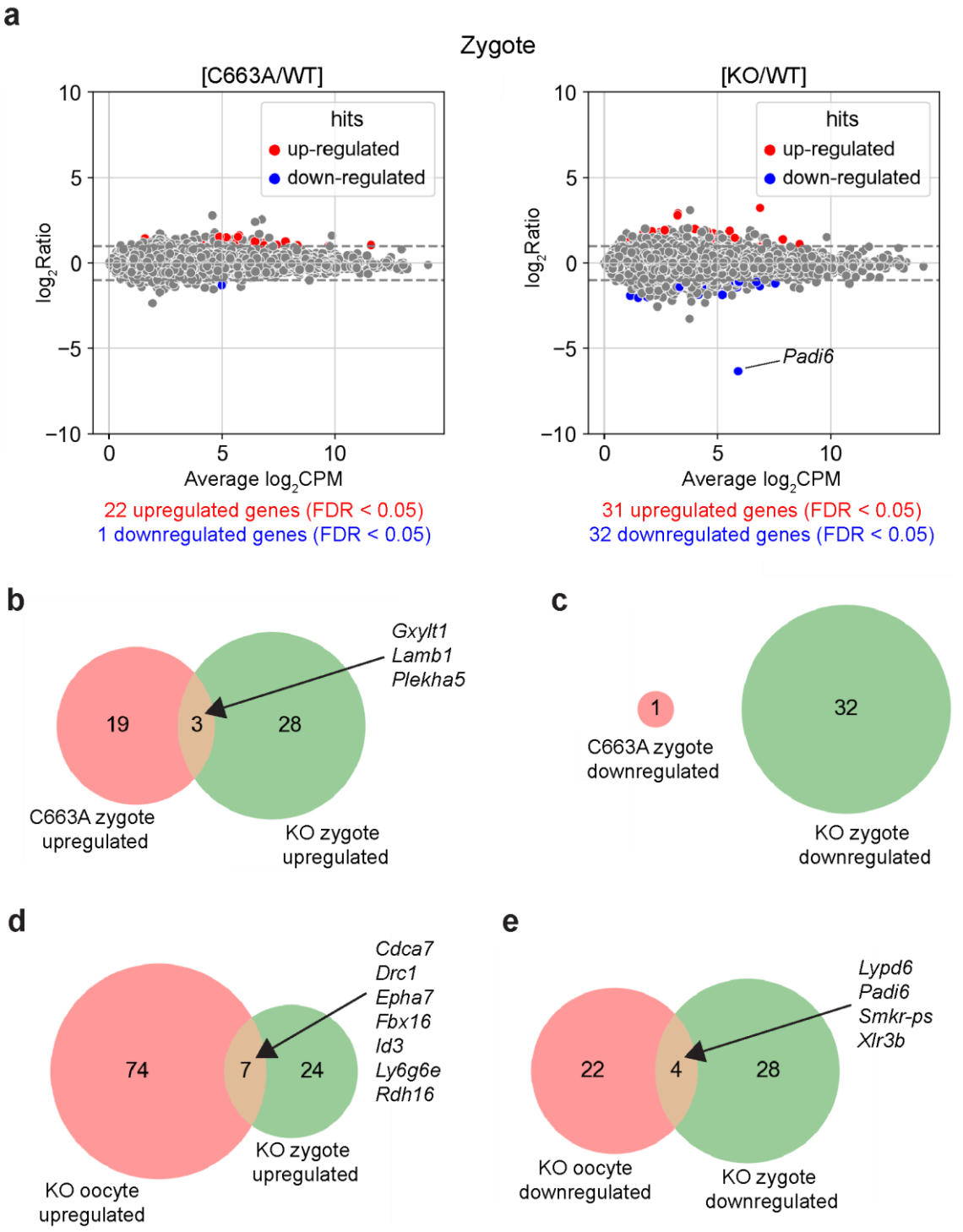


Figure 5.12 Up and downregulation of genes in the zygote

a) Left Panel: MA plot of [mat-*Padi6*$^{C663A/C663A}$/WT] log_2Ratio vs. average log_2CPM of genes in zygotes. Right panel: MA plot of [mat-*Padi6*$^{-/-}$/WT] log_2Ratio versus average log_2CPM of genes in zygotes. Genes were selected that were more than 2-fold upregulated (red) or downregulated (blue) versus WT with an FDR of 0.05. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. **b)-c)** Venn diagrams of common upregulated and downregulated genes in mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ zygotes. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. **d)-e)** Venn diagram of common upregulated and downregulated genes in *Padi6*$^{-/-}$ oocytes and mat-*Padi6*$^{-/-}$ zygotes. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

Three genes were commonly upregulated between mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ zygotes: *Gxylt1, Lamb1, and Plekha5* **(Figure 5.12b)**, no genes were commonly downregulated (**Figure 5.12c**). Comparing mat-*Padi6*$^{-/-}$ zygotes to *Padi6*$^{-/-}$ MII oocytes, seven genes were commonly upregulated between *Padi6*$^{-/-}$ MII oocytes and mat-*Padi6*$^{-/-}$ zygotes: *Cdca7, Drc1, Epha7, Fbx16, Id3, Ly6g6e,* and *Rdh16* **(Figure 5.12d)**. Four genes were commonly downregulated between *Padi6*$^{-/-}$ MII oocytes and mat-*Padi6*$^{-/-}$ zygotes: *Lypd6, Padi6, Smkr-ps,* and *Xlr3b* **(Figure 5.12e)**.

5.5.4.3 2-cell embryo

At the 2-cell stage, up and downregulation of hundreds of genes was observed in both mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ 2-cell embryos (**Figure 5.13a**). 1061 genes were upregulated in mat-*Padi6*$^{C663A/C663A}$ 2-cell blastomeres, with 459 downregulated. Only 604 genes were upregulated in mat-*Padi6*$^{-/-}$ 2-cell blastomeres compared to 1993 genes being downregulated. There were few genes commonly up or downregulated in mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ 2-cell blastomeres, at 15 and 26 respectively (**Figure 5.13b-c**). However, 579 genes that were upregulated in mat-*Padi6*$^{C663A/C663A}$ 2-cell blastomeres were downregulated in mat-*Padi6*$^{-/-}$, with only 53 common between mat-*Padi6*$^{C663A/C663A}$ downregulated and mat-*Padi6*$^{-/-}$ upregulated (**Figure 5.13d-e**).

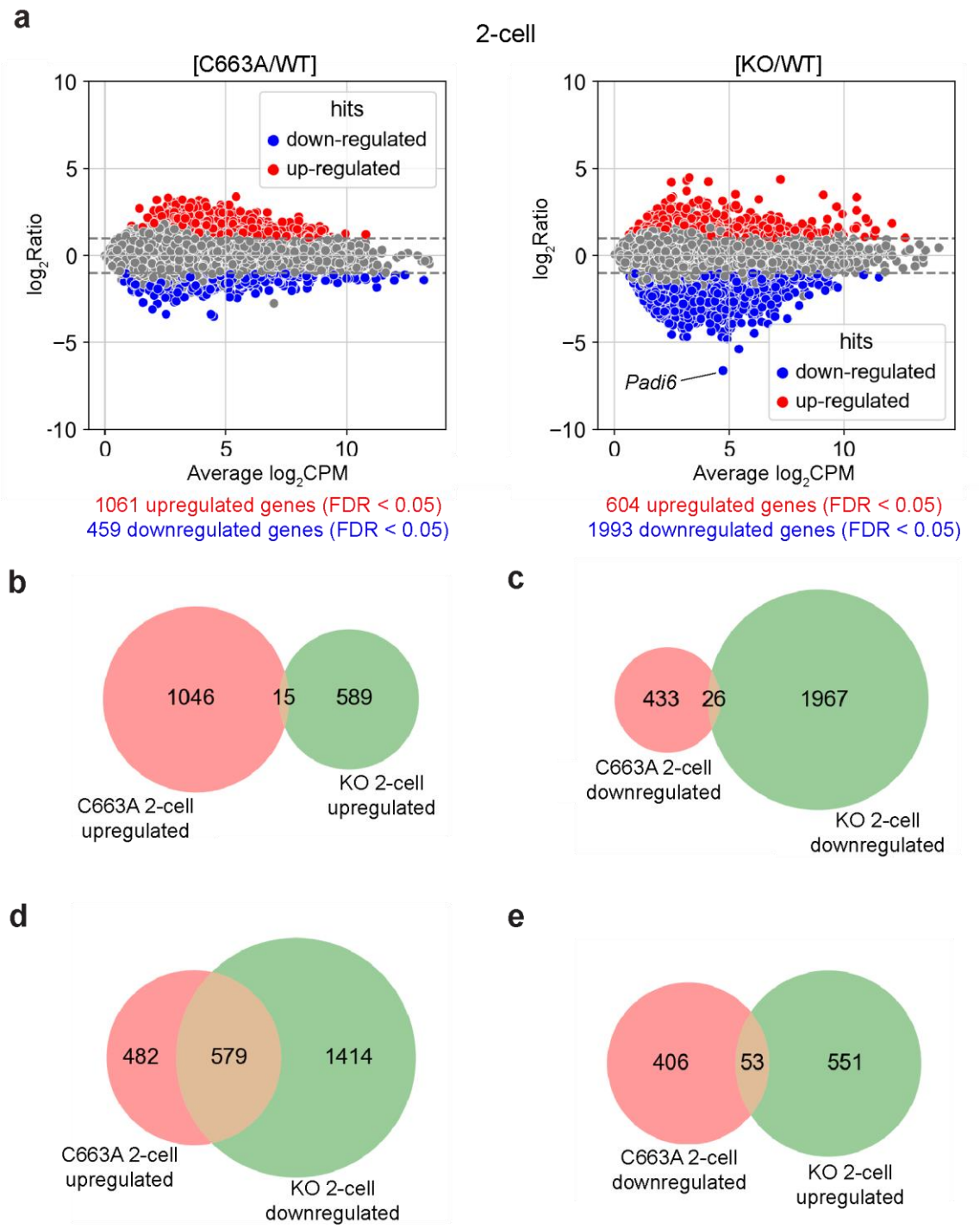


Figure 5.13 Up and downregulation of genes in the 2-cell blastomere
a) Left Panel: MA plot of [mat-*Padi6*$^{C663A/C663A}$/WT] log_2Ratio vs. average log_2CPM of genes in 2-cell blastomeres. Right panel: MA plot of [mat-*Padi6*$^{-/-}$/WT] log_2Ratio versus average log_2CPM of genes in 2-cell blastomeres. Genes were selected that were more than 2-fold upregulated (red) or downregulated (blue) versus WT with an FDR of 0.05. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. **b)-c)** Venn diagrams of common upregulated and downregulated genes in mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ 2-cell embryos. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. **d)** Venn diagram of common upregulated genes in mat-*Padi6*$^{C663A/C663A}$ 2-cell blastomeres and downregulated genes in mat-*Padi6*$^{-/-}$ 2-cell blastomeres. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. **e)** Venn diagram of common downregulated genes in mat-*Padi6*$^{C663A/C663A}$ 2-cell

blastomeres and upregulated genes in mat-*Padi6*$^{-/-}$ 2-cell blastomeres. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

GO analysis of biological processes (BP) of up and downregulated genes from mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ 2-cell blastomeres showed a similar opposite effect in transcript regulation. Mat-*Padi6*$^{C663A/C663A}$ 2-cell downregulated and mat-*Padi6*$^{-/-}$ 2-cell upregulated genes both showed slight enrichment for the mRNA export from nucleus (**Figure 5.14a-b**). On the other hand, mat-*Padi6*$^{C663A/C663A}$ 2-cell upregulated genes showed moderate enrichment for translation, with slight enrichment for protein transport, cytoplasmic translation and mitochondrial translation (**Figure 5.14c**). Mat-*Padi6*$^{-/-}$ 2-cell downregulated genes showed a similar but much stronger enrichment, with cytoplasmic translation, translation, ribosomal small subunit biogenesis, mRNA processing, RNA splicing, protein transport, and the cell cycle as the top terms (**Figure 5.14d**).

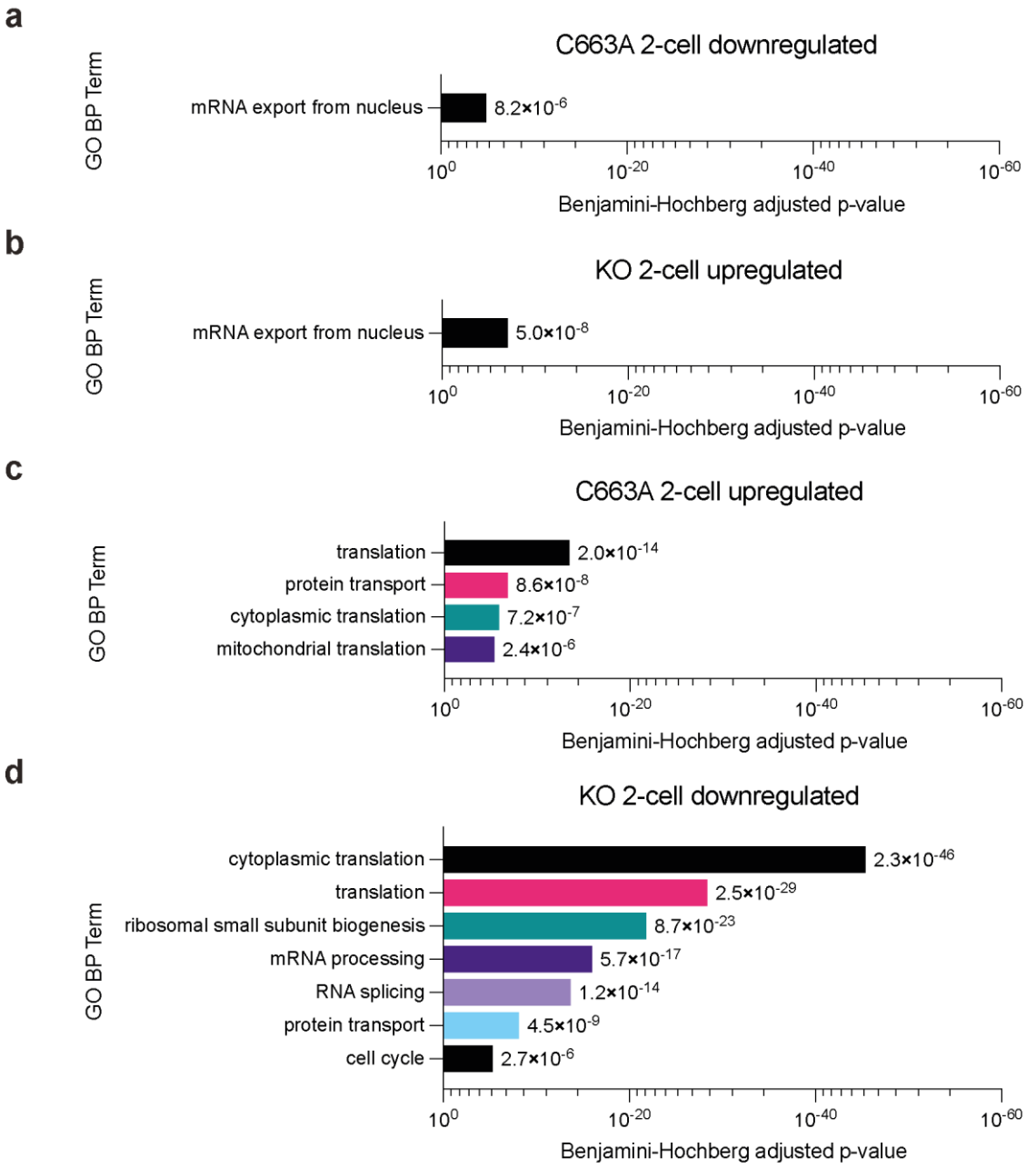


Figure 5.14 GO analysis of up and downregulated genes in 2-cell blastomeres
a)-d) GO analysis most enriched biological process (BP) terms for mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ 2-cell blastomere upregulated and downregulated genes. GO analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (246). p values were adjusted with the Benjamini-Hochberg procedure and only BP terms with an adjusted p value of lower than 1 x10^{-5} were selected. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

5.5.5 Investigation of EGA impairment

As discussed previously, it is predicted that PADI6 has a critical function in EGA. EGA is comprised of a minor wave of transcription at the late 1-cell stage, followed by a major wave of transcription at the 2-cell stage (113). To better

understand the impact that a lack of mPADI6 function or protein has on EGA, it is of interest to look at the transcript levels of minor and major EGA genes at the zygote and 2-cell stage respectively.

5.5.5.1 Identification of minor and major EGA genes

First minor and major EGA genes were identified. Genes that are upregulated 5-fold at the WT zygote stage compared with the MII oocyte, with an FDR of 0.05, were classed as minor EGA genes. Genes that are upregulated 5-fold at the 2-cell stage compared to the zygote stage, with an FDR of 0.05 were classed as major EGA genes. This resulted in the identification of 61 potential minor EGA genes, and 1192 potential major EGA genes (**Figure 5.15**).

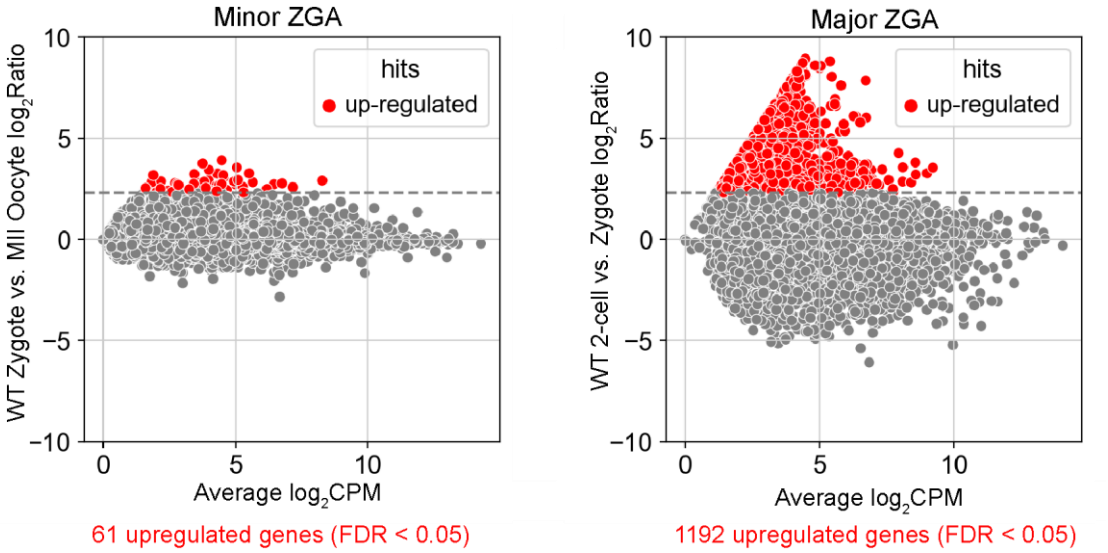


Figure 5.15 Identification of minor and major EGA genes
Left Panel: MA plot of [WT Zygote/WT MII oocyte] log_2Ratio vs. average log_2CPM of genes in WT MII oocytes and zygotes. Right panel: MA plot of [WT 2-cell/WT Zygote] log_2Ratio vs. average log_2CPM of genes in WT zygotes and 2-cell blastomeres. Genes were selected that were more than 5-fold upregulated (red) versus the preceding developmental stage with an FDR of 0.05.

5.5.5.2 Transcription of early zygotic genes in Padi6 transgenic embryos

The minor and major EGA genes determined in **Chapter 5.5.5.1** were then clustered based on their log fold-change in mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ compared to WT at the zygote and 2-cell embryo stage respectively. At the zygote stage, there were subtle differences in minor EGA genes between mat-*Padi6*$^{C663A/C663A}$ or mat-*Padi6*$^{-/-}$ and WT (**Figure 5.16a**). In clusters 1 and 2, transcript levels were

comparable between mat-*Padi6*$^{C663A/C663A}$ and WT but appear slightly downregulated in mat-*Padi6*$^{-/-}$ zygotes. Expression levels in gene cluster 3 were comparable in both mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ to WT. In cluster 4 and 5 however, expression was downregulated or comparable in mat-*Padi6*$^{-/-}$ but contrastingly slightly upregulated in mat-*Padi6*$^{C663A/C663A}$.

For major EGA genes at the 2-cell stage, clear differences were observed (**Figure 5.16b**). In mat-*Padi6*$^{C663A/C663A}$ 2-cell blastomeres, the majority of major EGA genes were upregulated compared to WT (cluster 1 and 2), with a moderate proportion comparable (cluster 3) and a much smaller proportion downregulated (cluster 4 and 5). Contrastingly, in mat-*Padi6*$^{-/-}$ 2-cell blastomeres, majority of the genes were downregulated compared to the WT (cluster 1, 2 and 3), with a much smaller proportion comparable (cluster 4) or upregulated (cluster 5). In the case of the mat-*Padi6*$^{-/-}$ 2-cell embryos, these data suggests that EGA is indeed impaired in the absence of mPADI6, validating the previous prediction of a critical function of PADI6 in EGA. The upregulation in mat-*Padi6*$^{C663A/C663A}$ 2-cell embryos mimics observations in **Chapter 5.5.4.3** that mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ embryos have an almost opposite transcriptomic phenotype at the 2-cell stage. This is an intriguing observation, and a biological explanation is not clear and requires further interrogation and analysis.

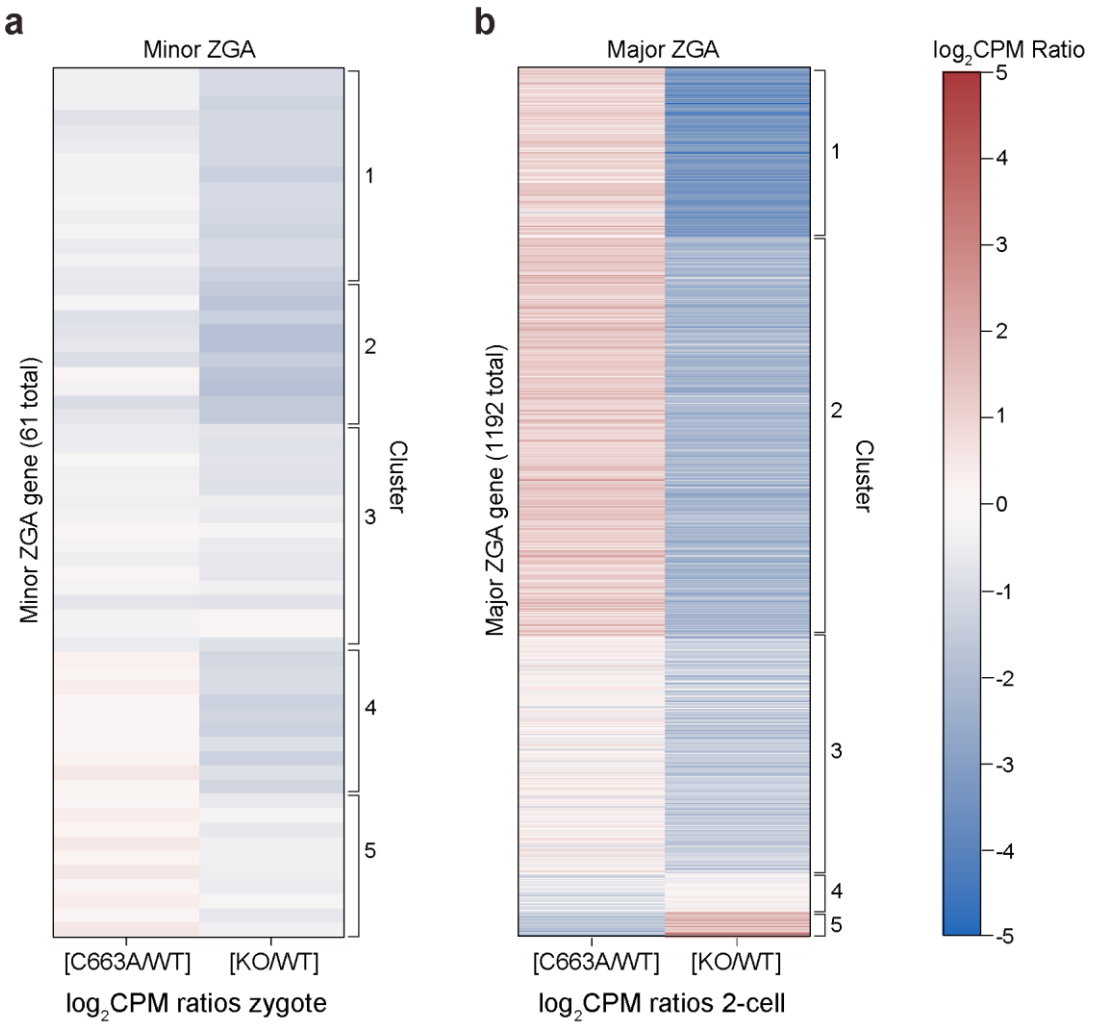


Figure 5.16 Clustering of minor and major EGA gene levels transgenic embryos
a) Cluster map of log_2CPM ratios [mat-*Padi6*$^{C663A/C663A}$/WT] and [mat-*Padi6*$^{-/-}$/WT] in zygotes of the 61 minor EGA genes identified in **Chapter 5.5.5.1**. Data was clustered into 5 groups using Python Seaborn package and ordered by cluster. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$. **b)** Cluster map of log_2CPM ratios [mat-*Padi6*$^{C663A/C663A}$/WT] and [mat-*Padi6*$^{-/-}$/WT] in 2-cell blastomeres of the 1192 major EGA genes identified in **Chapter 5.5.5.1**. Data was clustered into 5 groups using Python Seaborn package and ordered by cluster. C663A = mat-*Padi6*$^{C663A/C663A}$, KO = mat-*Padi6*$^{-/-}$.

5.6 Establishment of single embryo proteomics platform

The proteins in a cell are responsible for almost all biological processes, and the proteome is closer to observed phenotypes than the transcriptome (**Figure 5.17**). As such, when characterising cellular phenotypes, a cells proteome is most often a more valuable model than the transcriptome.

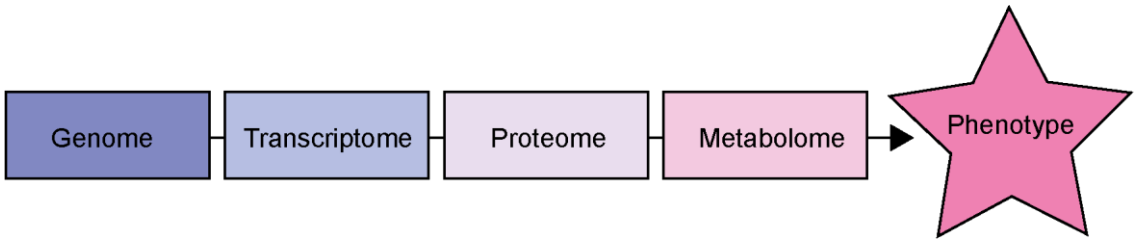


Figure 5.17 Omics organisation
Omics organisation from genome to transcriptome to proteome to metabolome to phenotype.

Whilst RNA-sequencing can give an indication of a protein's level in a cell, the correlation between mRNA and protein is not always linear (275). Furthermore, the transcriptome gives no information on the translation rate, stability, and post-translational modification state of a protein. In the case of early embryos, which are subjected to massive transcription and mRNA degradation programs during EGA, it is highly likely that there is little correlation between mRNA and protein levels. This is exemplified by data in 2017 work by Gao *et al.* where tandem-mass-tag (TMT) proteomic analysis of embryos from different stages shows that PADI6 protein level remains steady from the zygote to morula stage before falling (61). In contrast, *Padi6* mRNA level rapidly decreases between the zygote and 2-cell stage before remaining constant at a minimum.

Proteomics, however, commonly requires much more input material than RNA-sequencing as proteins cannot be amplified, in contrast to mRNA that can be amplified from tiny input quantities by polymerase chain reaction. This can be a challenge when working with limited cell numbers as is the case for oocytes and embryos. When reducing input quantity, instrument sensitivity and sample preparation become limiting factors, with large amounts of input protein being lost during transfer steps during sample preparation. This obstacle was circumvented by Gao *et al.* who employed a bulk proteomic approach using thousands of embryos to generate their dataset. An approach such as this is unfeasible for this work, and in the majority of cases, rendering the proteomic study of oocytes and early embryos highly inaccessible.

An alternative approach when cell number is limiting is to apply single-cell proteomics workflows. Single-cell or low input proteomics is a growing field of research, although one that is still in its infancy and that is limited to research groups who have access to the correct instrumentation and expertise. It is therefore far less accessible as a technique than scRNA-seq which can be done quickly and easily using a wide variety of commercial all-in-one kits. Performing single cell proteomics at The Francis Crick Institute was not a possibility until the latter stages of this work, however the opportunity arose in the latter stages to collaborate with Tania Auchynnikava of the Crick Proteomics STP to set up a single embryo proteomics platform.

5.6.1 Single embryo proteomics workflow

When this project was initiated, no reliable low input proteomic workflow existed and hence some initial tests were done to establish suitable workflow. Single cell proteomics workflows require special consideration as excessive sample handling, such as pipetting, can lead to significant sample loss. It is estimated that there is between 10 and 30 ng of protein in a mouse oocyte, which theoretically makes in-solution or S-trap digestion feasible (276). In-solution digestion involves a minimal number of processing steps during sample preparation, which reduces the potential for sample loss. The S-trap protocol, on the other hand, leads to complete sample solubilisation and despite several sample processing steps could be viable for larger cells, such as oocytes. The S-Trap protocol involves solubilisation and denaturation of lysate proteins using sodium dodecyl sulphate (SDS) and low pH, followed by binding to a derivatized silica S-Trap™ column (**Figure 5.18**). After washing away contaminants, addition of protease then allows digestion of peptides on resin, with peptides naturally eluting off the resin. The eluted peptides can then be loaded onto Evotips for subsequent liquid chromatography (LC) and tandem mass spectrometry (MS/MS) analysis.

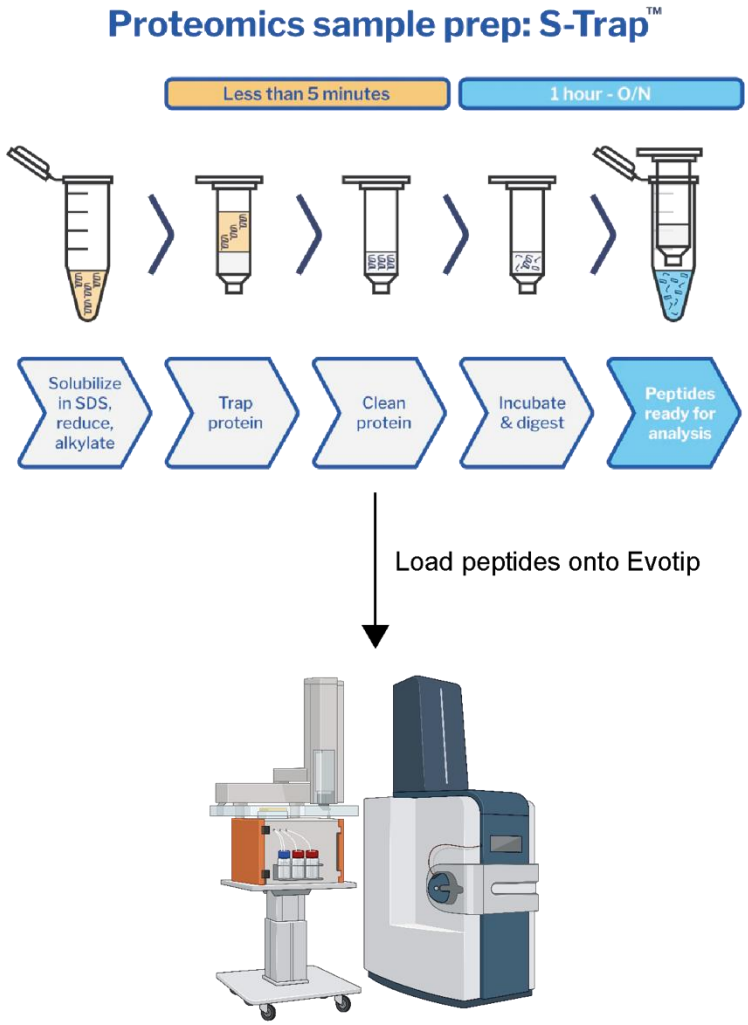


Figure 5.18 Initial single oocyte proteomics workflow
Lysate proteins are solubilized and denatured using SDS and low pH before binding to S-Trap™ column and subsequent washing steps. Protease is then added to digest peptides on resin with peptides naturally eluting off to be loaded onto Evotips. Evotip bound peptides are then LC separated using an Evosep+ 40SPD Whisper method followed by MS/MS analysis using a TIMS TOF Bruker Pro 2 MS/MS system.

Initially, I harvested a large pool of WT GV oocytes for testing and a modified S-trap protocol was trialled by Tania Auchynnikava. Samples consisting of one, two, five or ten oocytes were trialled separately (**Figure 5.19**). This initial workflow showed good potential, with 2485 and 2720 proteins identified from single GV oocytes. Increasing the number of oocytes per sample only resulted in small improvements in total proteins identified, with ten GV oocytes identifying 3150 total proteins.

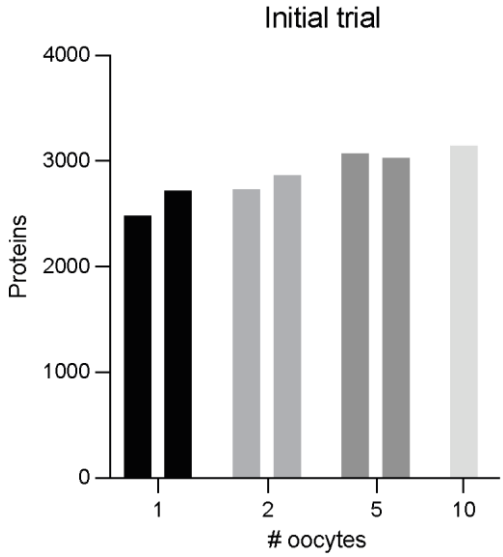


Figure 5.19 Initial WT GV oocyte proteomics trial
Total proteins identified using the initial single oocyte proteomics workflow. Two replicates of one, two, and five WT GV oocytes, and a single replicate of ten WT GV oocytes were trialled.

5.6.2 GV oocyte differential protein analysis

As the initial tests on WT GV oocytes were successful in identifying an average of 2500 proteins, the platform was tested to determine whether it was capable of identifying differentially regulated proteins between WT and *Padi6*$^{C663A/C663A}$ or *Padi6*$^{-/-}$ GV oocytes. 15 GV oocytes were harvested from homozygous females of each line, totalling 45 oocytes. These oocytes were then handed to Tania Auchynnikava for sample processing and data acquisition, before being returned to myself for subsequent analysis.

The number of proteins identified in each oocyte was highly variable, suggesting the platform requires further optimisation. In out of the 45 oocytes analysed, only 25 samples had more than 1000 proteins identified, these sample were kept for the future comparisons. This initial QC filtering resulted in dataset consisting of 10 WT, 7 *Padi6*$^{C663A/C663A}$, and 8 *Padi6*$^{-/-}$ oocytes, with the number of proteins identified ranging from 1163 to 3311 (**Figure 5.20a**).

Proteins that were present in GV oocytes from *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females at different levels to oocytes from WT females were then identified. These were selected as those with a signal intensity fold-change [*Padi6*^{mutant}/WT] greater than 2 and a p-value less than 0.01 were classed as upregulated, and with a fold-change less than 0.5 and a p-value less than 0.01 downregulated (**Figure 5.20b, c**). 15

proteins were up-regulated, and 65 downregulated in GV oocytes from *Padi6*$^{C663A/C663A}$ females versus those from WT females. On the other hand, 19 proteins were upregulated, and 35 downregulated in GV oocytes from *Padi6*$^{-/-}$ females versus those from WT females. No proteins were commonly up or downregulated between *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ GV oocytes versus WT. Pleasingly, UHRF1 was identified as downregulated almost 2-fold in *Padi6*$^{-/-}$ GV oocytes relative to WT. Whilst this downregulation is not as dramatic as the approximate 90% downregulation reported by Schuh and co-workers in *Padi6*$^{-/-}$ GV oocytes, it shows that this initial platform is capable of identifying up and downregulated proteins (67). Furthermore, whilst fewer proteins were identified as up or downregulated compared to the reported work, it must be noted that the stringencies of significant up or downregulation implemented were different.

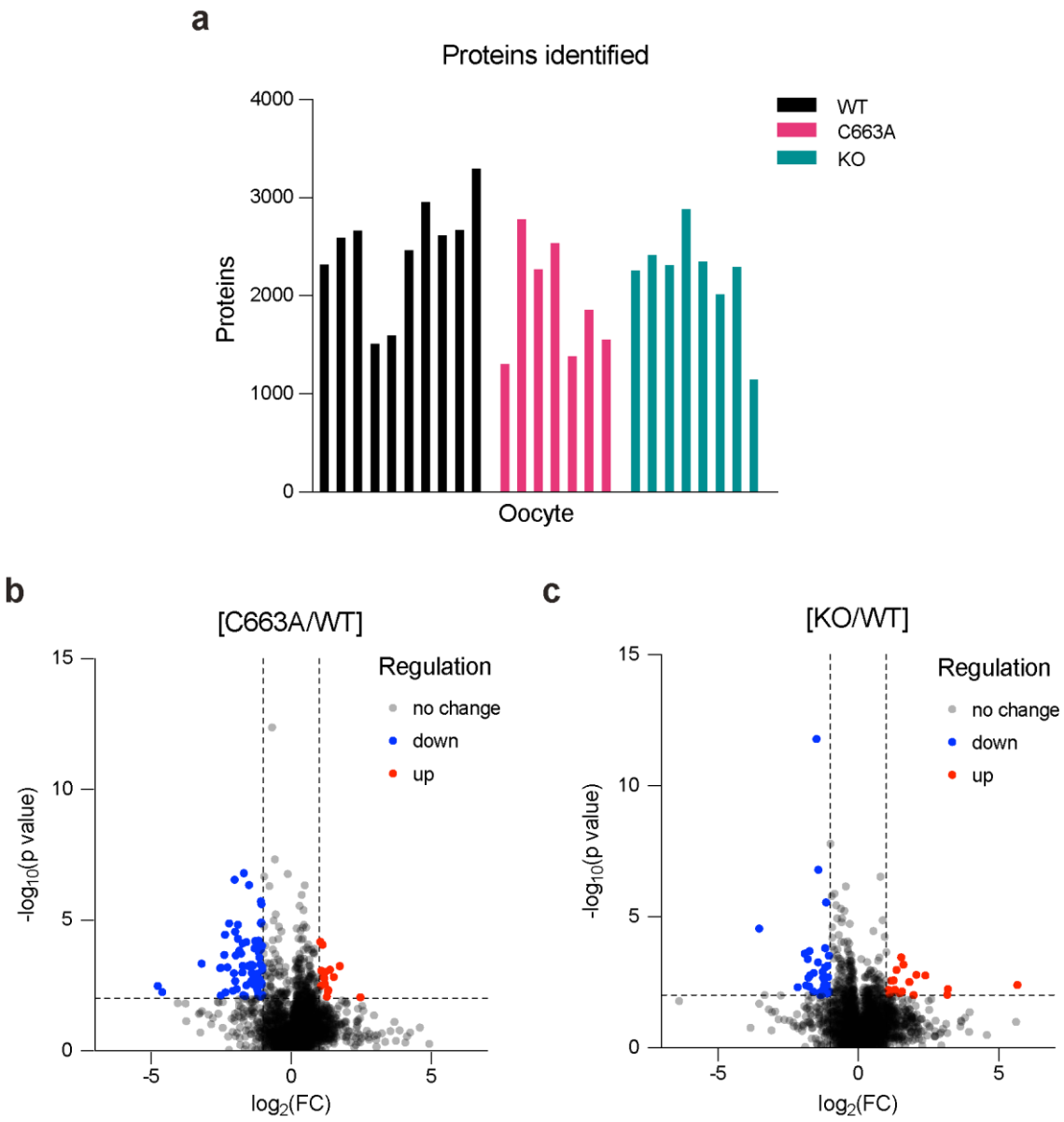


Figure 5.20 Initial workflow comparative proteomics of *Padi6*^{mutant} GV oocytes
a) Total proteins identified from each GV oocyte after filtering oocytes that captured fewer than 1000 proteins. C663A = *Padi6*$^{C663A/C663A}$, KO = *Padi6*$^{-/-}$. **b)** Volcano plot of [*Padi6*$^{C663A/C663A}$/WT] log_2(fold-change) and -log_{10}(p-value). Proteins were selected that were more than 2-fold upregulated (red) or downregulated (blue) versus WT with a p-value < 0.01. C663A = *Padi6*$^{C663A/C663A}$. **c)** Same as **b)** but KO instead of C663A. KO = mat-*Padi6*$^{-/-}$.

5.6.3 One-Tip adapted workflow

Due to the inconsistency in total proteins identified between samples in using the original platform, it was decided that the sample preparation could be optimised further. Since it became clear that lysing cells in SDS was beneficial for the sample preparation we explored using other detergents in our workflow. Many commonly used detergents cannot be used for mass spectrometry sample preparations due to presence of polyethylene glycol which results in contamination of the spectra across the LC gradient. N-Dodecyl β-D-maltoside (DDM) was proposed as a possible alternative, as it elutes as a single peak during data acquisition. Several labs have reported the use of DDM recently in low input proteomic workflows, including the One-Tip protocol (247). The One-Tip protocol involves performing cell lysis and digestion in solution and directly loading onto an Evotip for subsequent LC MS/MS analysis. Crucially, this eliminates the S-Trap processing steps, reducing the amount of sample lost during sample preparation. The reported work, however, utilises an MS/MS instrument with the latest generation of mass analysers, affording greater sensitivity and speed. Such an instrument was not currently available at the Crick at the time of this work. As such, the reported One-Tip protocol was modified to reduce potential sample loss during preparation and improve sample acidification and by extension binding to the Evotip C18 matrix. Crucial changes include:

- Lysis was promoted by brief sonication
- 5 μL lysates were not diluted further before loading onto the Evotip
- Higher quantity of Trypsin used during digestion
- Evotips incubated for 10 mins in a bath of 20% formic acid before addition of minimal formic acid to acidify the peptides without diluting the sample further

The adapted One-Tip protocol, with these changes implemented resulted in dramatic improvements to the workflow in both peptide material recovery and the consistency of sample preparation (**Figure 5.21a**). Out of 9 oocytes analysed using the modified One-Tip method all cells consistently had more than 5150 proteins identified with average 5332 proteins per sample. Furthermore, the proteins identified, and their intensities were highly replicable between oocytes (**Figure 5.21b**). In-line with previous data that PADI6 is one of the most abundant proteins in the oocyte and early embryo (61), it was on average the fourth most abundant protein identified in our dataset. Interestingly, SCMC proteins KHDC3 and TLE6 were the third and fifth most abundant

proteins across the 9 GV oocytes analysed with NALP5 and OOEP were 17th and 90th respectively. I am now working to harvest not only GV oocytes, but also MII oocytes, zygotes and 2-cell embryos to analyse using this optimised workflow.

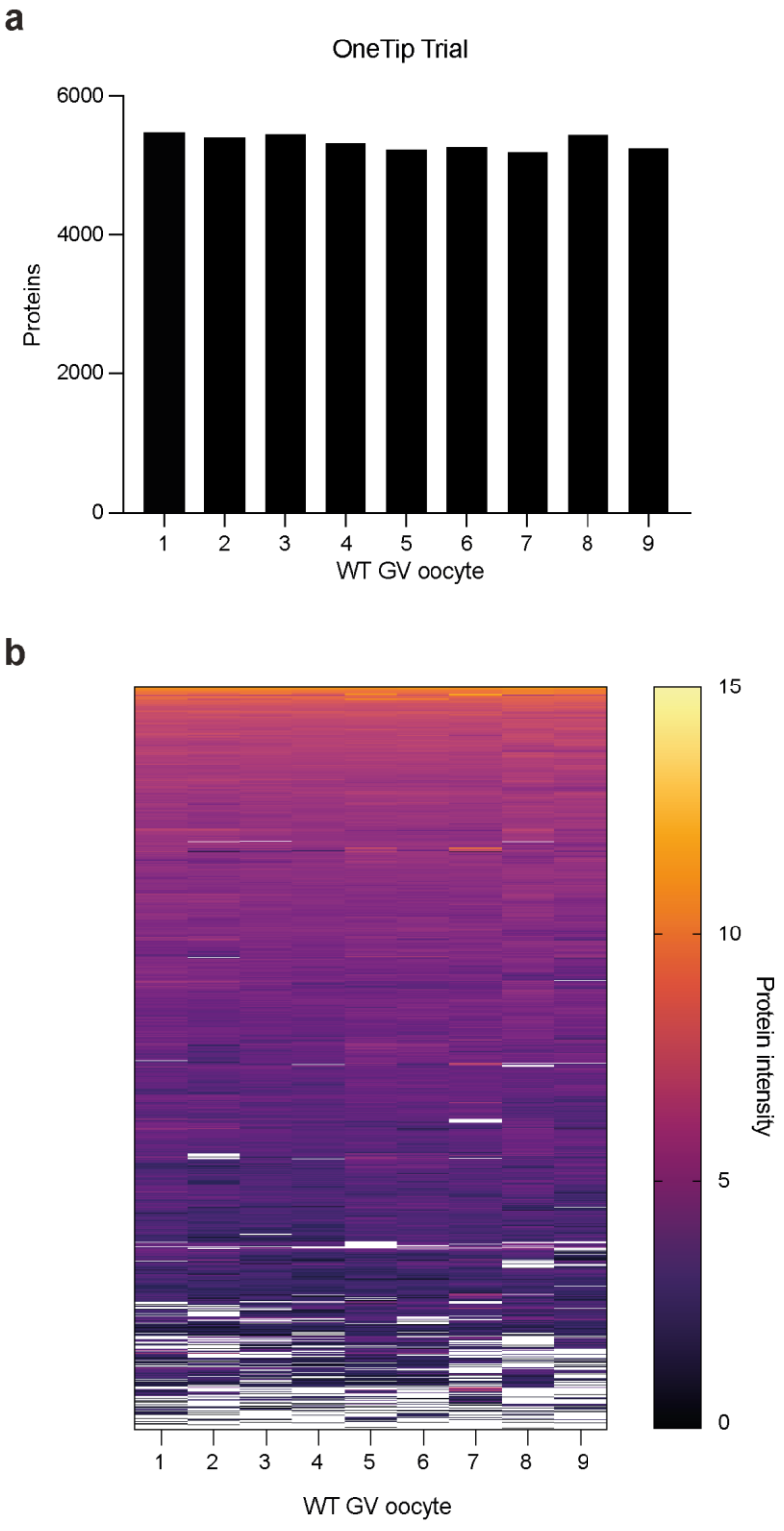


Figure 5.21 Adapted One-Tip method WT GV oocyte trial
a) Total proteins identified from nine single WT GV oocytes using the adapted One-Tip method. **b)** Heatmap of protein intensities identified from nine single WT GV oocytes using the adapted One-Tip method.

5.7 Summary

To better understand the function of PADI6 in the early embryo, as well as gather more data on the possibility of PADI6 catalytic activity, we used two transgenic mouse models. One, a *Padi6* knock-out line, the other a single cysteine mutant predicted to abolish PADI6 catalytic activity, if it has any. I then characterised the developmental phenotypes of these mutant lines using primarily time-lapse imaging, and scRNA-seq. In collaboration with the EM facility at the Crick, I additionally confirmed that the mPADI6 C663A substitution has no effect on CPL formation.

Crucially, catalytic mutant PADI6 females were fertile, suggesting that PADI6 catalytic activity is not required for embryo development. Interestingly however, fewer mat-*Padi6*$^{C663A/C663A}$ zygotes successfully reached the blastocyst stage than WT zygotes, with a larger subset arresting or degenerating at the 4-cell and morula stages. Consistent with this, mat-*Padi6*$^{C663A/C663A}$ embryos were delayed in reaching the 4-cell, 8-cell and morula stages compared to WT embryos. This delay comes after EGA and suggests a possible catalytic involvement of PADI6, although its absence is not wholly detrimental to the embryo's development.

The developmental phenotypes in mat-*Padi6*$^{-/-}$ embryos on the other hand were much more profound. Consistent with the suggested function of PADI6 in EGA, the majority of mat-*Padi6*$^{-/-}$ embryos arrested or degenerated at the 2-cell and 4-cell stages, with significant delays in reaching both stages. Interestingly, a very small subset of mat-*Padi6*$^{-/-}$ zygotes went on to form blastocyst like embryos. These embryos appeared highly abnormal, however, and as shown by the previous fertility study, fail to develop further. Together these data suggests that PADI6 primarily functions at the 2-cell stage, the time of major EGA in mice. Regarding the contrasting developmental arrest times between the literature, and the *Padi6* knock-out mice in this experiment, it is possible that the previously reported developmental phenotypes were accentuated by superovulation. In this work, all mice were naturally mated over a period of 4 days. This did reduce the number of embryos gathered; however, it should have eliminated

any artificial effects of superovulation. Therefore, we propose that embryos *Padi6* knock-out female mice do not have a hard 2-cell block but are majorly inhibited in the development at the 2-cell stage, and ultimately this results in female infertility.

To further characterise the developmental phenotypes of mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ oocytes and embryos, I performed scRNA-seq on MII oocytes, zygotes and 2-cell embryos from females of each line, in collaboration with the Advanced Sequencing Facility at the Crick. Few transcriptomic differences were observed in mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ MII oocytes, and similarly in zygotes when compared to WT cells. On the other hand, at the 2-cell stage, the time of major EGA in mice, dramatic up and downregulation of hundreds of genes was observed for both mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ 2-cell embryos compared to WT. Most interestingly, over a quarter of the genes that were downregulated in mat-*Padi6*$^{-/-}$ 2-cell embryos, were upregulated in mat-*Padi6*$^{C663A/C663A}$ 2-cell embryos. GO analysis of these genes showed strong enrichment for genes involved in translation, both cytoplasmic and mitochondrial, protein transport, and mRNA processing. This opposite effect is intriguing, and a biological explanation is not clear at this stage.

Identification of minor and major EGA genes based on the WT embryo data showed that the majority of these genes were downregulated in mat-*Padi6*$^{-/-}$ 2-cell embryos indicating that EGA is indeed defective when *Padi6* is knocked out. Contrastingly, the majority of major EGA genes were upregulated in mat-*Padi6*$^{C663A/C663A}$ 2-cell embryos. This aligns with the opposite transcriptomic observed between mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ 2-cell embryos when looking at all differentially expressed genes when compared with WT embryos. The exact mechanism of PADI6's function in EGA is unknown, and this is a question I hope to answer using the single embryo proteomics workflow developed in this work. Proteomic data will also provide clarity on the opposite transcriptomic phenotypes observed in mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ 2-cell embryos. A potential function of PADI6 in EGA is discussed further in **Chapter 7.2**.

Towards this, in collaboration with the Proteomics STP at the Crick, we have established a single embryo proteomics workflow that can consistently identify upwards of 5000 proteins from single GV oocytes. Furthermore, even in using the initial non-optimised proteomic workflow, we were able to replicate reported data showing UHRF1

was downregulated in mat-*Padi6*$^{-/-}$ GV oocytes (67). It must also be noted that the reported work that identified UHRF1 as downregulated in *Padi6*$^{-/-}$ GV oocytes utilised 120 GV oocytes per sample, compared to the one GV oocyte per sample in our own. The next step in this work is to apply the optimised pipeline to not only GV oocytes, but MII oocytes, zygotes and 2-cell embryos from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females, and this work is ongoing.

Together, this work provides further evidence for a function of PADI6 in EGA. Developmental defects were observed in embryos from our catalytic mutant females; however, these females were fertile, suggesting that any catalytic function is likely non-essential. The question of whether PADI6 has an additional non-essential catalytic function is discussed further in **Chapter 7.2**. Finally, we have established a single oocyte or embryo proteomics pipeline, which can replicate published data performed on large pools of oocytes and has opened the door for us to perform comprehensive proteomic profiling of oocytes and embryos from *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females.

Chapter 6. PADI6 and maternal imprinted signature establishment in mice

6.1 Introduction

Discussed in **Chapter 1.1.4**, it has been suggested that PADI6 is a *trans*-acting factor in maternal imprinting. *PADI6* variants in nine mothers have been attributed as the cause of MLID in their children (20). However, our understanding of the molecular mechanisms behind PADI6 MLID, and imprinting disorders in general, is limited. The most common way to investigate imprinting defects is through characterisation of the methylation state of imprinting centres (159). In the case of PADI6 and SCMC attributed MLID, characterisation is commonly done by methylome sequencing of peripheral blood lymphocytes of patients suffering from the disorders (159). If the patient suffers from an imprinting disorder or MLID, dramatic methylation defects (most commonly loss of methylation) are often observed at imprinting centres. These defects are presumed to originally occur in the oocyte or early embryo, caused by either the improper establishment of the imprinted signature in the oocyte, or the faulty maintenance of the imprinted signature through early embryo development.

The establishment and maintenance of the imprinted methylation signature are mediated by different cellular pathways, however, and are both complete by full-term pregnancy. As such, sequencing the methylomes of children fails to tease apart the true cause of the methylation defects. The study of the causes of these imprinting disorders in humans is therefore challenging as it requires methylation sequencing of oocytes and early embryos from donors carrying relevant variants. Such an experiment has not been reported for PADI6; however, it has in the case of the SCMC protein KHDC3L (see **Chapter 1.3.4**) (231). In oocytes from a patient homozygous for a loss-of-function *KHDC3L* variant, hypomethylation was observed across the genome, suggesting loss of KHDC3L function resulted in global defects in the establishment of the oocyte methylome. Methylation sequencing of a single blastocyst-like embryo from the patient displayed further losses in methylation, suggesting an additional function of KHDC3L in methylation maintenance. Given the theorised tight functional link between PADI6 and the SCMC, it is reasonable to suggest that PADI6 is also involved in the establishment of the maternal methylome, and its subsequent maintenance during early embryo development. Whilst it is clear that certain *PADI6* gene variants can lead

to methylation defects and by extension, imprinting disorders, there is limited understanding of the molecular mechanisms surrounding this phenotype.

The mouse genome also contains imprinting centres and has previously been used as a model for studying maternal methylome establishment in the oocyte, and methylome maintenance through early embryo development (167,168,179,277). Given the conservation of PADI6 and the SCMC between humans and mice (80), it is conceivable that their role in maternal imprinting is also conserved.

As discussed above, when oocytes from a female carrying a *KHDC3L* variant were subject to BS-Seq, dramatic methylation defects were observed suggesting that the SCMC and potentially also PADI6 are involved primarily in the establishment of the maternal methylome. As the maternal methylome is established at the GV oocyte stage, then any defects in establishment would be observable at this point.

In this work, I aimed to understand the function of PADI6 in the establishment of the maternal imprinted signature using a mouse model. I chose GV oocytes as the stage to investigate as it would give a clear answer to whether PADI6 is involved in the establishment of the maternal methylation signature in mice. The knock-out model previously described (see **Chapter 5.2**) will allow the interrogation of whether absence of PADI6 in the maturing oocyte results in defects in the maternal imprinted methylation signature in the GV oocyte. Furthermore, the catalytic mutant model could provide insight into whether PADI6 catalytic activity is required.

Studying the potential role of PADI6 in the maintenance of the maternal methylome is more difficult. If the defect results in the maternally methylated residues not being maintained through early embryo development, then successive rounds of cleavage would result in the original methylated residues becoming more and more diluted and the phenotype becoming clearer. At the 2-cell stage, four copies of each maternal gDMR will be present: two paternal, two maternal. The paternal copies would be expected to have a methylation level of 0% consistent with the maternally imprinted gene being expressed from the paternal genome. In a healthy 2-cell embryo, both maternal copies would be highly methylated. If there is a defect in the maintenance of the maternal imprinted signature however, it would be expected that this methylation level would be reduced by half consistent with the replicated genome not being

methylated during mitosis. With high CpG coverage, this decrease would be observable. Unfortunately, it was decided at the time that it was unfeasible to gather enough 2-cell embryos to acquire a dataset with sufficient coverage to confidently probe this decrease in methylation. Especially as many *Padi6* knock-out embryos arrest at the 2-cell stage. Consequently, only the involvement of PADI6 in the establishment of the maternal methylome was investigated.

6.2 Low-input bisulphite sequencing protocol establishment

6.2.1 Experimental design

Historically, the most commonly used technique to sequence the methylome is bisulphite sequencing (BS-seq). All BS-Seq techniques centre on the capacity of bisulphite to specifically convert cytosine to uracil, while methylated cytosine is protected from conversion (**Figure 6.1a**). Treatment of DNA with bisulphite (commonly sodium bisulphite) results in DNA retaining only the methylated cytosines with all unmethylated cytosines being converted to uracil (**Figure 6.1b**). Different methods can then be used to characterise the conversion. Most commonly however the bisulphite treated DNA is amplified and sequenced by NGS. Comparison of the sequencing data to a reference genome allows the identification of methylated cytosine residues, as they will be read as cytosine during sequencing and match the reference genome. Unmethylated and converted cytosines are further converted to thymine during PCR amplification of NGS libraries and therefore do not match the reference genome.

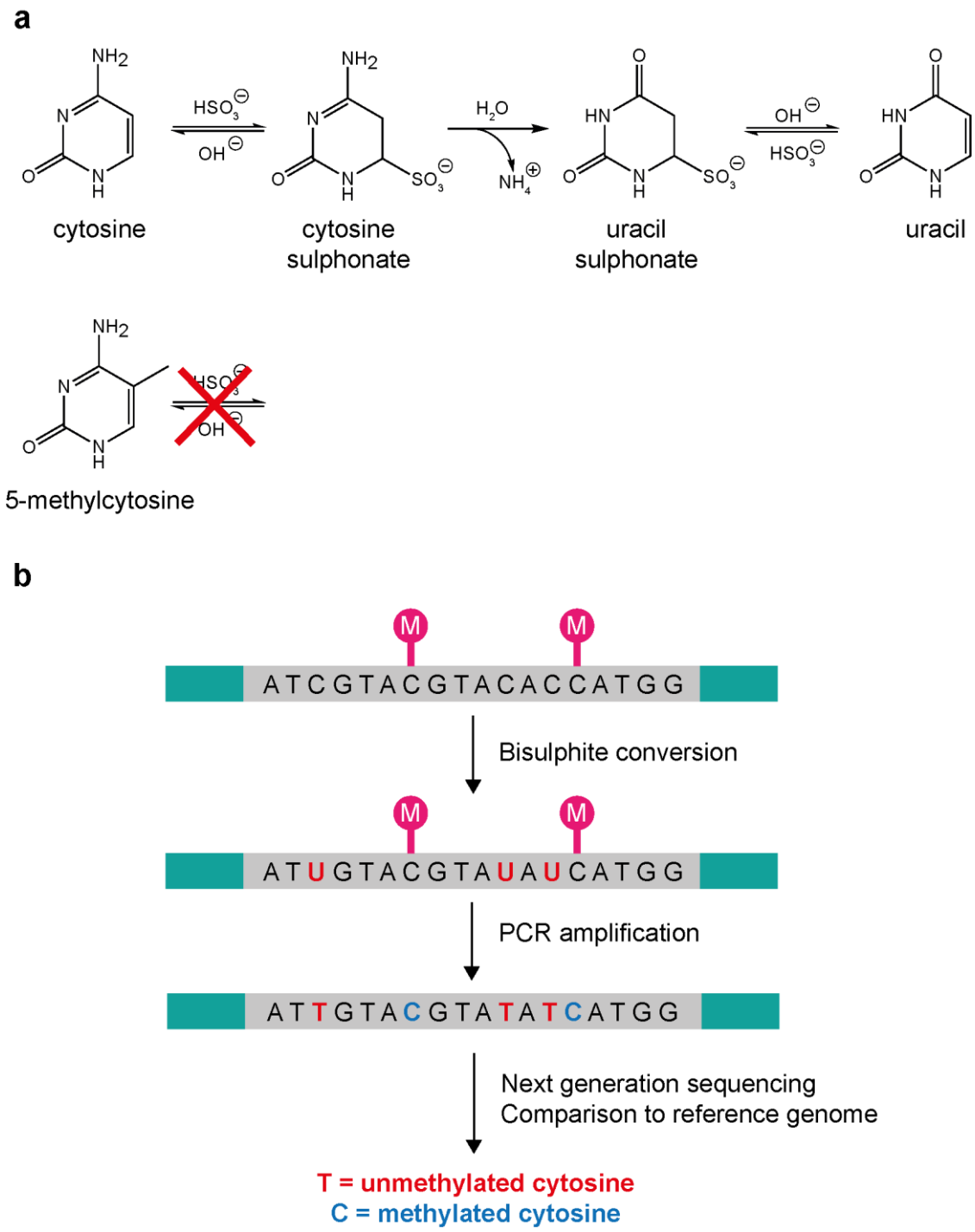


Figure 6.1 Bisulphite sequencing overview
a) Reaction scheme of bisulphite with cytosine and methylated cytosine. **b)** Schematic of bisulphite-sequencing workflow.

Classically, in whole genome bisulphite sequencing (WGBS), DNA is extracted, fragmented, the ends repaired, and adapters added before bisulphite conversion and amplification. The major drawback to this method is that it requires large starting amounts of DNA. Bisulphite treatment is inherently harsh; the conditions required for

full conversion are often high temperatures and long incubation times which can result in degradation of up to 90% of the DNA (278). This sample loss during library preparation can lead to poor depth of coverage and underrepresentation of methylation sites in the genome unless sufficient input quantities are used. Due to the high quantities of DNA required, the study of oocytes and embryos by classical BS-Seq is challenging as it requires far too many oocytes or embryos to acquire an adequate dataset.

Significant advances to the BS-Seq procedure have allowed a reduced amount of input material, however. In standard bisulphite sequencing, DNA is fragmented and sequencing adaptors are ligated to DNA before bisulphite treatment (279). The harshness of the treatment can then result in a significant loss of information, in either indexing or genomic sequences, due to DNA degradation. In low input BS-seq approaches, DNA is bisulphite treated and fragmented first in the same step (279–281). The bisulphite treated and fragmented DNA is then subjected to three rounds of PCR. The crucial first PCR round involves random priming to amplify the library, followed by adapterization second and finally the addition of indexes to the library. The resulting libraries can then be immediately sequenced by NGS.

Zymo Research provides the all-in-one Pico Methyl-Seq Library Prep Kit, which implements the above-described low input workflow and caters to input DNA quantities as low as 10 pg. I selected this method to apply to PADI6 samples to as the validated all-in-one nature of the kit would reduce the need for, and time required, to optimise an in-house low input bisulphite sequencing approach.

Regarding sample sizes, I decided on four replicates of 20 oocytes per mouse strain as 20 oocytes should bring the DNA quantity well above the lower limit of 10 pg, whilst still being feasible to harvest four replicates from each line (240 total oocytes). The Zymo Pico Methyl-Seq Library Prep Kit only allows for multiplexing of up to 6 samples, thus alternative sequencing primers were designed based on IDT for Illumina-TruSeq UDIs to allow multiplexing of up to 13. The sequencing primers used are detailed in **Appendix 13**.

6.2.2 Sample harvesting

Over the course of three sessions, GV oocytes were harvested from two three-week-old mice from each genetic strain. I conducted the harvesting with the aid of Sugako Ogushi, formerly of the Turner lab at the Crick to increase the speed of harvesting so many oocytes at once. Three-week-old mice were chosen as they possess significantly higher numbers of GV oocytes than the 8–10-week-old females. As the maternal methylation signature is fully established in GV oocytes, the age of the females should not have any effect. Oocytes from each pair of mice were pooled giving a total of 160 WT oocytes, 100 *Padi6*$^{C663A/C663A}$ oocytes and 100 *Padi6*$^{-/-}$ oocytes were harvested, divided into aliquots of 20 and stored in RLT+ buffer at -80 ^{o}C to be used in initial trials and the full experiment.

6.2.3 Initial library preparation trial

An initial trial of the library preparation was completed to highlight any steps that might require optimisation, as well as confirm the in-house designed sequencing primers were working. 1 ng of HeLa DNA, a sample of 20 WT GV oocytes, and a negative control of RLT+ lysis buffer were taken through bisulphite conversion and library preparation. The DNA quantity in the final libraries was quantified (**Table 6.1**) and library quality was analysed by TapeStation (**Figure 6.3, Appendix 14**).

Table 6.1 Initial library preparation trial library DNA concentrations

Sample	DNA concentration (ng/μL)
RLT+ buffer	14.5
10 ng HeLa DNA	15.9
20 WT GV oocytes	13.1

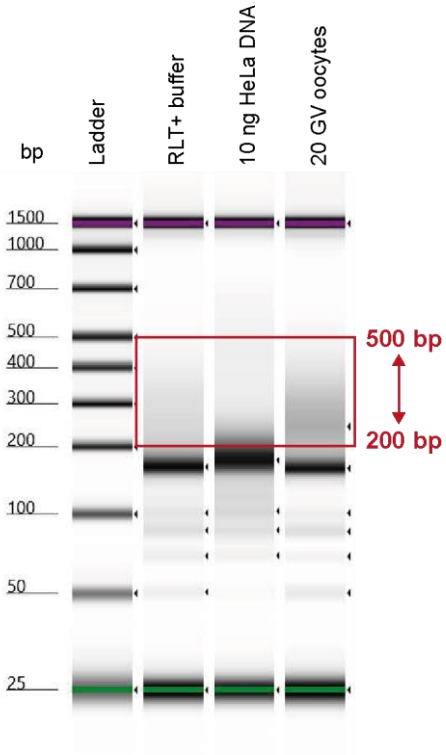


Figure 6.2 Initial library preparation trial fragment distributions
High Sensitivity DNA ScreenTape gel of libraries prepared from RLT+ buffer, 10 ng HeLa DNA, and 20 WT GV oocytes, run on 4200 TapeStation System. Red box indicates expected library fragment distribution.

The DNA quantity of all three samples ranged from 13.1 to 15.9 ng/uL, the presence of DNA in the RLT+ buffer negative control indicates either DNA contamination during library preparation, or the presence of large quantities of PCR side-products. In the TapeStation gels each library was dominated by a large peak at approximately 167 bp, overshadowing a smaller but wider peak of a shape and fragment size range that matches expected library size of 200 to 500 bp (**Figure 6.2**). Given the sharpness and consistent presence of the smaller peak in each sample, it is likely that this was a primer side-product from the PCRs. Furthermore, given the intensity of this peak, it is likely that this peak was the major DNA component in the final libraries, and likely why the negative control sample had a high DNA concentration. This intense band at 167 bp is accompanied by multiple less intense bands at lower fragment sizes, also likely to be PCR side-products. Supporting these assumptions, it is noted in the Zymo Pico Methyl-Seq Library Prep Kit manual that primer side-products can be present especially when input DNA quantities are low, and the presence of these can be alleviated by reducing the primer concentration during

the first amplification PCR (PreAMP Primer). Optimisation of this is further discussed in **Chapter 6.2.4**.

The final step of the purification in the kit involves a column clean-up of the library. However, it has been reported that bead-based clean-ups are much more efficient in the size-selective purification of DNA libraries (282). An additional way to remove primer side-products could thus be to subject the libraries to a final bead clean-up step. The expected final fragment sizes of the final libraries is between 200 bp to 500 bp and as such a ratio of 0.9x SPRIselect or two rounds of 1x SPRIselect volume: sample volume was trialled to remove PCR side-products. The final libraries were diluted, split and subjected to either no clean-up, one round of 0.9x, or two rounds of 1x SPRIselect volume: sample volume. Library fragment distributions before and after each bead clean-up condition were again analysed by TapeStation (**Figure 6.4, Appendix 15**). The large band at approximately 167 bp, and the smaller non-library bands were largely absent or reduced in intensity after both bead clean-ups suggesting these alternate bead clean-up steps had successfully removed the majority of the smaller primer side-products. The fragment distributions for the HeLa and GV oocyte samples displayed the expected library size range after either clean-up condition. Fragment sizes ranged from approximately 200 bp to 500 bp, with a near-Gaussian distribution with the highest concentration at the middle of the range. Furthermore, after bead clean-up the negative control lacked this characteristic fragment distribution suggesting the DNA quantified is dominated by the PCR side-products. This is exemplified by the bead clean-up failing to remove all the side-product fragments around 167 bp. Two rounds of 1x SPRIselect bead clean-up was more successful in removing the side-product peak and so it was decided all future library preparations would include this final clean-up step.

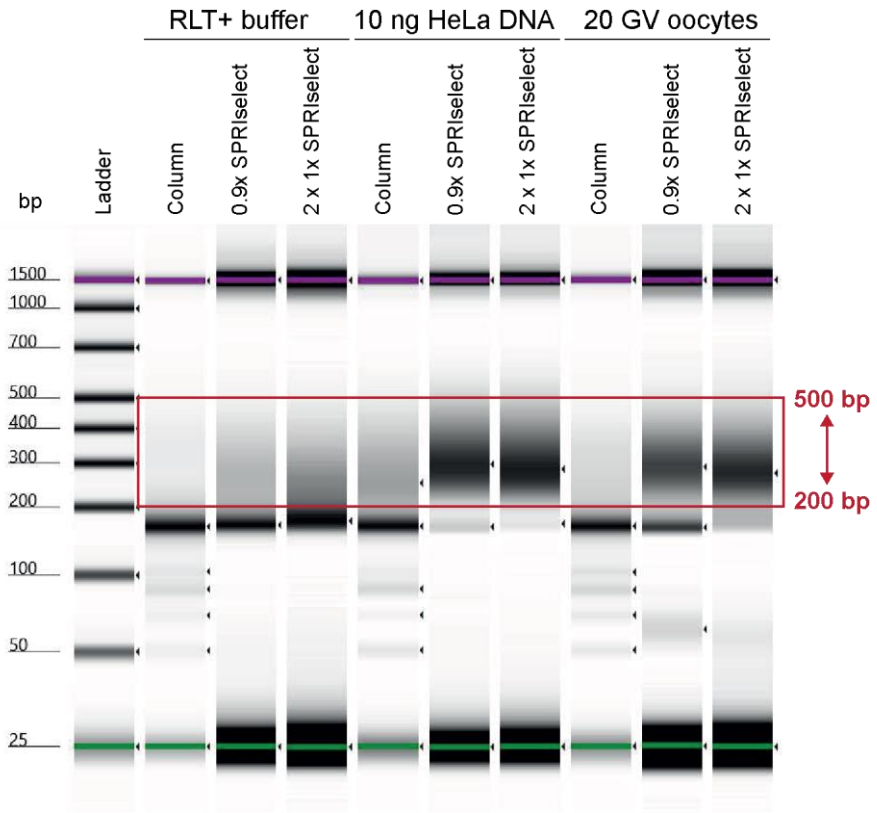


Figure 6.3 Bead clean-up optimisation fragment distributions
High Sensitivity DNA ScreenTape gel of libraries prepared from RLT+ buffer, 10 ng HeLa DNA, and 20 GV oocytes, run on 4200 TapeStation System. Libraries analysed before, and after either 0.9x or two 1x SPRIselect column clean-up rounds. Red box indicates expected library fragment distribution.

6.2.4 Further optimisation and Pilot Study 1

A follow-up library preparation trial was conducted, with equivalent starting samples to the initial trial, but reducing the PreAMP primer concentration from 40 µM to 10 µM. As discussed in the previous section, libraries were subjected to a final bead clean-up comprising two rounds of 1x SPRIselect clean-up. Final DNA library concentrations and TapeStation fragment distributions were again measured (**Table 6.2**, **Figure 6.5, Appendix 16**).

Table 6.2 Pilot Study 1 library DNA concentrations

Sample	DNA concentration (ng/µL)
RLT+ buffer	7.5
10 ng HeLa DNA	10.5
20 GV oocytes	9.4

.

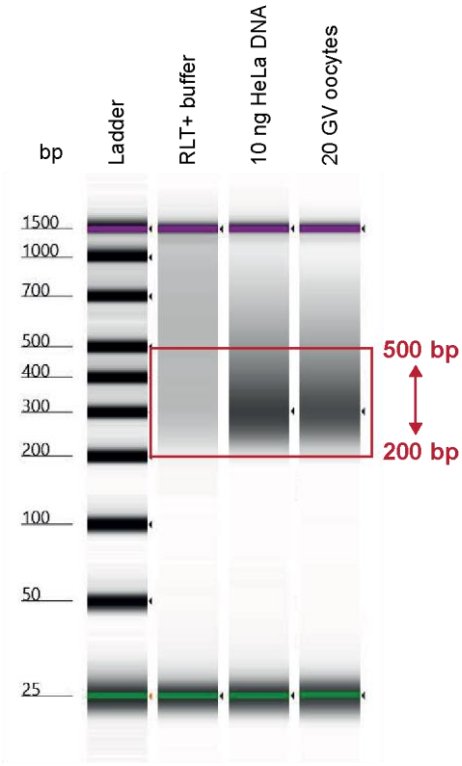


Figure 6.4 Pilot Study 1 library fragment distributions
High Sensitivity DNA ScreenTape gel of libraries prepared from RLT+ buffer, 10 ng HeLa DNA, and 20 GV oocytes, run on 4200 TapeStation System. Libraries were subjected to a final bead clean-up of two rounds of 1x SPRIselect. Red box indicates expected library fragment distribution.

Whilst the negative control RLT+ sample contained a high quantity of DNA, the library composition shown by TapeStation had no peak in the expected range, but a low intensity distribution of fragments from 200 to 1500 bp. Both the HeLa and GV oocyte libraries displayed the expected concentration and size distribution ranging from 200 to 500 bp. No peak around 167 bp was present suggesting both the reduced PreAMP primer concentration and bead clean-up had succeeded in eliminating the PCR side-products. The HeLa and GV oocyte libraries were then sequenced to 33 and 47.5 million 100 bp paired-end reads respectively by the Crick Advanced Sequencing Facility. Sequencing was successful indicating the in-house produced sequencing primers were effective. Raw sequencing reads were trimmed with TrimGalore! (248) and read 1 was mapped onto both the GRCh38 (human) and GRCm39 (mouse) reference genomes using Bismark (249). Alignments were deduplicated and extracted from Bismark before combined analysis with the R methylKit package (250). Sequencing and data processing statistics are summarised in **Table 6.3**.

Table 6.3 Pilot Study 1 sequencing and data processing statistics

	10 ng HeLa DNA	**20 WT GV oocytes**
Total sequences (million)	33	47.5
FastQC Duplicated reads (%)	53.4	65.3
GC content (%)	28	26
Mapping to human genome		
Mapping efficiency (%)	63.8	1.7
Reads removed in deduplication (%)	61.2	-
Sequences left after deduplication (million)	8.1	
CpG methylation (%)	60.4	-
CHG methylation (%)	19.5	-
CHH methylation (%)	25.1	-
Mapping to mouse genome		-
Mapping efficiency (%)	0.3	57.2
Reads removed in deduplication (%)	-	73.7
Sequences left after deduplication (million)		7.1
CpG methylation (%)	-	27.2
CHG methylation (%)	-	3.4
CHH methylation (%)	-	3.2

After sequencing, both the HeLa and GV oocyte samples had library duplication of 53.4% and 65.3% respectively. High duplication values are to be expected given the low input quantity and multiple rounds of PCR amplification. Furthermore, the higher duplication in the GV oocyte library is likely due the additional read depth of 14.5 million. The HeLa and GV oocyte had 28% and 26% GC% respectively. A lower GC percentage is to be expected given the majority of CpGs (those that are unmethylated) will have been converted to uracil and read as thymine, lowering the overall percentage of GC present in the DNA.

The HeLa methylation data mapped with 63.8% efficiency to the human genome and 0.3% to the mouse genome. On the other hand, the GV oocyte sample mapped with 57.2% efficiency to the mouse genome and 1.7% to the human genome. Both are comparable to mapping efficiencies for similar low-input experiments reported

in the literature (280,281). During deduplication of the mapped reads, 61.2% and 73.7% of each the HeLa and GV oocyte datasets were removed, leaving 8,122,042 and 7,110,036 unique mapped reads for each.

Taking the GV oocyte data, percentage methylation in CpG, CHG and CHH contexts were calculated. As expected, percentage methylation of CHG and CHH context cytosines was very low (3.4% and 3.2% respectively). Percentage methylation in the CpG context was calculated at 27.2%. Comparing these values to published data, both CHG and CHH match well, however, the CpG methylation was lower than the reported value of 37.9% (214).

To gain insight into the methylation levels at multiple genomic loci, a high coverage of CpG sites is required. This can be calculated by dividing the number of CpGs covered by reads in the data, by the total number of CpGs in the genome. Looking at the CpG coverage in the Pilot Study 1 GV oocyte data, 19.3% of all CpGs were captured at least once however only 3% of all CpGs were captured twice or more, dropping off further for three and four times captured (**Figure 6.6**). The large drop off between CpGs covered once compared to CpGs covered twice indicates that the library hasn't been sequenced to saturation. As such, before conducting the full experiment, it was decided a second pilot study would be completed with libraries prepared from oocytes from each genotype, with final sequencing to both 20 million and 200 million reads to compare any differences in coverage achieved.

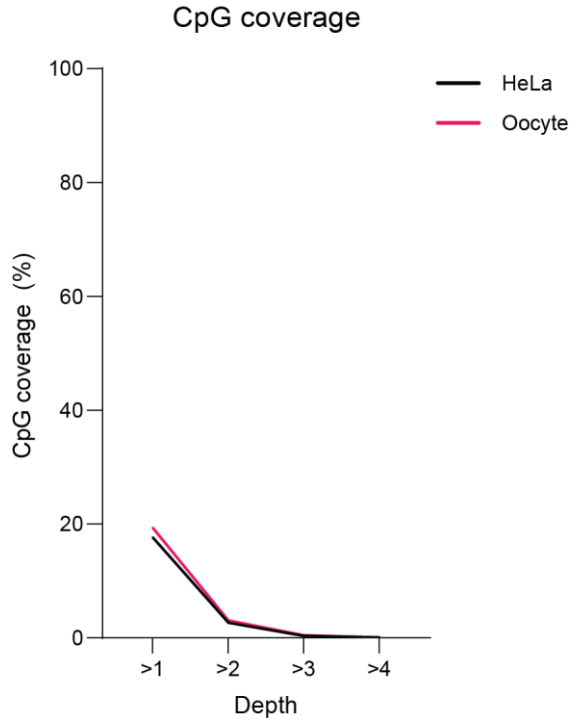


Figure 6.5 Pilot Study 1 CpG coverage
CpG coverage of HeLa and WT GV oocyte BS-seq reads.

6.2.5 Pilot study 2

A second pilot study was conducted with libraries prepared from 20 GV oocytes from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females as well as an RLT+ buffer negative control. DNA concentrations of the libraries were measured, ranging from 9.5 to 13.8 (**Table 6.4**). TapeStation analysis indicated the prepared libraries fell in the expected size range, with libraries ranging from 200 to 500 bp (**Figure 6.6, Appendix 17**).

Table 6.4 Pilot Study 2 library DNA concentrations

Sample	DNA concentration (ng/μL)
RLT+ buffer	11.4
WT	13.8
Padi6$^{C663A/C663A}$	9.5
Padi6$^{-/-}$	11.1

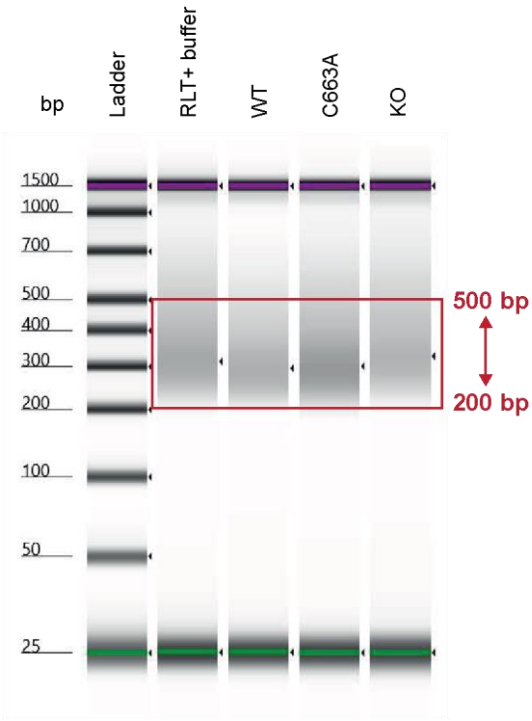


Figure 6.6 Pilot Study 2 library fragment distributions
High Sensitivity DNA ScreenTape gel of libraries prepared from RLT+ buffer, 20 WT, 20 *Padi6*$^{C663A/C663A}$, and 20 *Padi6*$^{-/-}$ GV oocytes, run on 4200 TapeStation System. Libraries were subjected to a final bead clean-up of two rounds of 1x SPRIselect. Black dashed lines indicate expected library fragment distribution.

Concerningly, the RLT+ buffer negative control sample had both high DNA concentration and a fragment distribution between 200 and 500 bp, indicating the sample had been contaminated with exogenous DNA. Each of the libraries were subsequently sequenced to 20 million reads by the Crick Advanced Sequencing Facility, to allow for coverage comparison at different read depths. Sequencing results were processed as before, and a summary of the library sequencing data and processing statistics is displayed in **Table 6.5.**

Table 6.5 Pilot Study 2 sequencing and data processing statistics after 20 million reads

	RLT+	WT	*Padi6*$^{C663A/C663A}$	*Padi6*$^{-/-}$
Total sequences (million)	27.6	22.1	19.6	25.6
FastQC Duplicated reads (%)	58.6	35	35.1	48.8
GC content (%)	33	28	29	28
Mapping to mouse genome				
Mapping efficiency (%)	0	43.8	44.8	40.3
Reads removed in deduplication (%)	-	40.7	41.8	56.4
Sequences left after deduplication (million)	-	5.8	5.1	4.5
CpG methylation (%)	-	39.9	40.1	39.8
CHG methylation (%)	-	4.2	6.9	4.3
CHH methylation (%)	-	4	6.6	4

The RLT+ buffer negative control sample mapped with 0% efficiency to the mouse genome suggesting no mouse DNA was present. The presence of DNA library within the expected fragment range in the TapeStation analysis, however, suggests that this sample has been contaminated during library preparation. This is possibly due to DNA contamination in the micropipettes, or surfaces used. To remedy this, it was decided that the full experiment library prep would be conducted in a UV-sterilised PCR hood, using UV-sterilised pipettes, to prevent any contamination.

The oocyte samples were then re-submitted for sequencing to 200 million reads to determine if this improved CpG coverage. Library sequencing data and processing statistics after sequencing to 200 million reads are displayed in **Table 6.6**.

Table 6.6 Pilot Study 2 sequencing and data processing statistics after 200 million reads

	RLT+	WT	*Padi6*$^{C663A/C663A}$	*Padi6*$^{-/-}$
Total sequences (million)	-	201	182.4	228.6
FastQC Duplicated reads (%)	-		35.1	48.8
GC content (%)	-			
Mapping to mouse genome				
Mapping efficiency (%)	-	43.6	44.6	39.9
Reads removed in deduplication (%)	-	77.1	80.2	88.2
Sequences left after deduplication (million)		20	16.1	10.7
CpG methylation (%)	-	39.2	39.5	39
CHG methylation (%)	-	4	6.9	4.2
CHH methylation (%)	-	3.8	6.6	3.9

At both 20 million and 200 million approximate reads, mapping efficiency across oocyte samples remained similar at around 40-45%. However, the percentage of duplicated reads and percentage of the library removed after deduplication was higher in the 200 million read dataset vs. 20 million. Despite this, sequencing to 200 million reads more than doubled the final sequences after deduplication for each genotype. The CpG methylation in this second pilot study matched the literature better than the initial study, with CpG methylation levels between 39% and 40.1% across all samples compared to the literature at 37.9% (214).

Sequencing to 200 million reads also increased CpG coverage dramatically, with between 26% and 43% of CpGs being covered at least once, 6% and 15% being covered twice, and 1.2% and 4.6% being covered three times (**Figure 6.7**). Based on these data, it was decided that the current protocol was appropriate to apply to the full sample set of collected oocytes.

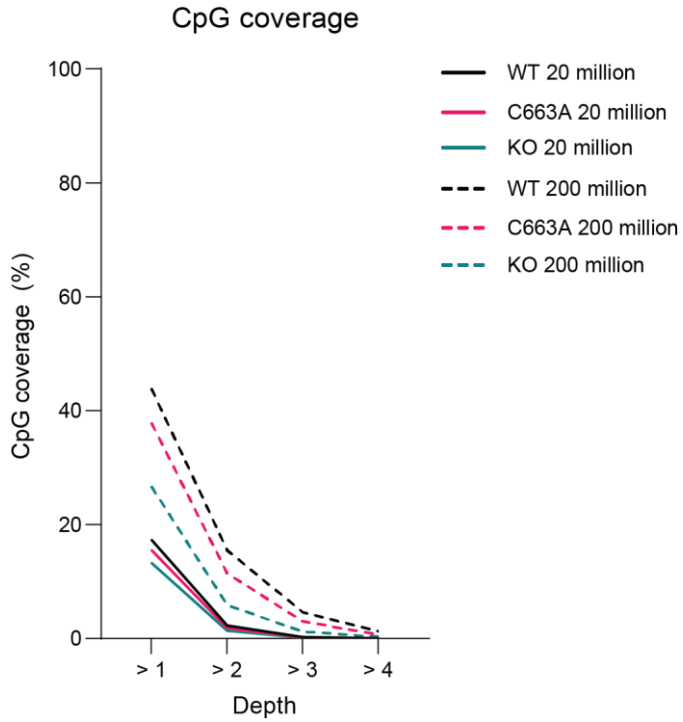


Figure 6.7 Pilot Study 2 CpG coverage
CpG coverage of WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ GV oocyte BS-seq reads.

6.3 Bisulphite sequencing of *Padi6* transgenic GV oocytes

Four 20 GV oocyte samples per genotype were bisulphite converted and libraries prepared with the Zymo Pico Methyl-Seq Library Prep Kit, incorporating the optimisations discussed in **Chapter 6.2**. After DNA quantification and TapeStation analysis (**Table 6.7, Figure 6.8, Appendix 18**), all libraries appeared to be as expected with fragment distributions between 200 and 500 bp, except for WT replicate 2 which had a very low DNA concentration and *Padi6*$^{C663A/C663A}$ replicate 4 which had a fragment distribution of 300 to 1000 bp. Despite concerns about these two libraries, it was elected to sequence all libraries, as further QC could be performed post-sequencing. As such, all samples were sequenced by the Crick Advanced Sequencing Facility to 200 million reads. The RLT+ buffer negative control sample had a very low DNA concentration and no fragment distribution in the expected range, suggesting preparing the library in a UV-hood with UV-sterilised pipettes had succeeded in preventing exogenous contamination.

Table 6.7 Full experiment library DNA concentrations

Sample	Replicate	DNA concentration (ng/μL)
WT	1	12.4
	2	1.3
	3	11.8
	4	11.2
Padi6$^{C663A/C663A}$	1	9.6
	2	7.8
	3	9.8
	4	7
Padi6$^{-/-}$	1	8.4
	2	5.4
	3	9.7
	4	9.9
RLT+	-	2.9

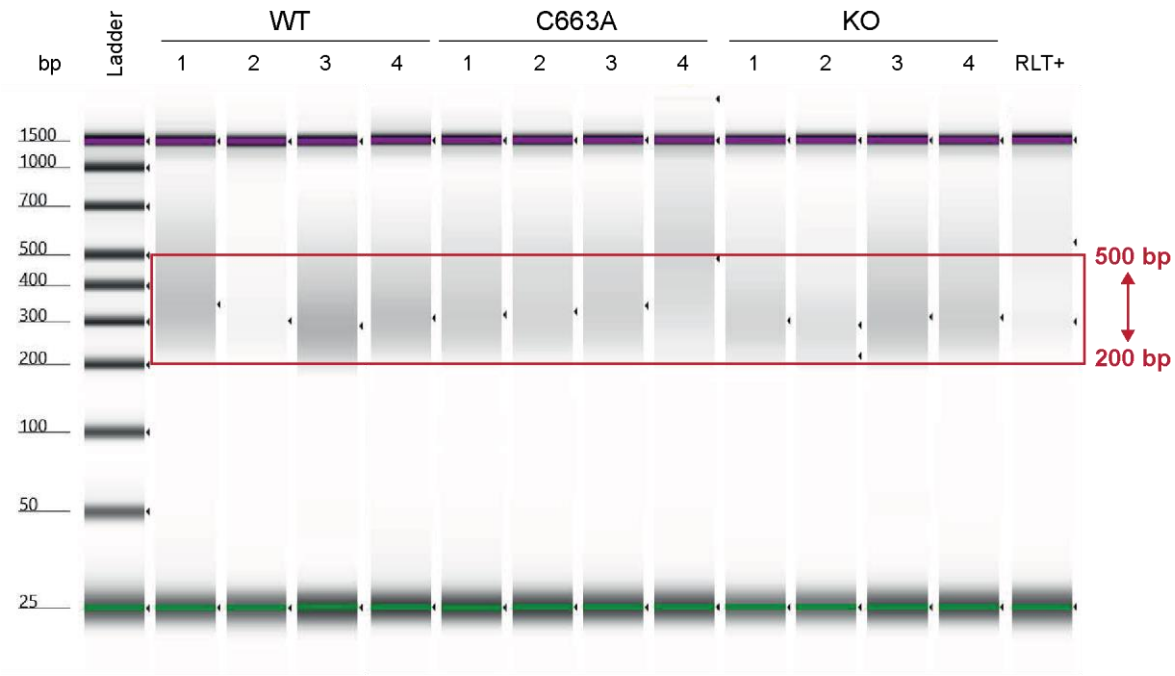


Figure 6.8 Full experiment library fragment distributions
High Sensitivity DNA ScreenTape gel of libraries prepared from RLT+ buffer, and four replicates of 20 WT, 20 *Padi6*$^{C663A/C663A}$, and 20 *Padi6*$^{-/-}$ GV oocytes, run on 4200 TapeStation System. Libraries were subjected to a final bead clean-up of two rounds of 1x SPRIselect. Red box indicates expected library fragment distribution.

Mapping efficiency, the number of unique sequences after deduplication and methylation percentages were all comparable to Pilot Study 2 conducted in **Chapter 6.2.5**, with methylation levels matching the literature (**Figure 6.9**, **Appendix 19**) (214). All four *Padi6*$^{-/-}$ replicates had lower duplication rates and more unique sequences left after deduplication than in Pilot Study 2. WT replicate 2 and *Padi6*$^{C663A/C663A}$ replicate 4 both had very high duplication rates with over 90% of the library removed after deduplication compared to an average of 76.6% across all other samples. Despite this their overall CpG, CHG, and CHH methylation percentages were as expected. Excluding WT replicate 2 and *Padi6*$^{C663A/C663A}$ replicate 4, which were left with 3.8 and 5.3 million respectively, all samples were left with between 16.3 and 23.6 million unique sequences after deduplication. The high duplication and few sequences left after deduplication of WT replicate 2 and *Padi6*$^{C663A/C663A}$ replicate 4 is likely due to sample loss during library preparation. Despite the limited number of reads, CpG, CHG, and CHH methylation percentages were as expected and comparable to the other replicates.

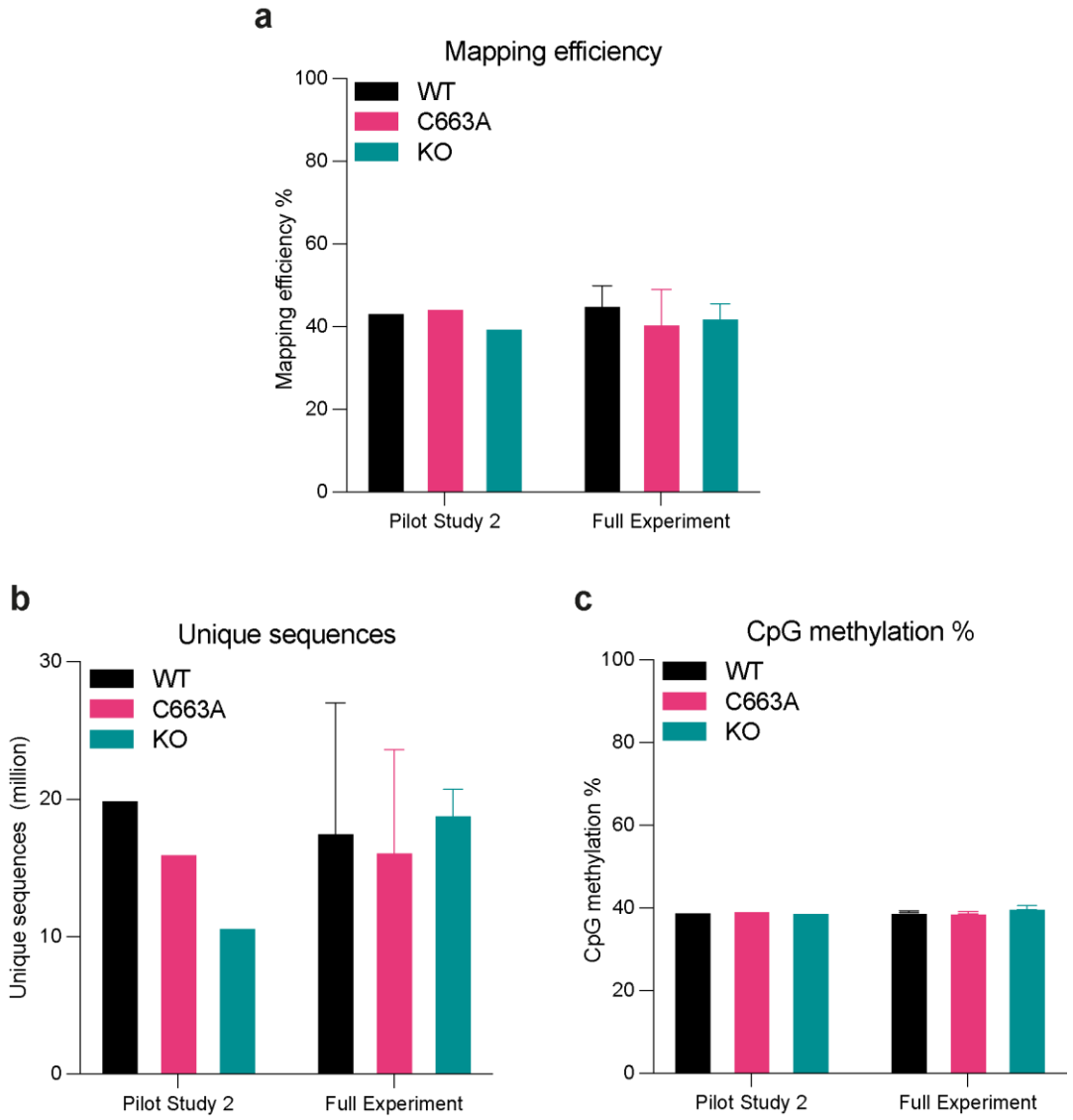


Figure 6.9 Data processing summary of full experiment compared with Pilot Study 2
a) - c) Mapping efficiency to the GRCm39 genome, unique sequences after deduplication, and CpG methylation % of Pilot Study 2 data (1 replicate of 20 WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ GV oocytes) and Full Experiment data (4 replicates of 20 WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ GV oocytes).

To get an initial view on any differences in CpG methylation between oocytes of each genotype, PCA analysis was conducted (**Figure 6.10**). Excluding one WT and one *Padi6*$^{C663A/C663A}$ replicate, all the WT and *Padi6*$^{C663A/C663A}$ samples appeared to cluster together. WT replicate 2 and *Padi6*$^{C663A/C663A}$ replicate 4 did not cluster with the other WT and *Padi6*$^{C663A/C663A}$ samples and is likely a result of the sample loss or contamination during library preparation. As such, it was decided that these replicates would be excluded from further analysis. Interestingly, all the *Padi6*$^{-/-}$ samples did not cluster with the other samples, and were dispersed along both PC1 and PC2, indicating a difference in the methylation states of these samples.

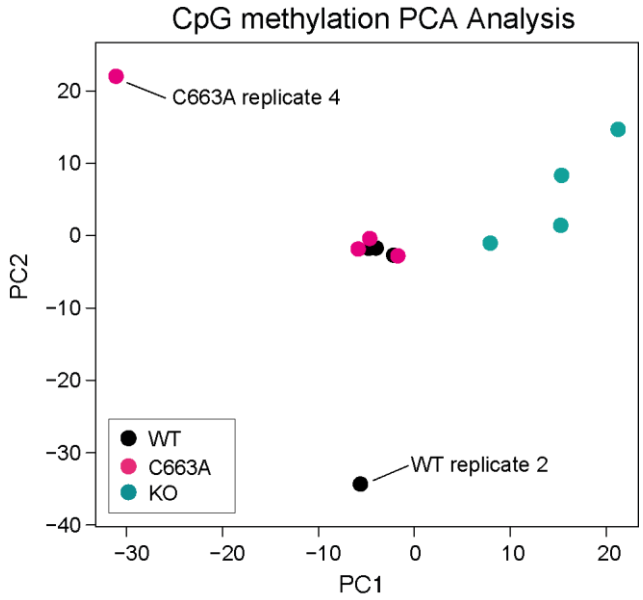


Figure 6.10 PCA of CpG methylation in GV oocytes
Plot of PC1 vs. PC2 after PCA of CpG methylation of four replicates of 20 GV oocytes from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females. PCA performed using the R package methylKit.

The CpG coverage of each sample was comparable to Pilot Study 2 (**Figure 6.11**), with averages for each dataset of around 38.4% of CpGs being covered once, 13.5% twice, 4.2% three times and 1.2% four times. It was decided that to further improve the overall coverage, replicate data from each genotype would be combined *in silico*. Combination of replicate data resulted in significant improvements to the CpG coverage with now around 77.8% of all CpGs being covered once for each genotype, 57.8% twice, 39.3% three times, and 24.7% four times (**Figure 6.11**).

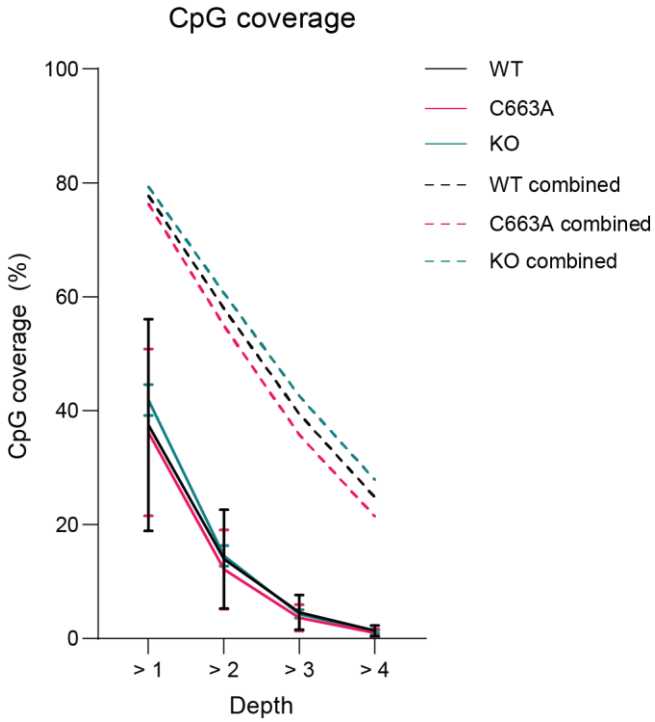


Figure 6.11 Full experiment CpG coverage
CpG coverage of full experiment BS-seq reads before and after *in silico* combination. Dashed lines indicate coverage after data from each of the four replicates for each genotype was combined.

Using R methylKit, percentage methylation in the combined datasets at 2000 bp non-overlapping bins (with >100 CpG reads), as well as at different genomic loci (with >20 CpG reads) were calculated (**Figure 6.12a-e**). Overall, there appeared to be little difference between the methylation states of oocytes from WT and *Padi6*$^{C663A/C663A}$ females across the genome and at CGIs, genes, promoters and intergenic regions. There does however appear to be a slight increase in overall methylation percentage between oocytes from *Padi6*$^{-/-}$ females over WT females, at each of the genomic loci tested.

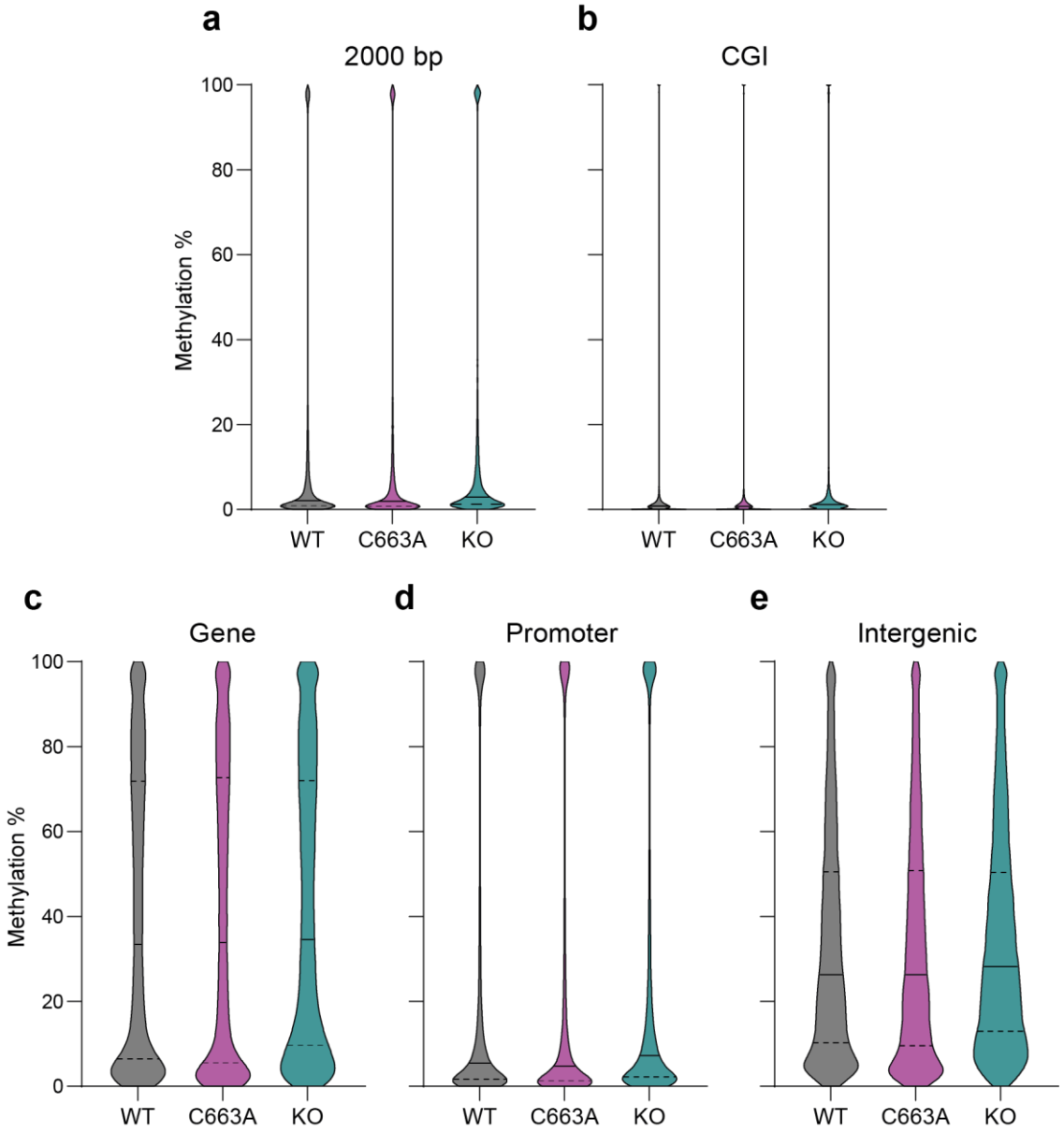


Figure 6.12 Methylation percentage at different genomic loci in GV oocytes
a) - e) Average methylation percentage in non-overlapping 2000 bp windows, CGIs, genes, promoters and intergenic regions. 2000 bp windows were limited to those covered by at least 100 CpG reads. CGIs, genes, promoters, and intergenic regions were limited to those that were covered by at least 20 CpG reads. WT = *in silico* combined data of three replicates of 20 WT GV oocytes. C663A = *in silico* combined data of three replicates of 20 *Padi6*$^{C663A/C663A}$ GV oocytes. KO = *in silico* combined data of four replicates of 20 *Padi6*$^{-/-}$ GV oocytes. Methylation percentage calculated using the R package methylKit.

To get a better understanding of potential methylation differences, scatter plots of methylation percentage at different loci were produced, comparing WT to *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$, and the Pearson correlation coefficient calculated for each (**Figure 6.13a-e**). Strong correlation between WT and *Padi6*$^{C663A/C663A}$ or *Padi6*$^{-/-}$ methylation levels was observed for every group of genomic loci. The lowest and most

dispersed correlation in methylation levels was observed between WT and *Padi6*$^{-/-}$ oocytes at CGIs, however a strong correlation with a Pearson coefficient of 0.9608 was still observed. The CGIs that were differentially methylated between WT and *Padi6*$^{-/-}$ GV oocytes were extracted, with differential methylation being defined as a 25% difference in methylation levels. A total of 90 CGIs were hypomethylated, and 110 CGIs hypermethylated in the *Padi6*$^{-/-}$ GV oocytes with a q-value less than 0.01, out of a total of 15,719 CGIs that had a minimum coverage of 20 CpGs. None of the differentially methylated CGIs aligned with known mouse DMRs and did not enrich for any terms after GO analysis. Given many proteins and mRNAs have been observed to be down and upregulated in *Padi6*$^{-/-}$ GV oocytes, it is conceivable that these methylation differences are an indirect effect of knocking out *Padi6* (67). As mentioned in **Chapter 1.3.1**, gene transcription has been identified as a driver of methylation in the developing oocyte. As such, differences in gene transcription in *Padi6*$^{-/-}$ oocytes could be leading to the altered methylation levels at these sites.

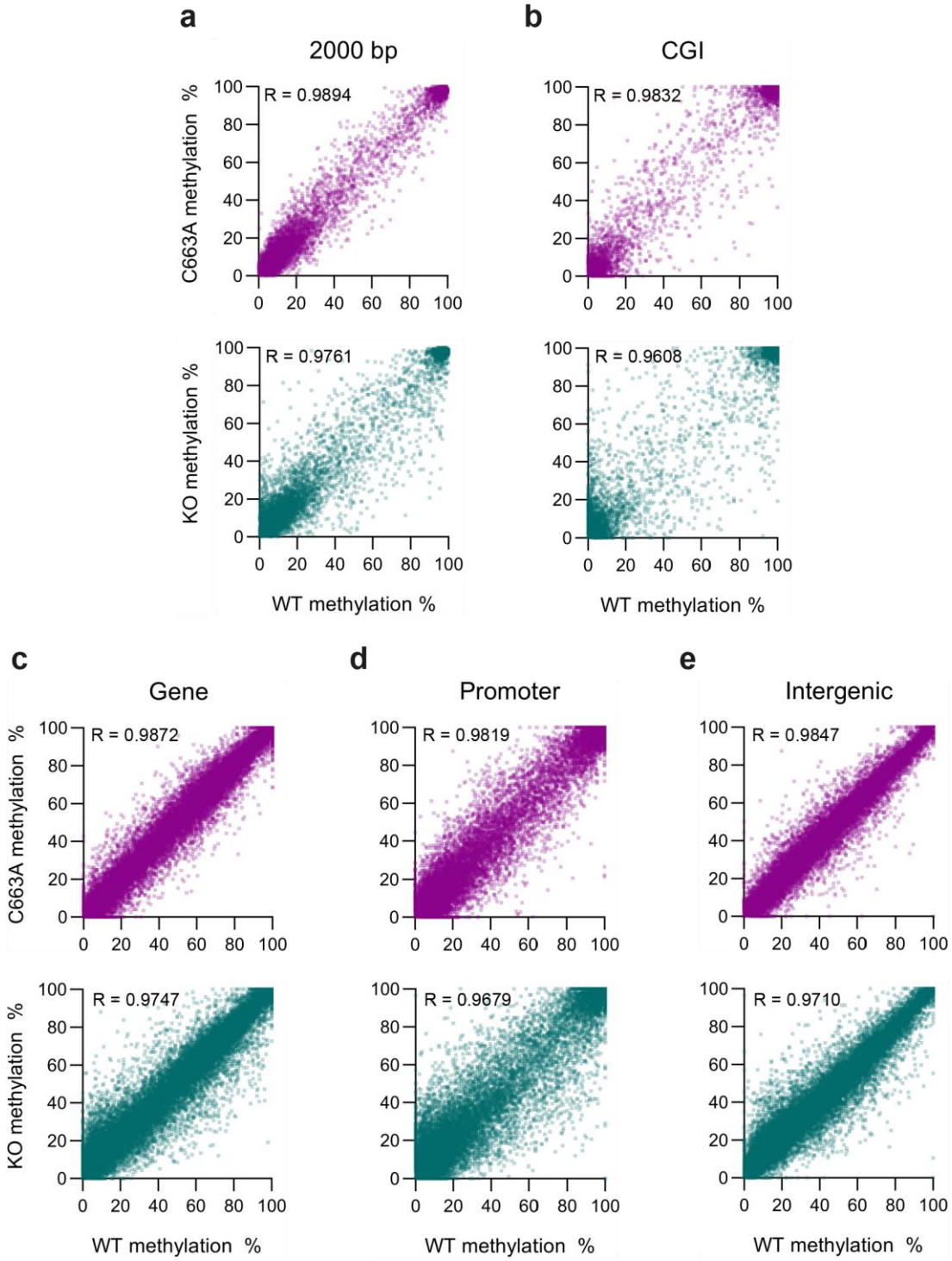


Figure 6.13 Scatter plot of methylation percentage at different genomic loci in GV oocytes

a) - e) Percent methylation across the genome divided into non-overlapping 2000 bp windows, CGIs, genes, promoters and intergenic regions. Top Panel: *Padi6*$^{C663A/C663A}$ vs. WT, Bottom Panel: *Padi6*$^{-/-}$ vs. WT. Individual loci were limited to those that were covered by at least 20 CpG reads. WT = *in silico* combined data of three replicates of 20 WT GV oocytes. C663A = *in silico* combined data of three replicates of 20 *Padi6*$^{C663A/C663A}$ GV oocytes. KO = *in silico* combined data of four replicates of 20 *Padi6*$^{-/-}$ GV oocytes. Methylation percentages calculated using the R package methylKit. R = Pearson correlation coefficients calculated in Prism.

The key phenotype observed in oocytes or offspring of women with PADI6 or SCMC gene variants, however, is hypomethylation at maternal DMRs leading to imprinting disorders. The level of hypomethylation, and the DMRs that are hypomethylated differ between patients and the causative variants in the mother, however, in the case of *PADI6* variants, reductions in methylation of over 50% have been reported (159). A list of the genomic locations of known mouse maternal germline DMRs as reported in (252) was lifted onto the GRCm39 genome coordinates and percentage methylation at each gDMR was calculated using methylKit (**Figure 6.14**). Out of the 25 maternal gDMRs, one did not pass the minimum coverage of 20 CpGs and so was excluded from the analysis. All these maternal gDMRs are known to be methylated on the maternal allele and as expected in the WT sample, all gDMRs had percentage methylation of above 80% with an average of 96.8%. Interestingly, all of the gDMRs in both the *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ had above 90% methylation, averaging 98% and 98.1% respectively, suggesting neither *Padi6* variant has in any defects in the establishment of the maternal imprinted methylome.

Together, these data suggest that there are no defects in the establishment of the methylation signature at maternal gDMRs in both *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ GV oocytes. Furthermore, there does not appear to be any dramatic methylation defects genome wide, except for a few CGIs. This contrasts to the genome wide, and dramatic hypomethylation observed in *KHDC3*L variant human oocytes (231).

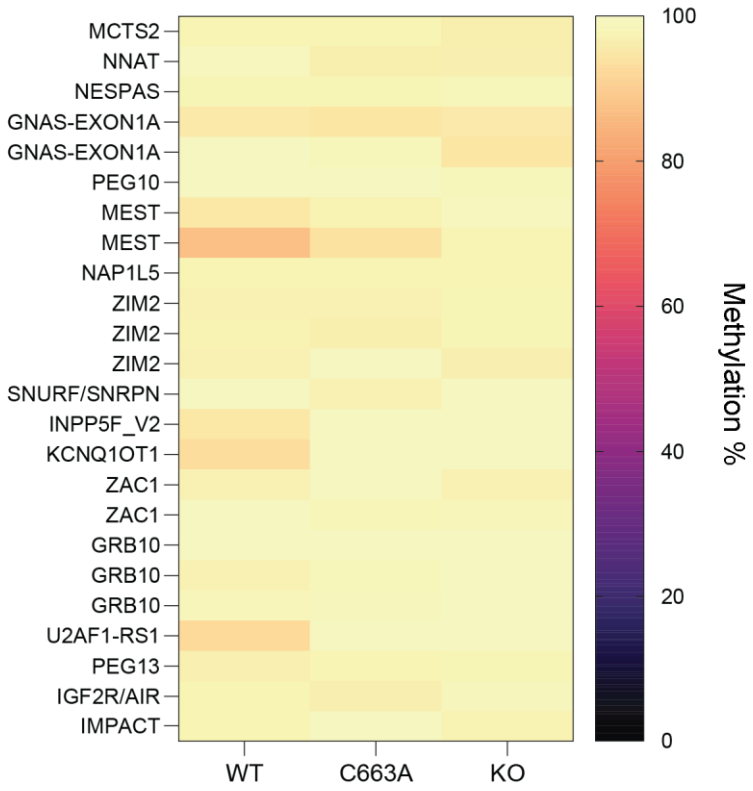


Figure 6.14 Mouse maternal gDMR methylation heatmap in GV oocytes
Heatmap of methylation percentage across 24 known mouse maternal gDMRs with a minimum coverage of 20 CpGs. Methylation percentage calculated using R methylKit. WT = *in silico* combined data of four replicates of 20 WT GV oocytes. C663A = *in silico* combined data of four replicates of 20 *Padi6*$^{C663A/C663A}$ GV oocytes. KO = *in silico* combined data of four replicates of 20 *Padi6*$^{-/-}$ GV oocytes. GV = germinal vesicle.

6.4 Summary

To investigate the function of PADI6 in the establishment of the maternal imprinted signature, I performed low input BS-seq on pools of GV oocytes from WT, *Padi6*$^{C663A/C663A}$ and *Padi6*$^{-/-}$ females. I had hypothesised that given SCMC protein KHDC3L has been shown to be vital for the establishment of the maternal imprinted methylation signature in humans, that PADI6 was likely also involved in this establishment. Contrastingly, the data suggests that absence of PADI6 potential catalytic activity, or PADI6 protein in mouse oocytes has little to no effect on the establishment of the methylome, and in particular, the methylation of maternal imprinting gDMRs. 200 CGIs out of the 15,719 that were covered by at least 20 CpGs were differentially methylated in *Padi6*$^{-/-}$ oocytes. This difference in methylation levels could be an indirect effect, attributed to differences in transcription caused by knocking

out *Padi6*. scRNA-seq analysis of oocytes at different stages of maturation would be a possible experiment to test this hypothesis. Altogether, this work suggests that PADI6 is not required for the establishment of the maternal imprinted methylation signature in mice.

This result is overall surprising given the dramatic methylation defects observed in human GV oocytes from an infertile woman carrying *KHDC3L* variants. Absent from this study, however, was an investigation into whether there are any defects in the maintenance of the maternal methylome when *Padi6* is knocked out. In this work, it was unfeasible to gather enough 2-cell embryos in a reasonable timeframe to probe possible defects in methylation maintenance using the implemented workflow. The recently developed Enzymatic Methyl-seq (EM-seq), however, offers a solution to this problem. EM-seq involves two enzyme catalysed reactions to sequence the methylome from picograms of input DNA (283). First, DNA is treated with TET2 and T4-BGT which oxidises both 5mC and hydroxy-methyl cytosine (5hmC). Secondly, DNA is treated with APOBEC3A which deaminates unmethylated cytosine residues to uracil, leaving the oxidised 5mC and 5hmC residues as cytosines. This treated DNA is then processed, sequenced and analysed as for bisulphite sequencing. Crucially, the enzymatic treatment steps do not damage and fragment the DNA as bisulphite does, meaning input quantities can be significantly lowered. This would allow higher coverage from fewer oocytes or embryos, and thus potentially allow the acquisition of high coverage methylation information from fewer mat-*Padi6*$^{-/-}$ embryos. Furthermore, time-lapse studies of the developmental potential of embryos from *Padi6*$^{-/-}$ females in **Chapter 5.3** highlighted that a small subset of embryos do develop beyond the 2-cell stage, with some even forming blastocyst like structures. BS-sequencing of these later stage embryos could pose an alternative to the harvesting of large quantities of 2-cell embryos as far more copies of the genome would be present. If PADI6 is only involved in the maintenance of the methylation signature and not the establishment, then this raises the question of whether PADI6 and the SCMC function together in maternal imprinting. Or, if the functions of PADI6 and the SCMC are not conserved between humans and mice, and in mice neither are required for proper maternal imprinting.

It is conceivable that the molecular mechanisms governing the establishment of the mouse and human maternal imprinted signatures are not conserved. It is interesting that in the oocytes of a woman with a *KHDC3L* variant, dramatic

methylation defects are observed (231). If PADI6 and the SCMC are indeed functioning together as other reported data suggests (80), then it would make sense that deletion of *Padi6* would lead to similar methylation defects in the oocyte. Given the lack of any methylation defects in *Padi6*$^{-/-}$ GV oocytes, it would then follow that the mechanisms regulating establishment of the maternal methylome are not conserved between humans and mice. One line of evidence to further support this is the absence of DNMT3L in human oocytes (167). DNMT3L, in conjunction with the de novo methylation enzyme DNMT3A, have been shown to be critical for the establishment of the maternal methylome during oocyte growth in mice (179). In humans on the other hand, DNMT3L is not expressed in the oocyte, suggesting that either DNMT3A is acting alone or relying on alternate co-factors for its regulation. With little evidence one cannot say whether these alternate co-factors include PADI6 and the SCMC, however a scenario such as this would explain the methylation differences observed between human and mouse oocytes with PADI6 or SCMC variants.

Overall, the molecular mechanisms behind the function of PADI6 in maternal imprinting remain unclear. Whilst differential methylation is observed in *Padi6*$^{-/-}$ GV oocytes at a small subset of CGIs, further work is required to understand whether these are differences are due to a direct effect, transcriptional dysregulation, or are just stochastic defects caused by an overall state of cellular stress. Additionally, the methylation sequencing of later stage mat-*Padi6*$^{-/-}$ embryos is of the upmost interest to confirm if PADI6 instead functions in the maintenance of the maternal methylome, not the establishment. This work also raises further questions regarding the conservation of maternal imprinting pathways in humans and mice, as well as conservation of function of PADI6 and the SCMC between the two species.

Chapter 7. Discussion

7.1 Overview

PADI6 is a maternal factor that is vital for early embryo development in mice and humans. Maternal variants in the *PADI6* gene most commonly leads to infertility and early embryonic developmental arrest. In some cases, women are fertile, but their pregnancies are often plagued by high incidences of complications such as miscarriages, and any offspring almost always born with MLID. Despite the clear importance of PADI6 to the early embryo, the molecular mechanisms behind this function remain poorly understood.

PADI6 belongs to a family of enzymes that catalyse the post-translational conversion of arginine to citrulline in a process known as citrullination. Whilst PADI6 has been identified as the fifth member of this family, there is currently no reported catalytic activity. Whether PADI6 has a catalytic function, and is indeed an arginine deiminase, is a question that must be answered to fully understand the function of PADI6. On top of this, PADI6 is less conserved with the other family members than they are with each other. Crucially, whilst three of the four catalytic tetrad residues are conserved, the key catalytic cysteine is replaced by an alanine, flanked by two cysteines in human PADI6. Only the first of these two cysteines is conserved in the sequence of PADI6 in multiple species suggesting that if PADI6 is catalytically active, then it would be through this first cysteine. Additionally, more than half the residues that are involved in calcium binding in PADIs 1 to 4 are not conserved in PADI6, raising the question of whether PADI6 binds, and by extension is activated by calcium.

Work has gone into trying to elucidate the function of PADI6 in the early embryo. In work by Coonrod and co-workers, it was found that PADI6, along with the SCMC was crucial for the for the formation of the CPLs, a cytoskeletal structure unique to the oocyte and early embryo. Absence of PADI6 or SCMC proteins in the oocyte or embryo, not only resulted in the absence of the CPLs, but defects in organelle and spindle localisation, and defects in ZGA. It was not until 2023, that work by Schuh and co-workers determined that the CPLs were made up of protein fibres, and that PADI6 and the SCMC, were actually within the CPLs, potentially constituting key structural components. It was further suggested that the CPLs are a storage site for maternal

proteins, and absence of the CPLs results in the downregulation of many important embryonic factors. Whether the phenotypes seen in *Padi6* knock-out mouse embryos, or *PADI6* variant human embryos are due to a downstream effect of the CLPLs being absent, or an absence of a direct PADI6 function is not clear.

In this thesis, I aimed to characterise any possible catalytic function of PADI6, as well as to elucidate the function of PADI6 in early embryo development. Towards this, I investigated the following questions:

- What effect does the lack of sequence conservation have on the overall structure of PADI6? What can we learn about the biochemical function of PADI6 from this structure? Investigated in **Chapter 3**.
- Does PADI6 interact with the SCMC? What is the nature and function of this interaction? Investigated in **Chapter 4**.
- What is the molecular function of PADI6 in early embryo development? Is PADI6 catalytic activity required for early embryo development? Investigated in **Chapter 5**.
- How is PADI6 involved in the establishment of the maternal imprinted methylation signature? Investigated in **Chapter 6**.

All together this work answers some of the outstanding questions regarding PADI6 and offers new perspectives regarding its function (**Figure 7.1, Figure 7.2**). I have confirmed that PADI6 does not bind Ca^{2+} like PADIs 1 to 4, it forms a dimer, and in the purified structural conformation is likely incapable of catalytic activity. Additionally, I have confirmed that human PADI6 does interact with the SCMC, and this work offers new insights to how the CPLs are formed and composed structurally. Recombinant expression of the SCMC and structural characterisation of the interactions could provide invaluable information to the question of how the CPLs are structurally arranged and remains a key future aim of this work. I also found that PADI6 is not required for the establishment of the maternally imprinted methylation signature in mouse oocytes. The question now remains as to whether PADI6 is involved in the maintenance of the methylation signature, or whether the mechanisms governing maternal imprinting are not conserved between mice and humans. This then raises the question of whether the functions of PADI6 and the SCMC are conserved between the two species, and furthermore, highlights the lack of conservation in imprinting mechanisms between mice and humans.

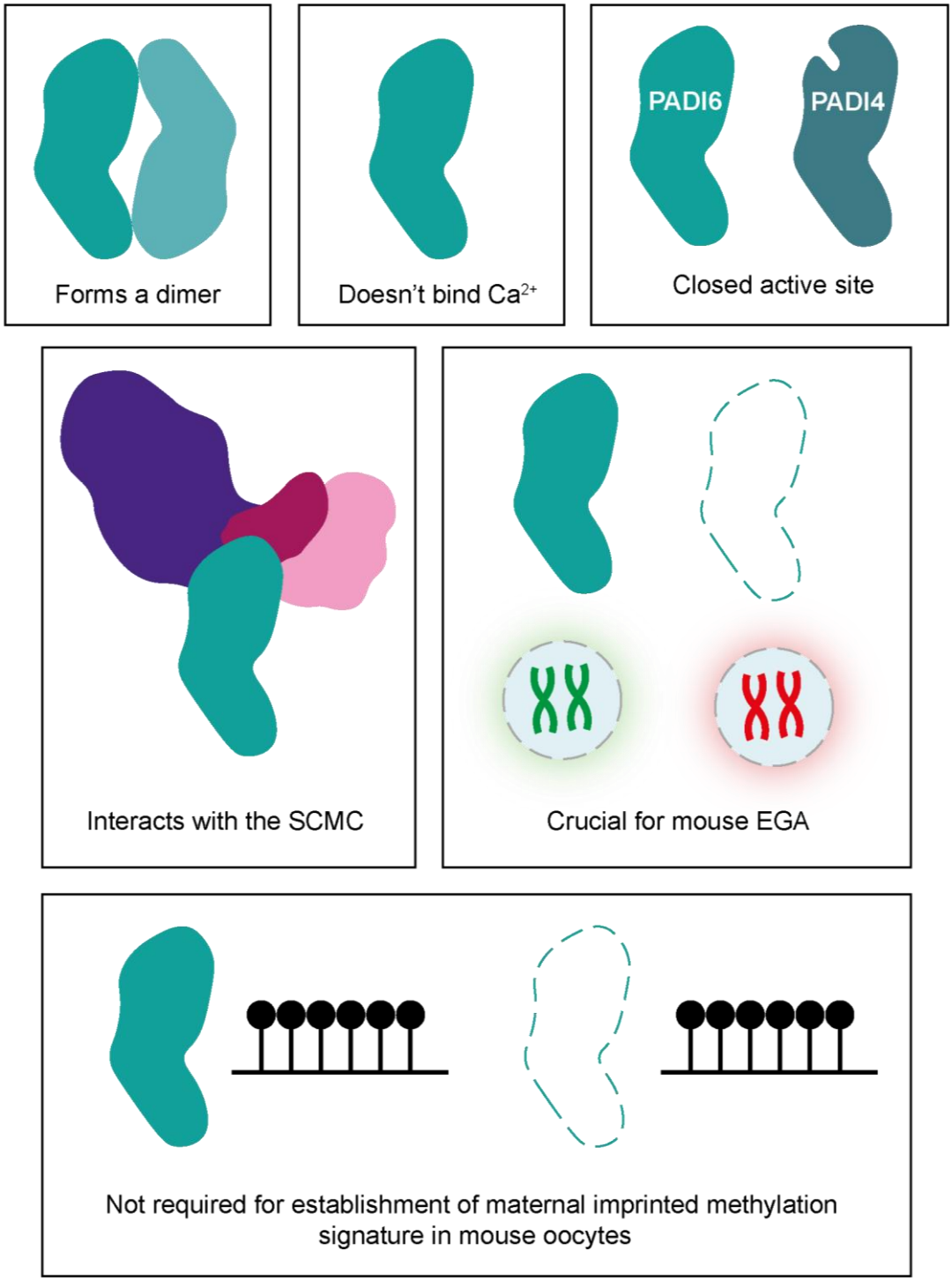


Figure 7.1 Questions answered in this work
Recombinant human and mouse PADI6 form dimers in solution and hPADI6 doesn't bind Ca^{2+}. The active site of hPADI6 is closed. hPADI6 interacts with the SCMC. mPADI6 is important for proper mouse EGA and is not required for the establishment of the maternal imprinted methylation signature in mouse oocytes.

The mouse models used in this work further validate the hypothesis that PADI6 is required for EGA in mice. Interestingly, the data suggests that there may be an additional non-essential catalytic function of mPADI6. Proteomics data is now required to further validate these findings and confirm if there are indeed proteomic defects in the early embryos of mice with catalytically dead PADI6.

Possibly the most significant outstanding question in my opinion now is, are all the observed embryonic phenotypes a downstream result of the lack of CPLs in the oocyte. Or are there additional unique functions of mPADI6 that contribute to the phenotypes observed. Confirmed in this work, *Padi6*$^{C663A/C663A}$ oocytes do possess an abundance of CPLs. If proteomic analysis does confirm that there are defects in mat-*Padi6*$^{C663A/C663A}$ embryos then this result would confirm that there is indeed an additional function of mPADI6, unique from forming the CPLs. Furthermore, given EGA appears dysregulated in mat-*Padi6*$^{C663A/C663A}$ 2-cell embryos, this model is a powerful platform for studying the function of mPADI6 in EGA and other early embryonic developmental processes, separate from the CPLs. Further experimentation will be required to confirm whether this additional function, is catalytic.

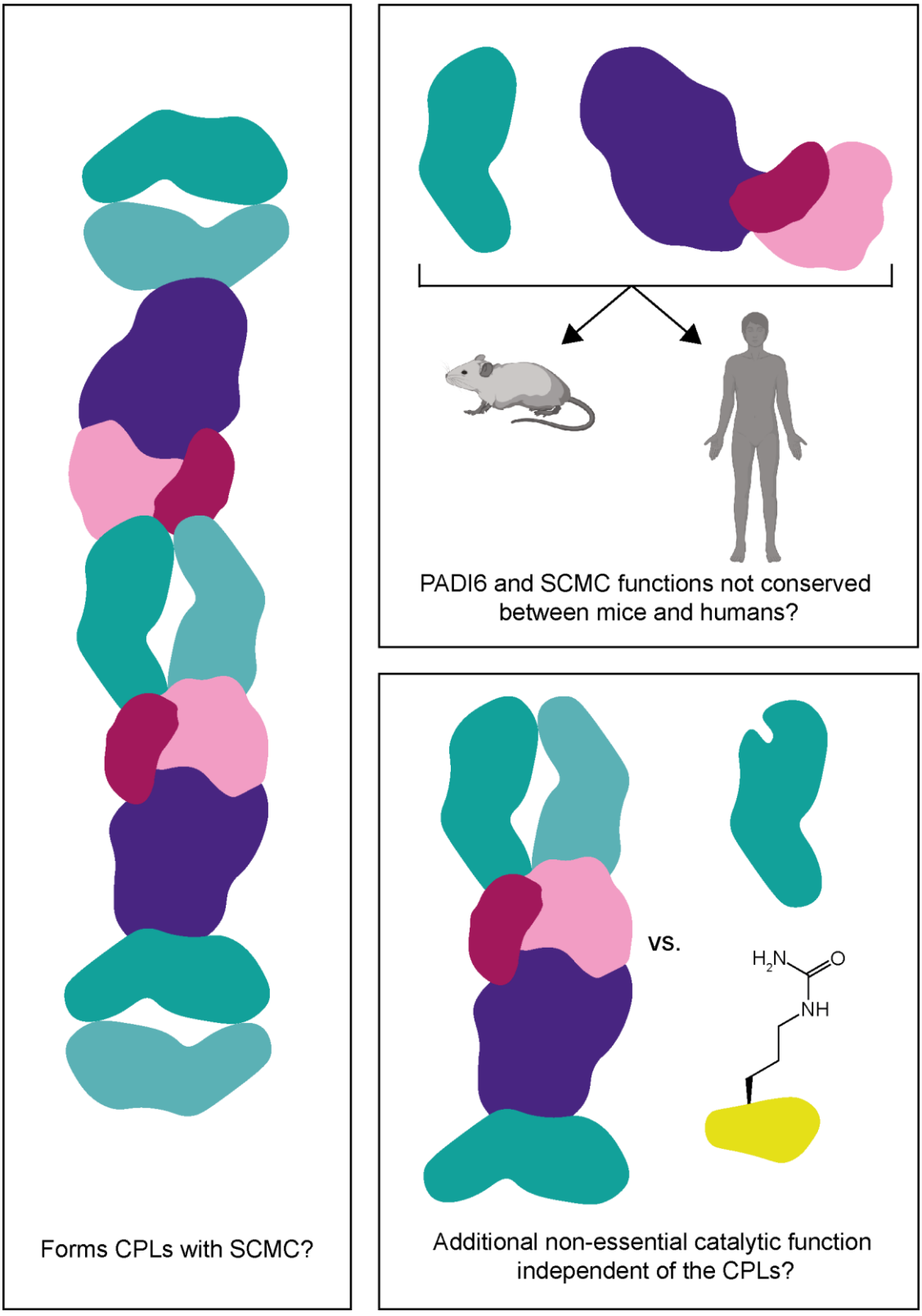


Figure 7.2 New questions and perspectives regarding the function of PADI6
Does PADI6 and SCMC interact together to form the CPL fibres, what is the structural arrangement of these fibres, and are other proteins required? Are the functions of PADI6 and the SCMC conserved between mice and humans? Is there an additional non-essential catalytic function of PADI6, independent of the CPLs?

7.2 Does PADI6 have a catalytic function?

The question of whether PADI6 has a catalytic function has been a focus throughout my work. I have replicated reported results that mPADI6 does not citrullinate BAEE *in vitro* and confirmed that this is also the case with hPADI6. Looking at the X-ray crystal structure of hPADI6, this is not surprising given that PADI6 does not appear to have an open active site pocket for a substrate arginine to fit into. Moving beyond this, whilst the positioning of the catalytic tetrad aspartic acids and histidine is conserved between hPADI6 and hPADI4, the predicted key catalytic cysteine is displaced further away in hPADI6. Together, it appears that without structural rearrangement, human PADI6 would be incapable of catalysing citrullination in this state. It must be noted however, that calcium binding causes structural rearrangements in PADIs 1 to 4 that open the active site pocket and move the key catalytic cysteine into a position where it can react with the guanidinium carbon of a substrate.

I have shown in this work that hPADI6 is not stabilised by increasing calcium concentrations, indicating that as predicted based on the lack of conservation in calcium binding residues, hPADI6 does not bind calcium like the other hPADIs. As mentioned previously, an evolutionary explanation for this could be to avoid PADI6 activity being altered during the large calcium transients induced by fertilisation. If this is indeed the case, then it is conceivable that PADI6 relies on interactions with co-activating proteins to induce the structural changes required for catalytic activity. I confirmed in this work that hPADI6 interacts with the human SCMC and subsequently tested my hypothesis that PADI6 requires protein co-factors for catalytic activation. As monitored by a citrulline reactive chemical probe, co-expression of hPADI6 with the SCMC proteins in HEK-293T cells did not result in any increases in citrullination. Whilst this result aligns with PADI6 not being catalytically active, it again cannot be ruled out that either the SCMC proteins tested are not the correct co-factors required for hPADI6 activity, or there are no natural hPADI6 substrates in the HEK-293T proteome. If PADI6 is catalytically active and does, as hypothesised, rely on protein co-factors for activation, without the knowledge of what these co-factors are, and furthermore, what the natural substrates are, it will be difficult to conclusively answer this question using biochemical experiments alone. With this in mind, we turned to a mouse model to gather further information on the possibility of PADI6 catalytic activity, now specifically in the embryo.

Using a mouse model with the predicted key catalytic cysteine of mouse PADI6 mutated to alanine, we sought to characterise whether PADI6 catalytic activity is required for early mouse embryo development. We hypothesised that this substitution would ablate any catalytic activity that PADI6 may have, and thus any developmental phenotypes observed would be a result of a lack of PADI6 catalytic activity, and not other structural functions. In testing this hypothesis, I first characterised the fertility of the catalytically inactive PADI6 female mice. *Padi6*$^{C663A/C663A}$ mice were fertile and produced litters of equal size to WT females. This immediately informed us that whilst there may be a catalytic function of PADI6, it is not essential for early embryo development in mice. Interestingly however, when cultured in vitro, fewer mat-*Padi6*$^{C663A/C663A}$ zygotes reached the blastocyst stage compared to WT zygotes. Furthermore, the point at which a large subset of the zygotes arrested or degenerated was at between 4-cell and morula stages, just after EGA. This increased incidence of development arrest or degeneration was accompanied by delays in development, however, by the time of cavity formation this delay was corrected. Together these data was the first indication that there may be a potential additional catalytic function of PADI6, that is not essential for embryo development. Furthermore, it appears that this function is possibly during EGA, and lack of catalytic activity results in downstream defects in the post 2-cell embryos post EGA. The onset of these defects aligns with the onset of defects in mat-*Padi6*$^{-/-}$ embryos, further substantiating the theory that PADI6 functions during EGA.

To further understand this potential partial EGA inhibition in mat-*Padi6*$^{C663A/C663A}$ embryos, I performed scRNA-seq on MII oocytes, zygotes and 2-cell embryos from these catalytically dead PADI6 females. Mat-*Padi6*$^{C663A/C663A}$ MII oocytes and zygotes displayed few differences to WT cells, however, at the 2-cell stage, hundreds of genes were up and downregulated. The most interesting result was that mat-*Padi6*$^{C663A/C663A}$ and mat-*Padi6*$^{-/-}$ 2-cell embryos displayed almost opposite phenotypes. Notably, many of the genes downregulated in mat-*Padi6*$^{-/-}$ 2-cell embryos, were upregulated in mat-*Padi6*$^{C663A/C663A}$. Knocking-out *Padi6* would result in the absence of any catalytic function, as well as any other structural or interactional functions. These data are confusing as one would therefore expect that if PADI6 does have a catalytic function, any genes that are up or downregulated in mat-*Padi6*$^{C663A/C663A}$ embryos would be similarly up or downregulated in mat-*Padi6*$^{-/-}$ embryos. A biological explanation for this opposite effect is not clear at this time.

It is possible that as the CPLs are present in *Padi6*$^{C663A/C663A}$ oocytes, the maternal factors required for EGA are present in similar levels to WT in the zygote. This is in contrast to *Padi6*$^{-/-}$ oocytes where these factors have been shown to be down-regulated (67), and EGA is majorly inhibited as shown in this work. As discussed in **Chapter 1.2.2**, EGA involves dramatic and highly regulated chromatin modifications and re-organisation, as well as the ordered expression of specific transcription factors. It is conceivable that PADI6 could have a catalytic function either in chromatin modification, like PADI2 (23) and PADI4 (24,284), or potentially in directly citrullinating transcription factors. A lack of PADI6 catalytic activity and citrullination could therefore result in mis-regulation of EGA, but not a complete inhibition as seen in mat-*Padi6*$^{-/-}$ embryos. As seen with PADI4 citrullination of DNMT3A, protein citrullination has the capacity to stabilise substrate proteins (6). Comprehensive proteomic information of the mat-*Padi6*$^{C663A/C663A}$ 2-cell embryo could provide answers to whether a potential lack of PADI6 catalytic activity results in the stabilisation, or destabilisation, of EGA required proteins, such as transcription factors or chromatin modellers. Furthermore, it will provide further validation of the scRNA-seq data acquired in this work and provide further clarification on the exact impact that a potential lack of PADI6 catalytic activity has on protein expression during EGA.

Alternatively, it is also conceivable that the function of PADI6 during EGA is not catalytic and that PADI6 functions through interactions with chromatin modellers or by directly interacting with DNA acting as a transcription factor. It would be of interest to determine whether PADI6 has any DNA binding capacity, a property of PADI4 noted by members of the Walport group. If so, then one possibility is that the C663A substitution affects the capacity of PADI6 to bind to DNA and thus disrupting any transcription activation function. On the other hand, the dysregulation of EGA observed in mat-*Padi6*$^{C663A/C663A}$ 2-cell embryos could be due to a disrupted function of PADI6 in chromatin remodelling. scNMT-sequencing could be employed to elucidate this function by characterising both the chromatin accessibility and genomic methylation state of mat-*Padi6*$^{C663A/C663A}$ 2-cell embryos (285).

Finally, to tease apart a possible catalytic or non-catalytic function of PADI6 during EGA, PADI inhibitors could be employed along with time-lapse imaging, scNMT-seq and proteomics. This would allow us to determine whether WT embryos treated with low doses of PADI inhibitors phenocopy mat-*Padi6*$^{C663A/C663A}$ embryos, indicating that

the function of PADI6 is indeed catalytic, although identification of the substrate citrullinated proteins will be required to fully validate any catalytic activity.

Together, this work raises further questions as to whether PADI6 has a catalytic function. It is clear that if PADI6 is catalytically active, then this catalytic activity is not required for early embryo development. However, the possibility remains, that PADI6 has an additional non-essential catalytic function, potentially during mouse EGA. Or alternatively, a non-catalytic function in chromatin remodelling or directly in transcriptional activation. Acquiring a comprehensive single embryo proteomics and chromatin accessibility dataset of catalytic mutant oocytes and embryos, and parallel treatment of WT embryos with PADI inhibitors will be invaluable in answering this question. This work is ongoing and a key aim of this work in the future.

Chapter 8. References

1. Barasa L, Thompson PR. Protein citrullination: inhibition, identification and insertion. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2023;378(1890): 20220240. https://doi.org/10.1098/rstb.2022.0240.

2. György B, Tóth E, Tarcsa E, Falus A, Buzás EI. Citrullination: A posttranslational modification in health and disease. *The International Journal of Biochemistry & Cell Biology*. 2006;38(10): 1662–1677. https://doi.org/10.1016/j.biocel.2006.03.008.

3. van Venrooij WJ, Pruijn GJM. Citrullination: a small change for a protein with great consequences for rheumatoid arthritis. *Arthritis Res.* 2000;2(4): 249–251. https://doi.org/10.1186/ar95.

4. Horibata S, Coonrod SA, Cherrington BD. Role of PADs in Disease and Female Reproduction. *Journal of Reproduction and Development*. 2012;58(3). https://doi.org/10.1262/jrd.2011-040.

5. Vossenaar ER, Zendman AJW, Van Venrooij WJ, Pruijn GJM. PAD, a growing family of citrullinating enzymes: Genes, features and involvement in disease. *BioEssays*. 2003;25(11): 1106–1118. https://doi.org/10.1002/bies.10357.

6. Deplus R, Denis H, Putmans P, Calonne E, Fourrez M, Yamamoto K, et al. Citrullination of DNMT3A by PADI4 regulates its stability and controls DNA methylation. *Nucleic Acids Research*. 2014;42(13): 8285–8296. https://doi.org/10.1093/nar/gku522.

7. Chavanas S, Méchin MC, Takahara H, Kawada A, Nachat R, Serre G, et al. Comparative analysis of the mouse and human peptidylarginine deiminase gene clusters reveals highly conserved non-coding segments and a new human gene, PADI6. *Gene*. 2004;330(1–2): 19–27. https://doi.org/10.1016/j.gene.2003.12.038.

8. Mondal S, Thompson PR. Protein Arginine Deiminases (PADs): Biochemistry and Chemical Biology of Protein Citrullination. *Accounts of Chemical Research*. 2019;52(3): 818–832. https://doi.org/10.1021/acs.accounts.9b00024.

9. Guerrin M, Ishigami A, Méchin MC, Nachat R, Valmary S, Sebbag M, et al. cDNA cloning, gene organization and expression analysis of human peptidylarginine deiminase type I. *Biochemical Journal*. 2003;370(1): 167–174. https://doi.org/10.1042/bj20020870.

10. Zhang J, Dai J, Zhao E, Lin Y, Zeng L, Chen J, et al. cDNA cloning, gene organization and expression analysis of human peptidylarginine deiminase type VI. *Acta Biochimica Polonica*. 2004;51(4): 1051–1058. https://doi.org/0451041051.

11. Wright PW, Bolling LC, Calvert ME, Sarmento OF, Berkeley EV, Shea MC, et al. ePAD, an oocyte and early embryo-abundant peptidylarginine deiminase-like protein that localizes to egg cytoplasmic sheets. *Developmental Biology*. 2003;256(1): 74–89. https://doi.org/10.1016/S0012-1606(02)00126-4.

12. Cummings TFM, Gori K, Sanchez-Pulido L, Gavriilidis G, Moi D, Wilson AR, et al. Citrullination Was Introduced into Animals by Horizontal Gene Transfer from Cyanobacteria. O'Connell M (ed.) *Molecular Biology and Evolution*. 2022;39(2): msab317. https://doi.org/10.1093/molbev/msab317.

13. Bicker KL, Thompson PR. The protein arginine deiminases: Structure, function, inhibition, and disease. *Biopolymers*. 2013;99(2): 155–163. https://doi.org/10.1002/bip.22127.

14. Beato M, Sharma P. Peptidyl Arginine Deiminase 2 (PADI2)-Mediated Arginine Citrullination Modulates Transcription in Cancer. *International Journal of Molecular Sciences*. 2020;21(4): 1351. https://doi.org/10.3390/ijms21041351.

15. Chang X, Xia Y, Pan J, Meng Q, Zhao Y, Yan X. PADI2 Is Significantly Associated with Rheumatoid Arthritis. Chopra A (ed.) *PLoS ONE*. 2013;8(12): e81259. https://doi.org/10.1371/journal.pone.0081259.

16. Darrah E, Giles JT, Ols ML, Bull HG, Andrade F, Rosen A. Erosive Rheumatoid Arthritis Is Associated with Antibodies That Activate PAD4 by Increasing Calcium Sensitivity. *Science Translational Medicine*. 2013;5(186). https://doi.org/10.1126/scitranslmed.3005370.

17. Kearney PL, Bhatia M, Jones NG, Yuan L, Glascock MC, Catchings KL, et al. Kinetic characterization of protein arginine deiminase 4: A transcriptional corepressor implicated in the onset and progression of rheumatoid arthritis. *Biochemistry*. 2005;44(31): 10570–10582. https://doi.org/10.1021/bi050292m.

18. Wu X, Wang Y, Wang W. The Role of Peptidyl Arginine Deiminase IV(PADI4) in Cancers. *Anti-Cancer Agents in Medicinal Chemistry*. 2023;23(3): 256–265. https://doi.org/10.2174/1871520622666220614115309.

19. Nachat R, Méchin MC, Takahara H, Chavanas S, Charveron M, Serre G, et al. Peptidylarginine Deiminase Isoforms 1–3 Are Expressed in the Epidermis and Involved in the Deimination of K1 and Filaggrin. *Journal of Investigative Dermatology*. 2005;124(2): 384–393. https://doi.org/10.1111/j.0022-202X.2004.23568.x.

20. Williams JPC, Walport LJ. PADI6: What we know about the elusive fifth member of the peptidyl arginine deiminase family. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2023;378(1890): 20220242. https://doi.org/10.1098/rstb.2022.0242.

21. Zheng L, Nagar M, Maurais AJ, Slade DJ, Parelkar SS, Coonrod SA, et al. Calcium Regulates the Nuclear Localization of Protein Arginine Deiminase 2. *Biochemistry*. 2019;58(27): 3042–3056. https://doi.org/10.1021/acs.biochem.9b00225.

22. Neira JL, Rizzuti B, Abián O, Araujo-Abad S, Velázquez-Campoy A, De Juan Romero C. Human Enzyme PADI4 Binds to the Nuclear Carrier Importin α3. *Cells*. 2022;11(14): 2166. https://doi.org/10.3390/cells11142166.

23. Zhang X, Bolt M, Guertin MJ, Chen W, Zhang S, Cherrington BD, et al. Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination

facilitates estrogen receptor α target gene activation. *Proceedings of the National Academy of Sciences*. 2012;109(33): 13331–13336. https://doi.org/10.1073/pnas.1203280109.

24. Christophorou MA, Castelo-Branco G, Halley-Stott RP, Oliveira CS, Loos R, Radzisheuskaya A, et al. Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature*. 2014;507(7490): 104–108. https://doi.org/10.1038/nature12942.

25. Sharma P, Lioutas A, Fernandez-Fuentes N, Quilez J, Carbonell-Caballero J, Wright RHG, et al. Arginine Citrullination at the C-Terminal Domain Controls RNA Polymerase II Transcription. *Molecular Cell*. 2019;73(1): 84-96.e7. https://doi.org/10.1016/j.molcel.2018.10.016.

26. Villanueva-Cañas JL, Fernandez-Fuentes N, Teyssier C, Rogalska ME, Pérez FP, Oliva B, et al. *Evolutionary analysis reveals the role of a non-catalytic domain of peptidyl arginine deiminase 2 in transcriptional regulation*. 2022. https://doi.org/10.1101/2022.09.19.508513.

27. Slade DJ, Fang P, Dreyton CJ, Zhang Y, Fuhrmann J, Rempel D, et al. Protein arginine deiminase 2 binds calcium in an ordered fashion: Implications for inhibitor design. *ACS Chemical Biology*. 2015;10(4): 1043–1053. https://doi.org/10.1021/cb500933j.

28. Arita K, Hashimoto H, Shimizu T, Nakashima K, Yamada M, Sato M. Structural basis for Ca2+-induced activation of human PAD4. *Nature Structural and Molecular Biology*. 2004;11(8): 777–783. https://doi.org/10.1038/nsmb799.

29. Liu YL, Chiang YH, Liu GY, Hung HC. Functional Role of Dimerization of Human Peptidylarginine Deiminase 4 (PAD4). Proost P (ed.) *PLoS ONE*. 2011;6(6): e21314. https://doi.org/10.1371/journal.pone.0021314.

30. Dreyton CJ, Knuckley B, Jones JE, Lewallen DM, Thompson PR. Mechanistic Studies of Protein Arginine Deiminase 2: Evidence for a Substrate-Assisted Mechanism. *Biochemistry*. 2014;53(27): 4426–4433. https://doi.org/10.1021/bi500554b.

31. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*. 2011;7(1): 539. https://doi.org/10.1038/msb.2011.75.

32. Esposito G, Vitale AM, Leijten FPJ, Strik AM, Koonen-Reemst AMCB, Yurttas P, et al. Peptidylarginine deiminase (PAD) 6 is essential for oocyte cytoskeletal sheet formation and female fertility. *Molecular and Cellular Endocrinology*. 2007;273(1–2): 25–31. https://doi.org/10.1016/j.mce.2007.05.005.

33. Liu J, Tan Z, He J, Jin T, Han Y, Hu L, et al. Two novel mutations in PADI6 and TLE6 genes cause female infertility due to arrest in embryonic development. *Journal of Assisted Reproduction and Genetics*. 2021;38(6): 1551–1559. https://doi.org/10.1007/s10815-021-02194-1.

34. Dong J, Fu J, Yan Z, Li L, Qiu Y, Zeng Y, et al. Novel biallelic mutations in PADI6 in patients with early embryonic arrest. *Journal of Human Genetics*. 2022;(September): 1–9. https://doi.org/10.1038/s10038-021-00998-8.

35. Qian J, Nguyen NMP, Rezaei M, Huang B, Tao Y, Zhang X, et al. Biallelic PADI6 variants linking infertility, miscarriages, and hydatidiform moles. *European Journal of Human Genetics*. 2018;26(7): 1007–1013. https://doi.org/10.1038/s41431-018-0141-3.

36. Zheng W, Chen L, Dai J, Dai C, Guo J, Lu C, et al. New biallelic mutations in PADI6 cause recurrent preimplantation embryonic arrest characterized by direct cleavage. *Journal of Assisted Reproduction and Genetics*. 2020;37(1): 205–212. https://doi.org/10.1007/s10815-019-01606-7.

37. Zhang T, Liu P, Yao G, Zhang X, Cao C. A complex heterozygous mutation in PADI6 causes early embryo arrest: A case report. *Frontiers in Genetics*. 2023;13: 1104085. https://doi.org/10.3389/fgene.2022.1104085.

38. Xu Y, Wang R, Pang Z, Wei Z, Sun L, Li S, et al. Novel Homozygous PADI6 Variants in Infertile Females with Early Embryonic Arrest. *Frontiers in Cell and Developmental Biology*. 2022;10(April): 1–8. https://doi.org/10.3389/fcell.2022.819667.

39. Xu Y, Shi Y, Fu J, Yu M, Feng R, Sang Q, et al. Mutations in PADI6 Cause Female Infertility Characterized by Early Embryonic Arrest. *American Journal of Human Genetics*. 2016;99(3): 744–752. https://doi.org/10.1016/j.ajhg.2016.06.024.

40. Wang X, Zhu H, He Y, Zeng J, Zhao J, Xia Q, et al. A novel homozygous mutation in the PADI6 gene causes early embryo arrest. *Reproductive Health*. 2022;19(1): 1–11. https://doi.org/10.1186/s12978-022-01495-7.

41. Rezaei M, Suresh B, Bereke E, Hadipour Z, Aguinaga M, Qian J, et al. Novel pathogenic variants in NLRP7, NLRP5, and PADI6 in patients with recurrent hydatidiform moles and reproductive failure. *Clinical Genetics*. 2021;99(6): 823–828. https://doi.org/10.1111/cge.13941.

42. Wang X, Song D, Mykytenko D, Kuang Y, Lv Q, Li B, et al. Novel mutations in genes encoding subcortical maternal complex proteins may cause human embryonic developmental arrest. *Reproductive BioMedicine Online*. 2018;36(6): 698–704. https://doi.org/10.1016/j.rbmo.2018.03.009.

43. Clift D, Schuh M. Restarting life: Fertilization and the transition from meiosis to mitosis. *Nature Reviews Molecular Cell Biology*. 2013;14(9): 549–562. https://doi.org/10.1038/nrm3643.

44. Wassarman PM. Zona Pellucida Glycoproteins. *Journal of Biological Chemistry*. 2008;283(36): 24285–24289. https://doi.org/10.1074/jbc.R800027200.

45. Wassarman PM, Litscher ES. The Mouse Egg's Zona Pellucida. In: *Current Topics in Developmental Biology*. Elsevier; 2018. p. 331–356. https://doi.org/10.1016/bs.ctdb.2018.01.003. [Accessed 20th November 2023].

46. Wassarman PM, Jovine L, Litscher ES. Mouse zona pellucida genes and glycoproteins. *Cytogenetic and Genome Research*. 2004;105(2–4): 228–234. https://doi.org/10.1159/000078193.

47. Eppig JJ. Oocyte control of ovarian follicular development and function in mammals. *Reproduction*. 2001;122: 829–838.

48. Buccione R, Schroeder AC, Eppig JJ. Interactions between Somatic Cells and Germ Cells Throughout Mammalian Oogenesis1. *Biology of Reproduction*. 1990;43(4): 543–547. https://doi.org/10.1095/biolreprod43.4.543.

49. Nakao S, Ito K, Sugahara C, Watanabe H, Kondoh G, Nakagata N, et al. Synchronization of the ovulation and copulation timings increased the number of in vivo fertilized oocytes in superovulated female mice. Hansen PJ (ed.) *PLOS ONE*. 2023;18(2): e0281330. https://doi.org/10.1371/journal.pone.0281330.

50. Ajayi AF, Akhigbe RE. Staging of the estrous cycle and induction of estrus in experimental rodents: an update. *Fertility Research and Practice*. 2020;6(1): 5. https://doi.org/10.1186/s40738-020-00074-3.

51. *Manipulating the mouse embryo: a laboratory manual*.. 4th ed. Cold Spring Harbor (N.Y.): Cold Spring Harbor laboratory press; 2014.

52. Li L, Zheng P, Dean J. Maternal control of early mouse development. *Development*. 2010;137(6): 859–870. https://doi.org/10.1242/dev.039487.

53. Mitchell LE. Maternal effect genes: Update and review of evidence for a link with birth defects. *Human Genetics and Genomics Advances*. 2022;3(1): 100067. https://doi.org/10.1016/j.xhgg.2021.100067.

54. Christians E, Davis AA, Thomas SD, Benjamin IJ. Maternal effect of Hsf1 on reproductive success. *Nature*. 2000;407(6805): 693–694. https://doi.org/10.1038/35037669.

55. Tong ZB, Gold L, Pfeifer KE, Dorward H, Lee E, Bondy CA, et al. Mater, a maternal effect gene required for early embryonic development in mice. *Nature Genetics*. 2000;26(3): 267–268. https://doi.org/10.1038/81547.

56. Gurtu VE, Verma S, Grossmann AH, Liskay RM, Skarnes WC, Baker SM. Maternal Effect for DNA Mismatch Repair in the Mouse. *Genetics*. 2002;160(1): 271–277. https://doi.org/10.1093/genetics/160.1.271.

57. Burns KH, Viveiros MM, Ren Y, Wang P, DeMayo FJ, Frail DE, et al. Roles of NPM2 in Chromatin and Nucleolar Organization in Oocytes and Embryos. *Science*. 2003;300(5619): 633–636. https://doi.org/10.1126/science.1081813.

58. Payer B, Saitou M, Barton SC, Thresher R, Dixon JPC, Zahn D, et al. stella Is a Maternal Effect Gene Required for Normal Early Development in Mice. *Current Biology*. 2003;13(23): 2110–2117. https://doi.org/10.1016/j.cub.2003.11.026.

59. Wu X, Viveiros MM, Eppig JJ, Bai Y, Fitzpatrick SL, Matzuk MM. Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. *Nature Genetics*. 2003;33(2): 187–191. https://doi.org/10.1038/ng1079.

60. Roest HP, Baarends WM, De Wit J, Van Klaveren JW, Wassenaar E, Hoogerbrugge JW, et al. The Ubiquitin-Conjugating DNA Repair Enzyme HR6A Is a Maternal Factor Essential for Early Embryonic Development in Mice. *Molecular and Cellular Biology*. 2004;24(12): 5485–5495. https://doi.org/10.1128/MCB.24.12.5485-5495.2004.

61. Gao Y, Liu X, Tang B, Li C, Kou Z, Li L, et al. Protein Expression Landscape of Mouse Embryos during Pre-implantation Development. *Cell Reports*. 2017;21(13): 3957–3969. https://doi.org/10.1016/j.celrep.2017.11.111.

62. Madgwick S, Jones KT. How eggs arrest at metaphase II: MPF stabilisation plus APC/C inhibition equals cytostatic factor. *Cell Division*. 2007;2. https://doi.org/10.1186/1747-1028-2-4.

63. Biswas A, Ng BH, Prabhakaran VS, Chan CJ. Squeezing the eggs to grow: The mechanobiology of mammalian folliculogenesis. *Frontiers in Cell and Developmental Biology*. 2022;10: 1038107. https://doi.org/10.3389/fcell.2022.1038107.

64. McGaughey RW, Capco DG. Specialized cytoskeletal elements in mammalian eggs: Structural and biochemical evidence for their composition. *Cell Motility and the Cytoskeleton*. 1989;13(2): 104–111. https://doi.org/10.1002/cm.970130205.

65. Ian Gallicano G, Larabell CA, McGaughey RW, Capco DG. Novel cytoskeletal elements in mammalian eggs are composed of a unique arrangement of intermediate filaments. *Mechanisms of Development*. 1994;45(3): 211–226. https://doi.org/10.1016/0925-4773(94)90009-4.

66. Capco DG, Gallicano GI, McGaughey RW, Downing KH, Larabell CA. Cytoskeletal sheets of mammalian eggs and embryos: A lattice-like network of intermediate filaments. *Cell Motility and the Cytoskeleton*. 1993;24(2): 85–99. https://doi.org/10.1002/cm.970240202.

67. Jentoft IMA, Bäuerlein FJB, Welp LM, Cooper BH, Petrovic A, So C, et al. Mammalian oocytes store proteins for the early embryo on cytoplasmic lattices. *Cell*. 2023; S0092867423010851. https://doi.org/10.1016/j.cell.2023.10.003.

68. Yurttas P, Vitale AM, Fitzhenry RJ, Cohen-Gould L, Wu W, Gossen JA, et al. Role for PADI6 and the cytoplasmic lattices in ribosomal storage in oocytes and translational control in the early mouse embryo. *Development*. 2008;135(15): 2627–2636. https://doi.org/10.1242/dev.016329.

69. Sternlicht AL, Schultz RM. Biochemical studies of mammalian oogenesis: Kinetics of accumulation of total and poly(A)-containing RNA during growth of the mouse oocyte. *Journal of Experimental Zoology*. 1981;215(2): 191–200. https://doi.org/10.1002/jez.1402150209.

70. Kan R, Yurttas P, Kim B, Jin M, Wo L, Lee B, et al. Regulation of mouse oocyte microtubule and organelle dynamics by PADI6 and the cytoplasmic lattices. *Developmental Biology*. 2011;350(2): 311–322. https://doi.org/10.1016/j.ydbio.2010.11.033.

71. Van Blerkom J, Runner MN. Mitochondrial reorganization during resumption of arrested meiosis in the mouse oocyte. *American Journal of Anatomy*. 1984;171(3): 335–355. https://doi.org/10.1002/aja.1001710309.

72. FitzHarris G, Marangos P, Carroll J. Changes in endoplasmic reticulum structure during mouse oocyte maturation are controlled by the cytoskeleton and cytoplasmic dynein. *Developmental Biology*. 2007;305(1): 133–144. https://doi.org/10.1016/j.ydbio.2007.02.006.

73. Friedman JR, Webster BM, Mastronarde DN, Verhey KJ, Voeltz GK. ER sliding dynamics and ER–mitochondrial contacts occur on acetylated microtubules. *Journal of Cell Biology*. 2010;190(3): 363–375. https://doi.org/10.1083/jcb.200911024.

74. Chen S, Owens GC, Makarenkova H, Edelman DB. HDAC6 Regulates Mitochondrial Transport in Hippocampal Neurons. Combs C (ed.) *PLoS ONE*. 2010;5(5): e10848. https://doi.org/10.1371/journal.pone.0010848.

75. Liu X, Morency E, Li T, Qin H, Zhang X, Zhang X, et al. Role for PADI6 in securing the mRNA-MSY2 complex to the oocyte cytoplasmic lattices. *Cell Cycle*. 2017;16(4): 360–366. https://doi.org/10.1080/15384101.2016.1261225.

76. Yu J, Deng M, Medvedev S, Yang J, Hecht NB, Schultz RM. Transgenic RNAi-mediated reduction of MSY2 in mouse oocytes results in reduced fertility. *Developmental Biology*. 2004;268(1): 195–206. https://doi.org/10.1016/j.ydbio.2003.12.020.

77. Yu J, Hecht NB, Schultz RM. Requirement for RNA-binding activity of MSY2 for cytoplasmic localization and retention in mouse oocytes. *Developmental Biology*. 2003;255(2): 249–262. https://doi.org/10.1016/S0012-1606(02)00094-5.

78. Yu J, Hecht NB, Schultz RM. RNA-Binding Properties and Translation Repression In Vitro by Germ Cell-Specific MSY2 Protein1. *Biology of Reproduction*. 2002;67(4): 1093–1098. https://doi.org/10.1095/biolreprod67.4.1093.

79. Zhu K, Yan L, Zhang X, Lu X, Wang T, Yan J, et al. Identification of a human subcortical maternal complex. *Molecular Human Reproduction*. 2014;21(4): 320–329. https://doi.org/10.1093/molehr/gau116.

80. Lu X, Gao Z, Qin D, Li L. A Maternal Functional Module in the Mammalian Oocyte-To-Embryo Transition. *Trends in Molecular Medicine*. 2017;23(11): 1014–1023. https://doi.org/10.1016/j.molmed.2017.09.004.

81. Mahadevan S, Sathappan V, Utama B, Lorenzo I, Kaskar K, Van Den Veyver IB. Maternally expressed NLRP2 links the subcortical maternal complex (SCMC) to fertility, embryogenesis and epigenetic reprogramming. *Scientific Reports*. 2017;7(November 2016). https://doi.org/10.1038/srep44667.

82. Gao Z, Zhang X, Yu X, Qin D, Xiao Y, Yu Y, et al. Zbed3 participates in the subcortical maternal complex and regulates the distribution of organelles. *Journal of Molecular Cell Biology*. 2018;10(1): 74–88. https://doi.org/10.1093/jmcb/mjx035.

83. Li L, Baibakov B, Dean J. A Subcortical Maternal Complex Essential for Preimplantation Mouse Embryogenesis. *Developmental Cell*. 2008;15(3): 416–425. https://doi.org/10.1016/j.devcel.2008.07.010.

84. Tashiro F, Kanai-Azuma M, Miyazaki S, Kato M, Tanaka T, Toyoda S, et al. Maternal-effect gene Ces5/Ooep/Moep19/Floped is essential for oocyte cytoplasmic lattice formation and embryonic development at the maternal-zygotic stage transition. *Genes to Cells*. 2010;15(8): 813–828. https://doi.org/10.1111/j.1365-2443.2010.01420.x.

85. Kim B, Kan R, Anguish L, Nelson LM, Coonrod SA. Potential role for MATER in cytoplasmic lattice formation in murine oocytes. *PLoS ONE*. 2010;5(9): 1–8. https://doi.org/10.1371/journal.pone.0012587.

86. Qin D, Gao Z, Xiao Y, Zhang X, Ma H, Yu X, et al. The subcortical maternal complex protein Nlrp4f is involved in cytoplasmic lattice formation and organelle distribution. *Development*. 2019;146: dev.183616. https://doi.org/10.1242/dev.183616.

87. Yu XJ, Yi Z, Gao Z, Qin D, Zhai Y, Chen X, et al. The subcortical maternal complex controls symmetric division of mouse zygotes by regulating F-actin dynamics. *Nature communications*. 2014;5: 4887. https://doi.org/10.1038/ncomms5887.

88. Zheng P, Dean J. Role of Filia, a maternal effect gene, in maintaining euplo i dy dur i ng cleavage-stage mouse embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106(18): 7473–7478. https://doi.org/10.1073/pnas.0900519106.

89. Han J, Zhang N, Cao Q, Shi X, Wang C, Rui X, et al. NLRP7 participates in the human subcortical maternal complex and its variants cause female infertility characterized by early embryo arrest. *Journal of Molecular Medicine*. 2023; https://doi.org/10.1007/s00109-023-02322-7.

90. Tong X, Jin J, Hu Z, Zhang Y, Fan HY, Zhang YL, et al. Mutations in OOEP and NLRP5 identified in infertile patients with early embryonic arrest. *Human Mutation*. 2022;(July). https://doi.org/10.1002/humu.24448.

91. Cozette C, Scheffler F, Lombart M, Massardier J, Bolze PA, Hajri T, et al. Pregnancy after oocyte donation in a patient with NLRP7 gene mutations and recurrent molar hydatidiform pregnancies. *Journal of Assisted Reproduction and Genetics*. 2020;37(9): 2273–2277. https://doi.org/10.1007/s10815-020-01861-z.

92. Uemura S, Maenohara S, Inoue K, Ogonuki N, Matoba S, Ogura A, et al. UHRF1 is essential for proper cytoplasmic architecture and function of mouse oocytes and derived embryos. *Life Science Alliance*. 2023;6(8): e202301904. https://doi.org/10.26508/lsa.202301904.

93. Bostick M, Kim JK, Estève PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*. 2007;317(5845): 1760–1764. https://doi.org/10.1126/science.1147939.

94. Wassarman PM, Jovine L, Litscher ES. A profile of fertilization in mammals. *Nature Cell Biology*. 2001;3(2): E59–E64. https://doi.org/10.1038/35055178.

95. Miyazaki S, Shirakawa H, Nakada K, Honda Y. Essential Role of the Inositol 1,4,5-Trisphosphate Receptor/Ca2+ Release Channel in Ca2+ Waves and Ca2+ Oscillations at Fertilization of Mammalian Eggs. *Developmental Biology*. 1993;158(1): 62–78. https://doi.org/10.1006/dbio.1993.1168.

96. Ducibella T, Huneau D, Angelichio E, Xu Z, Schultz RM, Kopf GS, et al. Egg-to-Embryo Transition Is Driven by Differential Responses to Ca2+ Oscillation Number. *Developmental Biology*. 2002;250(2): 280–291. https://doi.org/10.1006/dbio.2002.0788.

97. Horner VL, Wolfner MF. Transitioning from egg to embryo: Triggers and mechanisms of egg activation. *Developmental Dynamics*. 2008;237(3): 527–544. https://doi.org/10.1002/dvdy.21454.

98. Saunders CM, Larman MG, Parrington J, Cox LJ, Royse J, Blayney LM, et al. PLCζ: a sperm-specific trigger of Ca2+ oscillations in eggs and embryo development. *Development*. 2002;129(15): 3533–3544. https://doi.org/10.1242/dev.129.15.3533.

99. Rauh NR, Schmidt A, Bormann J, Nigg EA, Mayer TU. Calcium triggers exit from meiosis II by targeting the APC/C inhibitor XErp1 for degradation. *Nature*. 2005;437(7061): 1048–1052. https://doi.org/10.1038/nature04093.

100. Torres-Padilla ME, Bannister AJ, Hurd PJ, Kouzarides T, Zernicka-Goetz M. Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. *The International Journal of Developmental Biology*. 2006;50(Next). https://doi.org/10.1387/ijdb.052073mt.

101. Sassone-Corsi P. Unique Chromatin Remodeling and Transcriptional Regulation in Spermatogenesis. *Science*. 2002;296(5576): 2176–2178. https://doi.org/10.1126/science.1070963.

102. Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, et al. Active demethylation of the paternal genome in the mouse zygote. *Current Biology*. 2000;10(8): 475–478. https://doi.org/10.1016/S0960-9822(00)00448-6.

103. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. *Nature*. 2000;403(6769): 501–502. https://doi.org/10.1038/35000656.

104. Tachibana-Konwalski K, Godwin J, Van Der Weyden L, Champion L, Kudo NR, Adams DJ, et al. Rec8-containing cohesin maintains bivalents without turnover during the growing phase of mouse oocytes. *Genes & Development*. 2010;24(22): 2505–2516. https://doi.org/10.1101/gad.605910.

105. Watanabe Y, Nurse P. Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature*. 1999;400(6743): 461–464. https://doi.org/10.1038/22774.

106. Schuh M, Ellenberg J. Self-Organization of MTOCs Replaces Centrosome Function during Acentrosomal Spindle Assembly in Live Mouse Oocytes. *Cell*. 2007;130(3): 484–498. https://doi.org/10.1016/j.cell.2007.06.025.

107. So C, Seres KB, Steyer AM, Mönnich E, Clift D, Pejkovska A, et al. A liquid-like spindle domain promotes acentrosomal spindle assembly in mammalian oocytes. *Science*. 2019;364(6447). https://doi.org/10.1126/science.aat9557.

108. Chaigne A, Verlhac MH, Terret ME. Spindle positioning in mammalian oocytes. *Experimental Cell Research*. 2012;318(12): 1442–1447. https://doi.org/10.1016/j.yexcr.2012.02.019.

109. Zhang JM, Hou WB, Du JW, Zong M, Zheng KL, Wang WJ, et al. Argonaute 2 is a key regulator of maternal mRNA degradation in mouse early embryos. *Cell Death Discovery*. 2020;6(1). https://doi.org/10.1038/s41420-020-00368-x.

110. Sha QQ, Zheng W, Wu YW, Li S, Guo L, Zhang S, et al. Dynamics and clinical relevance of maternal mRNA clearance during the oocyte-to-embryo transition in humans. *Nature Communications*. 2020;11(1): 4917. https://doi.org/10.1038/s41467-020-18680-6.

111. Jukam D, Shariati SAM, Skotheim JM. Zygotic Genome Activation in Vertebrates. *Developmental Cell*. 2017;42(4): 316–332. https://doi.org/10.1016/j.devcel.2017.07.026.

112. Lee MT, Bonneau AR, Giraldez AJ. Zygotic Genome Activation During the Maternal-to-Zygotic Transition. *Annual Review of Cell and Developmental Biology*. 2014;30(1): 581–613. https://doi.org/10.1146/annurev-cellbio-100913-013027.

113. Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. *Nature Reviews Genetics*. 2019;20(4): 221–234. https://doi.org/10.1038/s41576-018-0087-x.

114. Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. *Development*. 2009;136(18): 3033–3042. https://doi.org/10.1242/dev.033183.

115. Dahl JA, Jung I, Aanes H, Greggains GD, Manaf A, Lerdrup M, et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature*. 2016;537(7621): 548–552. https://doi.org/10.1038/nature19360.

116. Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*. 2016;534(7609): 652–657. https://doi.org/10.1038/nature18606.

117. Lu F, Liu Y, Inoue A, Suzuki T, Zhao K, Zhang Y. Establishing chromatin regulatory landscape during mouse preimplantation development. *Cell*. 2016;165(6): 1375–1388. https://doi.org/10.1016/j.cell.2016.05.050.

118. Liu X, Wang C, Liu W, Li J, Li C, Kou X, et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature*. 2016;537(7621): 558–562. https://doi.org/10.1038/nature19362.

119. Zhang B, Zheng H, Huang B, Li W, Xiang Y, Peng X, et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature*. 2016;537(7621): 553–557. https://doi.org/10.1038/nature19361.

120. Gaume X, Torres-Padilla ME. Regulation of Reprogramming and Cellular Plasticity through Histone Exchange and Histone Variant Incorporation. *Cold Spring Harbor Symposia on Quantitative Biology*. 2015;80: 165–175. https://doi.org/10.1101/sqb.2015.80.027458.

121. Li L, Guo F, Gao Y, Ren Y, Yuan P, Yan L, et al. Single-cell multi-omics sequencing of human early embryos. *Nature Cell Biology*. 2018;20(7): 847–858. https://doi.org/10.1038/s41556-018-0123-2.

122. Wu J, Xu J, Liu B, Yao G, Wang P, Lin Z, et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature*. 2018;557(7704): 256–260. https://doi.org/10.1038/s41586-018-0080-8.

123. Le Bin GC, Muñoz-Descalzo S, Kurowski A, Leitch H, Lou X, Mansfield W, et al. Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. *Development*. 2014;141(5): 1001–1010. https://doi.org/10.1242/dev.096875.

124. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature Genetics*. 2017;49(6): 941–945. https://doi.org/10.1038/ng.3858.

125. Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nature Genetics*. 2017;49(6): 925–934. https://doi.org/10.1038/ng.3844.

126. Whiddon JL, Langford AT, Wong CJ, Zhong JW, Tapscott SJ. Conservation and innovation in the DUX4-family gene network. *Nature Genetics*. 2017;49(6): 935–940. https://doi.org/10.1038/ng.3846.

127. Eckersley-Maslin MA, Alda-Catalinas C, Reik W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nature Reviews Molecular Cell Biology*. 2018;19(7): 436–450. https://doi.org/10.1038/s41580-018-0008-z.

128. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Research*. 2011;21(3): 381–395. https://doi.org/10.1038/cr.2011.22.

129. Zhang B, Zheng H, Huang B, Li W, Xiang Y, Peng X, et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature*. 2016;537(7621): 553–557. https://doi.org/10.1038/nature19361.

130. Chan SH, Tang Y, Miao L, Darwich-Codore H, Vejnar CE, Beaudoin JD, et al. Brd4 and P300 Confer Transcriptional Competency during Zygotic Genome Activation. *Developmental Cell*. 2019;49(6): 867-881.e8. https://doi.org/10.1016/j.devcel.2019.05.037.

131. Eissenberg JC, Shilatifard A. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental Biology*. 2010;339(2): 240–249. https://doi.org/10.1016/j.ydbio.2009.08.017.

132. Xia W, Xu J, Yu G, Yao G, Xu K, Ma X, et al. Resetting histone modifications during human parental-to-zygotic transition. *Science*. 2019;365(6451): 353–360. https://doi.org/10.1126/science.aaw5118.

133. Weber CM, Henikoff S. Histone variants: dynamic punctuation in transcription. *Genes & Development*. 2014;28(7): 672–682. https://doi.org/10.1101/gad.238873.114.

134. Fu G, Ghadam P, Sirotkin A, Khochbin S, Skoultchi AI, Clarke HJ. Mouse Oocytes and Early Embryos Express Multiple Histone H1 Subtypes1. *Biology of Reproduction*. 2003;68(5): 1569–1576. https://doi.org/10.1095/biolreprod.102.012336.

135. Pérez-Montero S, Carbonell A, Morán T, Vaquero A, Azorín F. The Embryonic Linker Histone H1 Variant of Drosophila, dBigH1, Regulates Zygotic Genome Activation. *Developmental Cell*. 2013;26(6): 578–590. https://doi.org/10.1016/j.devcel.2013.08.011.

136. Gao L, Wu K, Liu Z, Yao X, Yuan S, Tao W, et al. Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. *Cell*. 2018;173(1): 248-259.e15. https://doi.org/10.1016/j.cell.2018.02.028.

137. Fogarty NME, McCarthy A, Snijders KE, Powell BE, Kubikova N, Blakeley P, et al. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature*. 2017;550(7674): 67–73. https://doi.org/10.1038/nature24033.

138. Sakashita A, Kitano T, Ishizu H, Guo Y, Masuda H, Ariura M, et al. Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nature Genetics*. 2023;55(3): 484–495. https://doi.org/10.1038/s41588-023-01324-y.

139. Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. *Genes & Development*. 2006;20(21): 2922–2936. https://doi.org/10.1101/gad.1477006.

140. Schultz RM, Davis W, Stein P, Svoboda P. Reprogramming of gene expression during preimplantation development. *Journal of Experimental Zoology*. 1999;285(3): 276–282. https://doi.org/10.1002/(SICI)1097-010X(19991015)285:3<276::AID-JEZ11>3.0.CO;2-K.

141. Gerri C, Menchero S, Mahadevaiah SK, Turner JMA, Niakan KK. Human Embryogenesis: A Comparative Perspective. *Annual Review of Cell and Developmental Biology*. 2020;36: 411–440. https://doi.org/10.1146/annurev-cellbio-022020-024900.

142. Gerri C, McCarthy A, Alanis-Lobato G, Demtschenko A, Bruneau A, Loubersac S, et al. Initiation of a conserved trophectoderm program in human, cow and mouse embryos. *Nature*. 2020;587(7834): 443–447. https://doi.org/10.1038/s41586-020-2759-x.

143. Huang B, Zhao Y, Zhou L, Gong T, Feng J, Han P, et al. PADI6 Regulates Trophoblast Cell Migration-Invasion Through the Hippo/YAP1 Pathway in Hydatidiform Moles. *Journal of Inflammation Research*. 2021;14(July): 3489–3500. https://doi.org/10.2147/jir.s313422.

144. Slim R, Mehio A. The genetics of hydatidiform moles: new lights on an ancient disease. *Clinical Genetics*. 2007;71(1): 25–34. https://doi.org/10.1111/j.1399-0004.2006.00697.x.

145. Zhang C, Wang M, Li Y, Zhang Y. Profiling and functional characterization of maternal mRNA translation during mouse maternal-to-zygotic transition. *Science advances*. 2022;8(5): eabj3967. https://doi.org/10.1126/sciadv.abj3967.

146. Zhang X, Liu X, Zhang M, Li T, Muth A, Thompson PR, et al. Peptidylarginine deiminase 1-catalyzed histone citrullination is essential for early embryo development. *Scientific Reports*. 2016;6(November): 1–11. https://doi.org/10.1038/srep38727.

147. Clancy KW, Weerapana E, Thompson PR. Detection and identification of protein citrullination in complex biological systems. *Current Opinion in Chemical Biology*. 2016;30: 1–6. https://doi.org/10.1016/j.cbpa.2015.10.014.

148. Kan R, Jin M, Subramanian V, Causey CP, Thompson PR, Coonrod SA. Potential role for PADI-mediated histone citrullination in preimplantation development. *BMC Developmental Biology*. 2012;12. https://doi.org/10.1186/1471-213X-12-19.

149. Chumanevich AA, Causey CP, Knuckley BA, Jones JE, Poudyal D, Chumanevich AP, et al. Suppression of colitis in mice by Cl-amidine: a novel peptidylarginine deiminase inhibitor. *American Journal of Physiology-Gastrointestinal and Liver Physiology*. 2011;300(6): G929–G938. https://doi.org/10.1152/ajpgi.00435.2010.

150. Bicker KL, Anguish L, Chumanevich AA, Cameron MD, Cui X, Witalison E, et al. D -Amino Acid-Based Protein Arginine Deiminase Inhibitors: Synthesis, Pharmacokinetics, and in Cellulo Efficacy. *ACS Medicinal Chemistry Letters*. 2012;3(12): 1081–1085. https://doi.org/10.1021/ml300288d.

151. Luo Y, Knuckley B, Bhatia M, Pellechia PJ, Thompson PR. Activity-Based Protein Profiling Reagents for Protein Arginine Deiminase 4 (PAD4): Synthesis and in vitro Evaluation of a Fluorescently Labeled Probe. *Journal of the American Chemical Society*. 2006;128(45): 14468–14469. https://doi.org/10.1021/ja0656907.

152. Snow AJ, Puri P, Acker-Palmer A, Bouwmeester T, Vijayaraghavan S, Kline D. Phosphorylation-dependent interaction of tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase activation protein (YWHA) with PADI6 following oocyte maturation in mice. *Biology of Reproduction*. 2008;79(2): 337–347. https://doi.org/10.1095/biolreprod.108.069328.

153. Johnson C, Crowther S, Stafford MJ, Campbell DG, Toth R, MacKintosh C. Bioinformatic and experimental survey of 14-3-3-binding sites. *Biochemical Journal*. 2010;427(1): 69–78. https://doi.org/10.1042/BJ20091834.

154. Fu H, Subramanian RR, Masters SC. 14-3-3 Proteins: Structure, Function, and Regulation. *Annual Review of Pharmacology and Toxicology*. 2000;40(1): 617–647. https://doi.org/10.1146/annurev.pharmtox.40.1.617.

155. Yang X, Lee WH, Sobott F, Papagrigoriou E, Robinson CV, Grossmann JG, et al. Structural basis for protein–protein interactions in the 14-3-3 protein family. *Proceedings of the National Academy of Sciences*. 2006;103(46): 17237–17242. https://doi.org/10.1073/pnas.0605779103.

156. Rose R, Rose M, Ottmann C. Identification and structural characterization of two 14-3-3 binding sites in the human peptidylarginine deiminase type VI. *Journal of Structural Biology*. 2012;180(1): 65–72. https://doi.org/10.1016/j.jsb.2012.05.010.

157. Eggermann T, Kadgien G, Begemann M, Elbracht M. Biallelic PADI6 variants cause multilocus imprinting disturbances and miscarriages in the same family. *European Journal of Human Genetics*. 2021;29(4): 575–580. https://doi.org/10.1038/s41431-020-00762-0.

158. Soellner L, Begemann M, Mackay DJG, Grønskov K, Tümer Z, Maher ER, et al. Recent Advances in Imprinting Disorders. *Clinical Genetics*. 2017;91(1): 3–13. https://doi.org/10.1111/cge.12827.

159. Eggermann T, Yapici E, Bliek J, Pereda A, Begemann M, Russo S, et al. Trans-acting genetic variants causing multilocus imprinting disturbance (MLID): common mechanisms and consequences. *Clinical Epigenetics*. 2022;14(1): 1–17. https://doi.org/10.1186/s13148-022-01259-x.

160. Monk D, Mackay DJG, Eggermann T, Maher ER, Riccio A. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nature Reviews Genetics*. 2019;20(4): 235–248. https://doi.org/10.1038/s41576-018-0092-0.

161. Begemann M, Rezwan FI, Beygo J, Docherty LE, Kolarova J, Schroeder C, et al. Maternal variants in NLRP and other maternal effect proteins are associated with multilocus imprinting disturbance in offspring. *Journal of Medical Genetics*. 2018;55(7): 497–504. https://doi.org/10.1136/jmedgenet-2017-105190.

162. Cubellis MV, Pignata L, Verma A, Sparago A, Del Prete R, Monticelli M, et al. Loss-of-function maternal-effect mutations of PADI6 are associated with familial and sporadic Beckwith-Wiedemann syndrome with multi-locus imprinting disturbance. *Clinical Epigenetics*. 2020;12(1): 139. https://doi.org/10.1186/s13148-020-00925-2.

163. Tannorella P, Calzari L, Daolio C, Mainini E, Vimercati A, Gentilini D, et al. Germline variants in genes of the subcortical maternal complex and Multilocus Imprinting Disturbance are associated with miscarriage / infertility or Beckwith – Wiedemann progeny. *Clinical Epigenetics*. 2022;14(43): 1–7. https://doi.org/10.1186/s13148-022-01262-2.

164. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*. 2019;20(10): 590–607. https://doi.org/10.1038/s41580-019-0159-6.

165. Schübeler D. Function and information content of DNA methylation. *Nature*. 2015;517(7534): 321–326. https://doi.org/10.1038/nature14192.

166. Panning B. X-chromosome inactivation: the molecular basis of silencing. *Journal of Biology*. 2008;7(8): 30. https://doi.org/10.1186/jbiol95.

167. Ferguson-Smith AC. Genomic imprinting: The emergence of an epigenetic paradigm. *Nature Reviews Genetics*. 2011;12(8): 565–575. https://doi.org/10.1038/nrg3032.

168. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, et al. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*. 2004;429(6994): 900–903. https://doi.org/10.1038/nature02633.

169. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. 2017;356(6337): eaaj2239. https://doi.org/10.1126/science.aaj2239.

170. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*. 1998;393(6683): 386–389. https://doi.org/10.1038/30764.

171. Chen Z, Chen Z, Chen Z, Zhang Y, Zhang Y, Zhang Y, et al. Role of Mammalian DNA Methyltransferases in Development. *Annual Review of Biochemistry*. 2020;89: 135–158. https://doi.org/10.1146/annurev-biochem-103019-102815.

172. Li E, Zhang Y. DNA Methylation in Mammals. *Cold Spring Harbor Perspectives in Biology*. 2014;6(5): a019133–a019133. https://doi.org/10.1101/cshperspect.a019133.

173. Smith ZD, Meissner A. DNA methylation: Roles in mammalian development. *Nature Reviews Genetics*. 2013;14(3): 204–220. https://doi.org/10.1038/nrg3354.

174. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes & Development*. 2011;25(10): 1010–1022. https://doi.org/10.1101/gad.2037511.

175. Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*. 2018;19(2): 81–92. https://doi.org/10.1038/nrg.2017.80.

176. Chédin F, Lieber MR, Hsieh CL. The DNA methyltransferase-like protein DNMT3L stimulates *de novo* methylation by Dnmt3a. *Proceedings of the National Academy of Sciences*. 2002;99(26): 16916–16921. https://doi.org/10.1073/pnas.262443999.

177. Okano M, Bell DW, Haber DA, Li E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*. 1999;99(3): 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6.

178. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*. 1992;69(6): 915–926. https://doi.org/10.1016/0092-8674(92)90611-F.

179. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the Establishment of Maternal Genomic Imprints. *Science*. 2001;294(5551): 2536–2539. https://doi.org/10.1126/science.1065848.

180. Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, et al. Chromatin Targeting of de Novo DNA Methyltransferases by the PWWP Domain. *Journal of Biological Chemistry*. 2004;279(24): 25447–25454. https://doi.org/10.1074/jbc.M312296200.

181. Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*. 2015;520(7546): 243–247. https://doi.org/10.1038/nature14176.

182. Wagner EJ, Carpenter PB. Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology*. 2012;13(2): 115–126. https://doi.org/10.1038/nrm3274.

183. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*. 2007;448(7154): 714–717. https://doi.org/10.1038/nature05987.

184. Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature*. 2015;517(7536): 640–644. https://doi.org/10.1038/nature13899.

185. Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*. 2002;129(8): 1983–1993. https://doi.org/10.1242/dev.129.8.1983.

186. Zhang ZM, Lu R, Wang P, Yu Y, Chen D, Gao L, et al. Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature*. 2018;554(7692): 387–391. https://doi.org/10.1038/nature25477.

187. Ishiyama S, Nishiyama A, Saeki Y, Moritsugu K, Morimoto D, Yamaguchi L, et al. Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance. *Molecular Cell*. 2017;68(2): 350-360.e7. https://doi.org/10.1016/j.molcel.2017.09.037.

188. Chuang LSH, Ian HI, Koh TW, Ng HH, Xu G, Li BFL. Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21 ^{WAF1}. *Science*. 1997;277(5334): 1996–2000. https://doi.org/10.1126/science.277.5334.1996.

189. Sharif J, Muto M, Takebayashi S ichiro, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*. 2007;450(7171): 908–912. https://doi.org/10.1038/nature06397.

190. Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, et al. Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nature Structural & Molecular Biology*. 2012;19(11): 1155–1160. https://doi.org/10.1038/nsmb.2391.

191. Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature*. 2013;502(7470): 249–253. https://doi.org/10.1038/nature12488.

192. Qin W, Wolf P, Liu N, Link S, Smets M, Mastra FL, et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Research*. 2015;25(8): 911–929. https://doi.org/10.1038/cr.2015.72.

193. Keverne EB. Genomic imprinting, action, and interaction of maternal and fetal genomes. *Proceedings of the National Academy of Sciences of the United States of America*. 2015;112(22): 6834–6840. https://doi.org/10.1073/pnas.1411253111.

194. Elbracht M, Mackay D, Begemann M, Kagan KO, Eggermann T. Disturbed genomic imprinting and its relevance for human reproduction: Causes and clinical consequences. *Human Reproduction Update*. 2020;26(2): 197–213. https://doi.org/10.1093/humupd/dmz045.

195. Spahn L, Barlow DP. An ICE pattern crystallizes. *Nature Genetics*. 2003;35(1): 11–12. https://doi.org/10.1038/ng0903-11.

196. Wan L, Bartolomei MS. Chapter 7 Regulation of Imprinting in Clusters: Noncoding RNAs Versus Insulators. In: *Advances in Genetics*. Elsevier; 2008. p. 207–223. https://doi.org/10.1016/S0065-2660(07)00007-7. [Accessed 26th November 2023].

197. Kelsey G, Feil R. New insights into establishment and maintenance of DNA methylation imprints in mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2013;368(1609): 20110336. https://doi.org/10.1098/rstb.2011.0336.

198. Wang T, Li J, Yang L, Wu M, Ma Q. The Role of Long Non-coding RNAs in Human Imprinting Disorders: Prospective Therapeutic Targets. *Frontiers in Cell and Developmental Biology*. 2021;9: 730014. https://doi.org/10.3389/fcell.2021.730014.

199. Kanduri C. Long noncoding RNAs: Lessons from genomic imprinting. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. 2016;1859(1): 102–111. https://doi.org/10.1016/j.bbagrm.2015.05.006.

200. Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*. 2012;484(7394): 339–344. https://doi.org/10.1038/nature10960.

201. Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, et al. The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse

Primordial Germ Cells. *Molecular Cell*. 2012;48(6): 849–862. https://doi.org/10.1016/j.molcel.2012.11.001.

202. Guibert S, Forné T, Weber M. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Research*. 2012;22(4): 633–641. https://doi.org/10.1101/gr.130997.111.

203. Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, et al. Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. *Science*. 2013;339(6118): 448–452. https://doi.org/10.1126/science.1229277.

204. Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, et al. The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell*. 2015;161(6): 1437–1452. https://doi.org/10.1016/j.cell.2015.05.015.

205. Gkountela S, Zhang KX, Shafiq TA, Liao WW, Hargan-Calvopiña J, Chen PY, et al. DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell*. 2015;161(6): 1425–1436. https://doi.org/10.1016/j.cell.2015.05.012.

206. Kagiwada S, Kurimoto K, Hirota T, Yamaji M, Saitou M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *The EMBO Journal*. 2012;32(3): 340–353. https://doi.org/10.1038/emboj.2012.331.

207. Ohno R, Nakayama M, Naruse C, Okashita N, Takano O, Tachibana M, et al. A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. *Development*. 2013;140(14): 2892–2903. https://doi.org/10.1242/dev.093229.

208. Yamaguchi S, Hong K, Liu R, Shen L, Inoue A, Diep D, et al. Tet1 controls meiosis by regulating meiotic gene expression. *Nature*. 2012;492(7429): 443–447. https://doi.org/10.1038/nature11709.

209. Yamaguchi S, Shen L, Liu Y, Sendler D, Zhang Y. Role of Tet1 in erasure of genomic imprinting. *Nature*. 2013;504(7480): 460–464. https://doi.org/10.1038/nature12805.

210. Hill PWS, Leitch HG, Requena CE, Sun Z, Amouroux R, Roman-Trufero M, et al. Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature*. 2018;555(7696): 392–396. https://doi.org/10.1038/nature25964.

211. Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, et al. The DNA methylation landscape of human early embryos. *Nature*. 2014;511(7511): 606–610. https://doi.org/10.1038/nature13544.

212. Singh P, Li AX, Tran DA, Oates N, Kang ER, Wu X, et al. De Novo DNA Methylation in the Male Germ Line Occurs by Default but Is Excluded at Sites of H3K4 Methylation. *Cell Reports*. 2013;4(1): 205–219. https://doi.org/10.1016/j.celrep.2013.06.004.

213. Marques CJ, João Pinho M, Carvalho F, Bièche I, Barros A, Sousa M. DNA methylation imprinting marks and DNA methyltransferase expression in human

spermatogenic cell stages. *Epigenetics*. 2011;6(11): 1354–1361. https://doi.org/10.4161/epi.6.11.17993.

214. Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, et al. Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. *PLoS Genetics*. 2013;9(4). https://doi.org/10.1371/journal.pgen.1003439.

215. Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, et al. Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. Reik W (ed.) *PLoS Genetics*. 2012;8(1): e1002440. https://doi.org/10.1371/journal.pgen.1002440.

216. Smallwood SA, Tomizawa S ichi, Krueger F, Ruf N, Carli N, Segonds-Pichon A, et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genetics*. 2011;43(8): 811–814. https://doi.org/10.1038/ng.864.

217. Okae H, Chiba H, Hiura H, Hamada H, Sato A, Utsunomiya T, et al. Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development. Oakey RJ (ed.) *PLoS Genetics*. 2014;10(12): e1004868. https://doi.org/10.1371/journal.pgen.1004868.

218. Veselovska L, Smallwood SA, Saadeh H, Stewart KR, Krueger F, Maupetit Méhouas S, et al. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biology*. 2015;16(1): 1–17. https://doi.org/10.1186/s13059-015-0769-z.

219. Xu Q, Xiang Y, Wang Q, Wang L, Brind'Amour J, Bogutz AB, et al. SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nature Genetics*. 2019;51(5): 844–856. https://doi.org/10.1038/s41588-019-0398-7.

220. Brind'Amour J, Kobayashi H, Richard Albert J, Shirane K, Sakashita A, Kamio A, et al. LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nature Communications*. 2018;9(1): 3331. https://doi.org/10.1038/s41467-018-05841-x.

221. Stewart KR, Veselovska L, Kim J, Huang J, Saadeh H, Tomizawa S ichi, et al. Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes & Development*. 2015;29(23): 2449–2462. https://doi.org/10.1101/gad.271353.115.

222. Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, et al. A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Developmental Cell*. 2008;15(4): 547–557. https://doi.org/10.1016/j.devcel.2008.08.014.

223. Takahashi N, Coluccio A, Thorball CW, Planet E, Shi H, Offner S, et al. ZNF445 is a primary regulator of genomic imprinting. *Genes & Development*. 2019;33(1–2): 49–54. https://doi.org/10.1101/gad.320069.118.

224. Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, et al. In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated

Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. *Molecular Cell*. 2011;44(3): 361–372. https://doi.org/10.1016/j.molcel.2011.08.032.

225. Guo F, Li X, Liang D, Li T, Zhu P, Guo H, et al. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell*. 2014;15(4): 447–459. https://doi.org/10.1016/j.stem.2014.08.003.

226. Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell*. 2014;157(4): 979–991. https://doi.org/10.1016/j.cell.2014.04.017.

227. Shen L, Inoue A, He J, Liu Y, Lu F, Zhang Y. Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. *Cell Stem Cell*. 2014;15(4): 459–471. https://doi.org/10.1016/j.stem.2014.09.002.

228. Auclair G, Guibert S, Bender A, Weber M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biology*. 2014;15(12): 545. https://doi.org/10.1186/s13059-014-0545-5.

229. Guenatri M, Duffié R, Iranzo J, Fauque P, Bourc'his D. Plasticity in Dnmt3L-dependent and -independent modes of *de novo* methylation in the developing mouse embryo. *Development*. 2013;140(3): 562–572. https://doi.org/10.1242/dev.089268.

230. Monk D, Sanchez-Delgado M, Fisher R. NLRPS, the subcortical maternal complex and genomic imprinting. *Reproduction*. 2017;154(6): R161–R170. https://doi.org/10.1530/REP-17-0465.

231. Demond H, Anvar Z, Jahromi BN, Sparago A, Verma A, Davari M, et al. A KHDC3L mutation resulting in recurrent hydatidiform mole causes genome-wide DNA methylation loss in oocytes and persistent imprinting defects post-fertilisation. *Genome Medicine*. 2019;11(1): 1–14. https://doi.org/10.1186/s13073-019-0694-y.

232. Martin SR, Schilstra MJ. Circular Dichroism and Its Application to the Study of Biomolecules. In: *Methods in Cell Biology*. Elsevier; 2008. p. 263–293. https://doi.org/10.1016/S0091-679X(07)84010-6. [Accessed 27th November 2023].

233. Winter G, Lobley CMC, Prince SM. Decision making in *xia* 2. *Acta Crystallographica Section D Biological Crystallography*. 2013;69(7): 1260–1273. https://doi.org/10.1107/S0907444913015308.

234. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*. 2010;66(2): 213–221. https://doi.org/10.1107/S0907444909052925.

235. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. *Acta Crystallographica Section D Biological Crystallography*. 2010;66(4): 486–501. https://doi.org/10.1107/S0907444910007493.

236. Vaguine AA, Richelle J, Wodak SJ. *SFCHECK*: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallographica Section D Biological Crystallography*. 1999;55(1): 191–205. https://doi.org/10.1107/S0907444998006684.

237. Scheres SHW. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*. 2012;180(3): 519–530. https://doi.org/10.1016/j.jsb.2012.09.006.

238. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*. 2017;14(3): 290–296. https://doi.org/10.1038/nmeth.4169.

239. Knipp M, Vašák M. A colorimetric 96-well microtiter plate assay for the determination of enzymatically formed citrulline. *Analytical Biochemistry*. 2000;286(2): 257–264. https://doi.org/10.1006/abio.2000.4805.

240. Bicker KL, Subramanian V, Chumanevich AA, Hofseth LJ, Thompson PR. Seeing citrulline: Development of a phenylglyoxal-based probe to visualize protein citrullination. *Journal of the American Chemical Society*. 2012;134(41): 17015–17018. https://doi.org/10.1021/ja308871v.

241. Lewallen DM, Bicker KL, Subramanian V, Clancy KW, Slade DJ, Martell J, et al. Chemical Proteomic Platform to Identify Citrullinated Proteins. *ACS Chemical Biology*. 2015;10(11): 2520–2528. https://doi.org/10.1021/acschembio.5b00438.

242. Hunt PA, Worthman C, Levinson H, Stallings J, LeMaire R, Mroz K, et al. Germ cell loss in the XXY male mouse: Altered X-chromosome dosage affects prenatal development. *Molecular Reproduction and Development*. 1998;49(2): 101–111. https://doi.org/10.1002/(SICI)1098-2795(199802)49:2<101::AID-MRD1>3.0.CO;2-T.

243. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. 2012;9(7): 671–675. https://doi.org/10.1038/nmeth.2089.

244. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*. 2020;38(3): 276–278. https://doi.org/10.1038/s41587-020-0439-x.

245. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*. 2018;19(1): 15. https://doi.org/10.1186/s13059-017-1382-0.

246. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. 2009;4(1): 44–57. https://doi.org/10.1038/nprot.2008.211.

247. Ye Z, Sabatier P, Martin-Gonzalez J, Eguchi A, Bekker-Jensen DB, Bache N, et al. *One-Tip enables comprehensive proteome coverage in minimal cells and single zygotes*. 2023 Aug [Accessed 27th November 2023]. https://doi.org/10.1101/2023.08.10.552756. [Accessed 27th November 2023].

248. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011;17(1): 10. https://doi.org/10.14806/ej.17.1.200.

249. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27(11): 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.

250. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*. 2012;13(10). https://doi.org/10.1186/gb-2012-13-10-R87.

251. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*. 2009;25(14): 1841–1842. https://doi.org/10.1093/bioinformatics/btp328.

252. Schulz R, Woodfine K, Menheniott TR, Bourc'his D, Bestor T, Oakey RJ. WAMIDEX: A web atlas of murine genomic imprinting and differential expression. *Epigenetics*. 2008;3(2): 89–96. https://doi.org/10.4161/epi.3.2.5900.

253. Hinrichs AS. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research*. 2006;34(90001): D590–D598. https://doi.org/10.1093/nar/gkj144.

254. Taki H, Gomi T, Knuckley B, Thompson PR, Vugrek O, Hirata K, et al. Purification of enzymatically inactive peptidylarginine deiminase type 6 from mouse ovary that reveals hexameric structure different from other dimeric isoforms. *Advances in Bioscience and Biotechnology*. 2011;02(04): 304–310. https://doi.org/10.4236/abb.2011.24044.

255. Saijo S, Nagai A, Kinjo S, Mashimo R, Akimoto M, Kizawa K, et al. Monomeric Form of Peptidylarginine Deiminase Type I Revealed by X-ray Crystallography and Small-Angle X-ray Scattering. *Journal of Molecular Biology*. 2016;428(15): 3058–3073. https://doi.org/10.1016/j.jmb.2016.06.018.

256. Rechiche O, Lee TV, Lott JS. Structural characterization of human peptidyl-arginine deiminase type III by X-ray crystallography. *Acta Crystallographica Section F Structural Biology Communications*. 2021;77(10): 334–340. https://doi.org/10.1107/S2053230X21009195.

257. Robichon C, Luo J, Causey TB, Benner JS, Samuelson JC. Engineering Escherichia coli BL21(DE3) Derivative Strains To Minimize E. coli Protein Contamination after Purification by Immobilized Metal Affinity Chromatography. *Applied and Environmental Microbiology*. 2011;77(13): 4634–4646. https://doi.org/10.1128/AEM.00119-11.

258. Walker PA, Leong LEC, Ng PWP, Tan SH, Waller S, Murphy D, et al. Efficient and Rapid Affinity Purification of Proteins Using Recombinant Fusion Proteases. *Nature Biotechnology*. 1994;12(6): 601–605. https://doi.org/10.1038/nbt0694-601.

259. Kim SH, Yoo HJ, Park EJ, Na DH. Nano Differential Scanning Fluorimetry-Based Thermal Stability Screening and Optimal Buffer Selection for Immunoglobulin G. *Pharmaceuticals*. 2021;15(1): 29. https://doi.org/10.3390/ph15010029.

260. Minton AP. Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. *Analytical Biochemistry*. 2016;501: 4–22. https://doi.org/10.1016/j.ab.2016.02.007.

261. Wyatt PJ. Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta*. 1993;272(1): 1–40. https://doi.org/10.1016/0003-2670(93)80373-S.

262. Young G, Hundt N, Cole D, Fineberg A, Andrecka J, Tyler A, et al. Quantitative mass imaging of single biological macromolecules. *Science*. 2018;360(6387): 423–427. https://doi.org/10.1126/science.aar5839.

263. Andrade F, Darrah E, Gucek M, Cole RN, Rosen A, Zhu X. Autocitrullination of human peptidyl arginine deiminase type 4 regulates protein citrullination during cell activation. *Arthritis & Rheumatism*. 2010;62(6): 1630–1640. https://doi.org/10.1002/art.27439.

264. Fan X, Wang J, Zhang X, Yang Z, Zhang JC, Zhao L, et al. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nature Communications*. 2019;10(1): 2386. https://doi.org/10.1038/s41467-019-10368-w.

265. Morão LG, Manzine LR, Clementino LOD, Wrenger C, Nascimento AS. A scalable screening of E. coli strains for recombinant protein expression. Beddoe T (ed.) *PLOS ONE*. 2022;17(7): e0271403. https://doi.org/10.1371/journal.pone.0271403.

266. Oswald C, Smits S, Bremer E, Schmitt L. Microseeding – A Powerful Tool for Crystallizing Proteins Complexed with Hydrolyzable Substrates. *International Journal of Molecular Sciences*. 2008;9(7): 1131–1141. https://doi.org/10.3390/ijms9071131.

267. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*. 2007;372(3): 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.

268. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873): 583–589. https://doi.org/10.1038/s41586-021-03819-2.

269. Tutturen AEV, Holm A, Jørgensen M, Stadtmüller P, Rise F, Fleckenstein B. A technique for the specific enrichment of citrulline-containing peptides. *Analytical Biochemistry*. 2010;403(1–2): 43–51. https://doi.org/10.1016/j.ab.2010.04.012.

270. Van Oers MM, Pijlman GP, Vlak JM. Thirty years of baculovirus–insect cell protein expression: from dark horse to mainstream technology. *Journal of General Virology*. 2015;96(1): 6–23. https://doi.org/10.1099/vir.0.067108-0.

271. Smits AH, Jansen PWTC, Poser I, Hyman AA, Vermeulen M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Research*. 2013;41(1): e28–e28. https://doi.org/10.1093/nar/gks941.

272. Piersimoni L, Kastritis PL, Arlt C, Sinz A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein–Protein Interactions─A Method for All Seasons. *Chemical Reviews*. 2022;122(8): 7500–7531. https://doi.org/10.1021/acs.chemrev.1c00786.

273. Bashtrykov P, Jankevicius G, Jurkowska RZ, Ragozin S, Jeltsch A. The UHRF1 Protein Stimulates the Activity and Specificity of the Maintenance DNA Methyltransferase DNMT1 by an Allosteric Mechanism. *Journal of Biological Chemistry*. 2014;289(7): 4106–4115. https://doi.org/10.1074/jbc.M113.528893.

274. Taher L, Israel S, Drexler HCA, Makalowski W, Suzuki Y, Fuellen G, et al. The proteome, not the transcriptome, predicts that oocyte superovulation affects embryonic phenotypes in mice. *Scientific Reports*. 2021;11(1): 23731. https://doi.org/10.1038/s41598-021-03054-9.

275. Buccitelli C, Selbach M. mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics*. 2020;21(10): 630–644. https://doi.org/10.1038/s41576-020-0258-4.

276. Li Q, Mu L, Yang X, Wang G, Liang J, Wang S, et al. Discovery of Oogenesis Biomarkers from Mouse Oocytes Using a Single-Cell Proteomics Approach. *Journal of Proteome Research*. 2023;22(6): 2067–2078. https://doi.org/10.1021/acs.jproteome.3c00157.

277. Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, Jaenisch R, et al. Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes & Development*. 2008;22(12): 1607–1616. https://doi.org/10.1101/gad.1667008.

278. Grunau C. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Research*. 2001;29(13): 65e–665. https://doi.org/10.1093/nar/29.13.e65.

279. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods*. 2014;11(8): 817–820. https://doi.org/10.1038/nmeth.3035.

280. Miura F, Enomoto Y, Dairiki R, Ito T. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Research*. 2012;40(17): e136–e136. https://doi.org/10.1093/nar/gks454.

281. Clark SJ, Smallwood SA, Lee HJ, Krueger F, Reik W, Kelsey G. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nature Protocols*. 2017;12(3): 534–547. https://doi.org/10.1038/nprot.2016.187.

282. Berensmeier S. Magnetic particles for the separation and purification of nucleic acids. *Applied Microbiology and Biotechnology*. 2006;73(3): 495–504. https://doi.org/10.1007/s00253-006-0675-0.

283. Vaisvila R, Ponnaluri VKC, Sun Z, Langhorst BW, Saleh L, Guan S, et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Research*. 2021;31(7): 1280–1289. https://doi.org/10.1101/gr.266551.120.

284. Arita K, Shimizu T, Hashimoto H, Hidaka Y, Yamada M, Sato M. Structural basis for histone N-terminal recognition by human peptidylarginine deiminase 4. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(14): 5291–5296. https://doi.org/10.1073/pnas.0509639103.

285. Clark SJ, Argelaguet R, Kapourani CA, Stubbs TM, Lee HJ, Alda-Catalinas C, et al. ScNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells e. *Nature Communications*. 2018;9(1): 1–9. https://doi.org/10.1038/s41467-018-03149-4.

Appendices

Appendix 1. Human PADI sequence alignments

Protein sequence alignments of the human PADIs produced with Clustal Omega (31). PADI4 side-chain Ca^{2+} binding residues highlighted magenta, those that are conserved in PADI6 highlighted green, those that are not highlighted red. Amino acids involved in backbone interactions with Ca^{2+} highlighted grey. Catalytic triad residues highlighted yellow.

```
sp|Q6TGC4|PADI6_HUMAN      MVSVEGRAMSFQSIIHLSLDSPVHAVCVLGTEICLDLSGCAPQKCQCFTIHGSGRVLIDV       60
sp|Q9Y2J8|PADI2_HUMAN      --------MLRERTVRLQYGSRVEAVYVLGTYLWTDVYSAAPAGAQTFSLKHSEHVWVEV       52
sp|Q9UM07|PADI4_HUMAN      --------MAQGTLIRVTPEQPTHAVCVLGTLTQLDICSSAPEDCTSFSINASPGVVVDI       52
sp|Q9ULC6|PADI1_HUMAN      --------MAPKRVVQLSLKMPTHAVCVVGVEAHVDIHSDVPKGANSFRVSGSSGVEVFM       52
sp|Q9ULW8|PADI3_HUMAN      --------MSLQRIVRVSLEHPTSAVCVAGVETLVDIYGSVPEGTEMFEVYGTPGVDIYI       52
                                    *     :::     . ** * *.    *: . .*     * :  :  * : :

sp|Q6TGC4|PADI6_HUMAN      ANTVI--SEKEDATIWWPLSDPTYATVKMTSPSPSVDADKVSVTYYGPNEDAPVGTAVLY       118
sp|Q9Y2J8|PADI2_HUMAN      VRDGEAEEVATNGKQRWLLSPSTTLRVTMSQASTEASSDKVTVNYYDEEGSIPIDQAGLF       112
sp|Q9UM07|PADI4_HUMAN      AHGPP-AKKKSTGSSTWPLDPGVEVTLTMKVASGSTGDQKVQISYYGPKTP--PVKALLY       109
sp|Q9ULC6|PADI1_HUMAN      VYNRT-RVKEPIGKARWPLDTDADMVVSVGTASKELKDFKVRVSYFGEQEDQALGRSVLY       111
sp|Q9ULW8|PADI3_HUMAN      SPNME-RGRERADTRRWRFDATLEIIVVMNSPSNDLNDSHVQISYHSSHEPLPLAYAVLY       111
                                       .  * :.      : :   * .    :* :.*.. .       : *:

sp|Q6TGC4|PADI6_HUMAN      LTGIEVSLEVDIYRNGQVEMSSDKQAKKKWIWGPSGWGAILLVNCNPADVGQQLEDKKDK       178
sp|Q9Y2J8|PADI2_HUMAN      LTAIEISLDVDADRDGVVE--KNNPKKASWTWGPEGQGAILLVNCDRETPWLPKEDCRDE       170
sp|Q9UM07|PADI4_HUMAN      LTGVEISLCADITRTGKVKPTRAVKDQRTWTWGPCGQGAILLVNCDRDNLESSAMDCEDD       169
sp|Q9ULC6|PADI1_HUMAN      LTGVDISLEVDTGRTGKVK--RSQGDKKTWRWGPEGYGAILLVNCDRDNHRSAEPDLTHS       169
sp|Q9ULW8|PADI3_HUMAN      LTCVDISLDCDLNCEGRQD--RNFVDKRQWVWGPSGYGGILLVNCDRDDPSCDVQDNCDQ       169
                           ** :::**  *    *  .       :  * *** * *.******:         *   .

sp|Q6TGC4|PADI6_HUMAN      KVIFSESITNLSQMTLNVQGPSCILKKYRLVLHTSKEESKKARVYWPQKDN--SSTFELV       236
sp|Q9Y2J8|PADI2_HUMAN      KVYSKEDLKDMSQMILRTKGPDRLPAGYEIVLYISMSDSDKVGVFYVENPFF-GQRYIHI       229
sp|Q9UM07|PADI4_HUMAN      EVLDSEDLQDMSLMTLSTKTPKDFFTNHTLVLHVARSEMDKVRVFQATRGKL-SSKCSVV       228
sp|Q9ULC6|PADI1_HUMAN      WLMSLADLQDMSPMLLSCNGPDKLFDSHKLVLNVPFSDSKRVRVFCARGGNS-LSDYKQV       228
sp|Q9ULW8|PADI3_HUMAN      HVHCLQDLEDMSVMVLRTQGPAALFDDHKLVLHTSSYDAKRAQVFHICGPEDVCEAYRHV       229
                            :    :: ::* * *  : *  :   : :**     : .:. *:         .    :

sp|Q6TGC4|PADI6_HUMAN      LGPDQHAYTLALLGNHLKETFYVEAIAFPSAEFSGLISYSVSLVEESQDPSIPETVLYKD       296
sp|Q9Y2J8|PADI2_HUMAN      LGRRKLYHVVKYTGGSAELLFFVEGLCFPDEGFSGLVSIHVSLLEYMA-QDIPLTPIFTD       288
sp|Q9UM07|PADI4_HUMAN      LGPKWPSHYLMVPGGKHNMDFYVEALAFPDTDFPGLITLTISLLDTSN-LELPEAVVFQD       287
sp|Q9ULC6|PADI1_HUMAN      LGPQCLSYEVERQPGEQEIKFYVEGLTFPDADFLGLVSLSVSLVDPG---TLPEVTLFTD       285
sp|Q9ULW8|PADI3_HUMAN      LGQDKVSYEVPRLHGDE-ERFFVEGLSFPDAGFTGLISFHVTLLDDSN-EDFSASPIFTD       287
                           **     : :    .     *:**.: **.  * **::  ::*::      :    :: *

sp|Q6TGC4|PADI6_HUMAN      TVVFRVAPCVFIPCTQVPLEVYLCRELQ----LQGFVDTVTKLSEKSNSQVASVYEDPNR       352
sp|Q9Y2J8|PADI2_HUMAN      TVIFRIAPWIMTPNILPPVSVFVCCMKD----NYLFLKEVKNLVEKTNCELKVCFQYLNR       344
sp|Q9UM07|PADI4_HUMAN      SVVFRVAPWIMTPNTQPPQEVYACSIFE----NEDFLKSVTTLAMKAKCKLTICPEEENM       343
sp|Q9ULC6|PADI1_HUMAN      TVGFRMAPWIMTPNTQPPEELYVCRVMDTHGSNEKFLEDMSYLTLKANCKLTICPQVENR       345
sp|Q9ULW8|PADI3_HUMAN      TVVFRVAPWIMTPSTLPPLEVYVCRVRN----NTCFVDAVAELARKAGCKLTICPQAENR       343
                           :* **:** :: *    * .:: *   :       *:. :  *  *: .::    :  *

sp|Q6TGC4|PADI6_HUMAN      LGRWLQDEMAFCYTQAPHKTTSLILDTPQAADLDEFPMKYSLSPGIGYMIQDTEDHKVAS       412
sp|Q9Y2J8|PADI2_HUMAN      GDRWIQDEIEFGYIEAPHKGFPVVLDSPRDGNLKDFPVKELLGPDFGYVTREPLFESVTS       404
sp|Q9UM07|PADI4_HUMAN      DDQWMQDEMEIGYIQAPHKTLPVVFDSPRNRGLKEFPIKRVMGPDFGYVTRGPQTGGISG       403
sp|Q9ULC6|PADI1_HUMAN      NDRWIQDEMEFGYIEAPHKSFPVVFDSPRNRGLKDFPYKRILGPDFGYVTREIPLPGPSS       405
sp|Q9ULW8|PADI3_HUMAN      NDRWIQDEMELGYVQAPHKTLPVVFDSPRNGELQDFPYKRILGPDFGYVTREPRDRSVSG       403
                            .:*:***: : * :****   :::*:*:   *.:** *  :.*.:**: :       :.

sp|Q6TGC4|PADI6_HUMAN      MDSIGNLMVSPPVKVQGKEYPLGRVLIGSSFYPSAEGRAMSKTLRDFLYAQQVQAPVELY       472
sp|Q9Y2J8|PADI2_HUMAN      LDSFGNLEVSPPVTVNGKTYPLGRILIGSS-FPLSGGRRMTKVVRDFLKAQQVQAPVELY       463
sp|Q9UM07|PADI4_HUMAN      LDSFGNLEVSPPVTVRGKEYPLGRILFGDSCYPSNDSRQMHQALQDFLSAQQVQAPVKLY       463
sp|Q9ULC6|PADI1_HUMAN      LDSFGNLDVSPPVTVGGTEYPLGRILIGSS-FPKSGGRQMARAVRNFLKAQQVQAPVELY       464
sp|Q9ULW8|PADI3_HUMAN      LDSFGNLEVSPPVVANGKEYPLGRILIGGN-LPGSSGRRVTQVVRDFLHAQKVQPPVELF       462
                           :**:*** ***** . *. *****:*:*..  *   .* : :.::** **:** **:*:

sp|Q6TGC4|PADI6_HUMAN      SDWLMTGHVDEFMCFIPTDDKNEGKKGFLLLLASPSACYKLFREKQKEGYGDALLFDELR       532
sp|Q9Y2J8|PADI2_HUMAN      SDWLTVGHVDEFMSFVPIPG----TKKFLLLMASTSACYKLFREKQKDGHGEAIMFKGLG       519
sp|Q9UM07|PADI4_HUMAN      SDWLSVGHVDEFLSFVPAPD----RKGFRLLLASPRSCYKLFQEQQNEGHGEALLFEGIK       519
sp|Q9ULC6|PADI1_HUMAN      SDWLSVGHVDEFLTFVPTSD----QKGFRLLLASPSACLKLFQEKKEEGYGEAAQFDGLK       520
sp|Q9ULW8|PADI3_HUMAN      VDWLAVGHVDEFLSFVPAPD----GKGFRMLLASPGACFKLFQEKQKCGHGRALLFQGVV       518
                            *** .******: *:*  .     * * :*:**  :* ***:*::: *:* *  *. :
```

(continued on next page)

```
sp|Q6TGC4|PADI6_HUMAN      ADQLLSNGREAKTIDQLLADESLKKQNEYVEKCIHLNRDILKTELGLVEQDIIEIPQLFC       592
sp|Q9Y2J8|PADI2_HUMAN      GMS-----SKRITINKILSNESLVQENLYFQRCLDWNRDILKKELGLTEQDIIDLPALFK       574
sp|Q9UM07|PADI4_HUMAN      KKK-------QQKIKNILSNKTLREHNSFVERCIDWNRELLKRELGLAESDIIDIPQLFK       572
sp|Q9ULC6|PADI1_HUMAN      HQA-------KRSINEMLADRHLQRDNLHAQKCIDWNRNVLKRELGLAESDIVDIPQLFF       573
sp|Q9ULW8|PADI3_HUMAN      DDEQ----VKTISINQVLSNKDLINYNKFVQSCIDWNREVLKRELGLAECDIIDIPQLFK       574
                                      .*.::*::. * . * . : *:. **::** ****.* **:::* **

sp|Q6TGC4|PADI6_HUMAN      LEKLTNIPSDQQPKRSFARPYFPDLLRMIVMGKNLGIPKPFGPQIKGTCCLEEKICCLLE       652
sp|Q9Y2J8|PADI2_HUMAN      MDED-----------HRARAFFPNMVNMIVLDKDLGIPKPFGPQVEEECCLEMHVRGLLE       623
sp|Q9UM07|PADI4_HUMAN      LKEF-----------SKAEAFFPNMVNMLVLGKHLGIPKPFGPVINGRCCLEEKVCSLLE       621
sp|Q9ULC6|PADI1_HUMAN      LKNF-----------YAEAFFPDMVNMVVLGKYLGIPKPYGPIINGRCCLEEKVQSLLE       621
sp|Q9ULW8|PADI3_HUMAN      TERK-----------KATAFFPDLVNMLVLGKHLGIPKPFGPIINGCCCLEEKVRSLLE       622
                            ..             *  :**:::.*:*:.* ******:** ::  **** ::  ***

sp|Q6TGC4|PADI6_HUMAN      PLGFKCTFINDFDCYLTEVGDICACANIRRVPFAFKWWKMVP  694
sp|Q9Y2J8|PADI2_HUMAN      PLGLECTFIDDISAYHKFLGEVHCGTNVRRKPFTFKWWHMVP  665
sp|Q9UM07|PADI4_HUMAN      PLGLQCTFINDFFTYHIRHGEVHCGTNVRRKPFSFKWWNMVP  663
sp|Q9ULC6|PADI1_HUMAN      PLGLHCIFIDDYLSYHELQGEIHCGTNVRRKPFPFKWWNMVP  663
sp|Q9ULW8|PADI3_HUMAN      PLGLHCTFIDDFTPYHMLHGEVHCGTNVCRKPFSFKWWNMVP  664
                           ***:.* **:*   *    *:: . :*: * ** ****:***
```

Appendix 2. PADI6 multi specie sequence alignments

Protein sequence alignments of PADI6 in mice, rats, cows, humans and pigs, produced with Clustal Omega (31). Catalytic triad residues highlighted yellow.

```
sp|Q8K3V4|PADI6_MOUSE       --------MSFQNSLSLSLVNPTHALCMVGMEITLDISKCAPDKCKSFTIRGSPRILIHI      52
tr|D4ABN7|D4ABN7_RAT        --------MSFQSTLSLSLDSPTHAICMVGMEITLDISGCAPDKCESFTIRGSPRVLIHI      52
tr|F1N404|F1N404_BOVIN      VVSPEGCAMAFRNIVPLSLDSPTHAVCVLGMEIFLDISGCAPRACESFSVDASLGVVVRV      60
sp|Q6TGC4|PADI6_HUMAN       MVSVEGRAMSFQSIIHLSLDSPVHAVCVLGTEICLDLSGCAPQKCQCFTIHGSGRVLIDV      60
tr|F1SUQ9|F1SUQ9_PIG        --------MSFQSIIHLSLDSPVHSTCMLGMEICLDLSGCAPQKCESFTIASSPGVLVDI      52
                                    *:*:. : *** .*.*: *::* ** **:* ***  *:.*:: .*  ::: :

sp|Q8K3V4|PADI6_MOUSE       SSSVIA--GKEDTVVWRSMNHPTVALVRMVAPSPTVDEDKVLVSYFCPDQEVPTATAVLF      110
tr|D4ABN7|D4ABN7_RAT        SSSVIT--GKEDAVVWRPMKQPTEALVRMVSPSPTVDEDKVLVSYYCPDKEVPTATAVLF      110
tr|F1N404|F1N404_BOVIN      CGADPVESREKSAAPRWPLSRPTDVLVAMTSPNFAEDEGKVMIFYHQPDEDAPMATAVLH      120
sp|Q6TGC4|PADI6_HUMAN       ANTVIS--EKEDATIWWPLSDPTYATVKMTSPSPSVDADKVSVTYYGPNEDAPVGTAVLY      118
tr|F1SUQ9|F1SUQ9_PIG        YNVASGKVTQEAALTQWPLSHPMDVLVKMTSPSSDVNGDKVLVSYFLPHEDVPVATAILY      112
                             .       :: :     :. *  . * *.:*.   : .** : *. *.::.* .**:*.

sp|Q8K3V4|PADI6_MOUSE       LTGIEISLEADIYRDGQLDMPSDKQAKKKWMWGMNGWGAILLVNCSPNAVGQPDEQS---      167
tr|D4ABN7|D4ABN7_RAT        LTGIEVSLEADIYRDGQLDMPSDKQAKKRWMWGLNGWGAILLVNCNPETPSQSDEQSG--      168
tr|F1N404|F1N404_BOVIN      LTGIGRTWLWP----EQV---FLSSLQKTWVWGPGGWGAILLVNCSPSVSGQPVD-----      168
sp|Q6TGC4|PADI6_HUMAN       LTGIEVSLEVDIYRNGQVEMSSDKQAKKKWIWGPSGWGAILLVNCNPADVGQQLEDKKTK      178
tr|F1SUQ9|F1SUQ9_PIG        LTGIEVSLDVDIYRNGDVEMANDKQAKKNWVWGPDGWGAILLVNCALTNRGQI-KDGKIT      171
                            ****  :         ::     .. :* *:** .**********     .*  .

sp|Q8K3V4|PADI6_MOUSE       FQEGPREIQNLSQMNVTVEGPTSILQNYQLILHTSEEEAKKTRVYWSQR-GSSAYELVVG      226
tr|D4ABN7|D4ABN7_RAT        AQESPREIQNLSQMTLTVEGPTSILRDYQLLLHTSKEEAKKTRVYWAHR-GASAYELVVG      227
tr|F1N404|F1N404_BOVIN      -----KKVKSLSQMVLKVQGPSCITKNYRVVLHTSKEESEKARVYRPQNDGSMTFELILG      223
sp|Q6TGC4|PADI6_HUMAN       KVIFSEEITNLSQMTLNVQGPSCILKKYRLVLHTSKEESKKARVYWPQKDNSSTFELVLG      238
tr|F1SUQ9|F1SUQ9_PIG        KVFSAEDIRNLSPMVLRVQGPSCILKKHRLLLHTSKEESDKARVYRPQKDSSSTFELVLG      231
                                 ..: .** * : *:**:.* :.::::*****:**:.*:***  :. .: ::**::*

sp|Q8K3V4|PADI6_MOUSE       PNKPVYLLPTF----ENRRKEAFYVEATEFPSPSFSGLISLSLSLVEKAHDECIPEIPLY      282
tr|D4ABN7|D4ABN7_RAT        PDKPFHILPTF----DTRGKETFYIEATEFPSPSFSGLISFSVSLIEQAHDQLIPETPLY      283
tr|F1N404|F1N404_BOVIN      PEQHAYTLAPQWDEVKGTLRETFYVEALEFPSASFSGLISFSASLVEESQDPLVPEAVLC      283
sp|Q6TGC4|PADI6_HUMAN       PDQHAYTLALL----GNHLKETFYVEAIAFPSAEFSGLISYSVSLVEESQDPSIPETVLY      294
tr|F1SUQ9|F1SUQ9_PIG        PGQHAYALGPL----ESDLKETFYVEALQFPSGSFPGLISYSVSLVEESKDGIGPQILVY      287
                            * :  : *           :*:**:**  *** .* **** * **:*:::*   *:  :

sp|Q8K3V4|PADI6_MOUSE       KDTVMFRVAPYIFMPSTQMPLEVYLCRELQLQGFVDSVTKLSEKSKVQVVKVYEDPNRQS      342
tr|D4ABN7|D4ABN7_RAT        KDTVMFRVAPYIFLPSTQLPLEVYLCRELQLQGFVDSVTKLSEKSNVQVASVYEDPNRQG      343
tr|F1N404|F1N404_BOVIN      KDTVLFRVAPCIFVPSTQMPLEVYLCRELQVQGFVSTVTELSERSNSQVASVYEDPNRLG      343
sp|Q6TGC4|PADI6_HUMAN       KDTVVFRVAPCVFIPCTQVPLEVYLCRELQLQGFVDTVTKLSEKSNSQVASVYEDPNRLG      354
tr|F1SUQ9|F1SUQ9_PIG        KDTVVFRVAPCIFTPSTQMPLEMYVCRELQVQGFVNAVTELSERSNSQVASVYEDPNRLG      347
                            ****:***** :* *.**:***:*:*****:****.:**:***:*: **..******* .

sp|Q8K3V4|PADI6_MOUSE       KWLQDEMAFCYTQAPHKTVSLILDTPRVSKLEDFPMKYTLTPGSGYLIRQTEDHRVASLD      402
tr|D4ABN7|D4ABN7_RAT        RWLQDEMAFCYTQAPHKTLSFILDTPRVAKLEDFPMKYSLSPGVGYLIQETEDHRVASLD      403
tr|F1N404|F1N404_BOVIN      RWLQDEMAFCYTQAPHKTVSLVLDTPRVAKLDDFPMKYSLSPGVGYLTLHTQDHTVASID      403
sp|Q6TGC4|PADI6_HUMAN       RWLQDEMAFCYTQAPHKTTSLILDTPQAADLDEFPMKYSLSPGIGYMIQDTEDHKVASMD      414
tr|F1SUQ9|F1SUQ9_PIG        RWLQDEMAFCYTQAPHKTTALVLDTPRVTKLEDFPMRYSLSPGIGYVTLNTLDHTVASID      407
                            :***************** :::****:.:.*::***:*:*:** **:  .* ** ***:*

sp|Q8K3V4|PADI6_MOUSE       SIGNLMVSPPVKAQGKDYPLGRVLIGGSFYPSSEGRDMNKGLREFVYAQQVQAPVELFSD      462
tr|D4ABN7|D4ABN7_RAT        SVGNLIVSPPVKAQGKDYPLGRILIGGSFYPSFEGRDMSKALRDFVYAQQVQAPVELFSD      463
tr|F1N404|F1N404_BOVIN      VVGKLMVSPPVKAQGREYPLGRVLIGSSFYPSKDSRNVSRSLQDFLRAQQVQAPVELFSD      463
sp|Q6TGC4|PADI6_HUMAN       SIGNLMVSPPVKVQGKEYPLGRVLIGSSFYPSAEGRAMSKTLRDFLYAQQVQAPVELYSD      474
tr|F1SUQ9|F1SUQ9_PIG        SIGNLMVSPPVKVKGKEYPLGRILIGSSFYPSKDGRDMSKALRDFLYAQQVQAPVELFSD      467
                             :*:*:******.:*::*****:***.***** :.* :.: *::*: **********:**

sp|Q8K3V4|PADI6_MOUSE       WLMTGHMDQFMCFVPTNDKNNDQKDFRLLLASPSACFELFEQKQKEGYGNVTLFEDIGAE      522
tr|D4ABN7|D4ABN7_RAT        WLMTGHVDQFLCFVPVDDKNSDRKGFRLLLASPSACYELFEQKQKEGFGDVTLFEDVRAE      523
tr|F1N404|F1N404_BOVIN      WLMTGHADEFMRFIPAQYKVEGKKDFRLLLASPSSCYKLFKERQKEGYGDAMLFEGLRKD      523
sp|Q6TGC4|PADI6_HUMAN       WLMTGHVDEFMCFIPTDDKNEGKKGFLLLLASPSACYKLFREKQKEGYGDALLFDELRAD      534
tr|F1SUQ9|F1SUQ9_PIG        WLMIGHIVEFMCFIPTQFKVEGKKDFRLLLASPSTCYKLFTEKQKQGYGDAALFGEIRKE      527
                            *** **  :*: *:*.: * ..:*.* *******:*::** ::**:*:*:. **  :  :

sp|Q8K3V4|PADI6_MOUSE       QLLSNGRESKTISQILADKSFREQNTYVEKCISLNRTLLKTELGLEDKDIILIPQLFCLE      582
tr|D4ABN7|D4ABN7_RAT        QLISNGREVKTISQILADENLREQNAYVEKCISLNRTLLKKELGLVDNDFIRIPQLFCLE      583
tr|F1N404|F1N404_BOVIN      QLLSNGREAITINQLLADEKMRKQNDYAEKCIHLNRGILKRELGLQEEDIIPVPQLFCLE      583
sp|Q6TGC4|PADI6_HUMAN       QLLSNGREAKTIDQLLADESLKKQNEYVEKCIHLNRDILKTELGLVEQDIIEIPQLFCLE      594
tr|F1SUQ9|F1SUQ9_PIG        QLLSNGREALTINQLLADENMKKQNDYVERCIDLNRAILKRELGLAEKDIIDIPQLFCLE      587
                            **:*****  **.*:***:.:::** *.*:** *** :** **** ::*:* :*******
```

(continued on next page)

```
sp|Q8K3V4|PADI6_MOUSE       QLTNVPSNQQSTKLFARPYFPDMLQIIVLGKNLGIPKPFGPKINGTCCLEEKVCGLLEPL      642
tr|D4ABN7|D4ABN7_RAT        PLTNVPSNQQTSKLFARPYFPDMLQMIVLGKNIGIPKPFGPQIKGTCCLEETVCGLLEPL      643
tr|F1N404|F1N404_BOVIN      HIANAPPSEQTKKLYARPYFPDLVRMVVMGLNLGIPKPFGPRINGTCCLEERICQLLEPL      643
sp|Q6TGC4|PADI6_HUMAN       KLTNIPSDQQPKRSFARPYFPDLLRMIVMGKNLGIPKPFGPQIKGTCCLEEKICCLLEPL      654
tr|F1SUQ9|F1SUQ9_PIG        QIVNVPSNELTGKLYARPYFPNLLQIIVMGQNLGIPKPFGPQINGACCLEEKIYELLEPL      647
                             :.* * .:   : :******:::::*:* *:********:*:*:***** :  *****

sp|Q8K3V4|PADI6_MOUSE       GLKCTFIDDFDCYLANIGDVCACAIINRVPFAFKWWKMTP  682
tr|D4ABN7|D4ABN7_RAT        GLKCTFIDDFDCYLANIGDACACVIVNRVPFAFKWWKMNP  683
tr|F1N404|F1N404_BOVIN      GFQCTFIDDFDCYLTEIGDFCSCANIRRVPFAFKWWGMVP  683
sp|Q6TGC4|PADI6_HUMAN       GFKCTFINDFDCYLTEVGDICACANIRRVPFAFKWWKMVP  694
tr|F1SUQ9|F1SUQ9_PIG        GFQCTFINDFDCYLTEIGDFCSCANIRRVPFAFKWWKMVP  687
                             *::****:******:::** *:.. :.********* * *
```

Appendix 3. InFusion cloning primers

Primers used for InFusion cloning. Bb = Backbone, I = Insert, F = Forward, R = Reverse.

Plasmid	Primer	Sequence
pMCSG_His-TEV-hPADI6	Bb-F	GAGAACCTGTACTTCCAATCCGCC
	Bb-R	CTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAAT
	I-F	AACTTTAAGAAGGAGATGGTCAGCGTGGAGGG
	I-R	GAAGTACAGGTTCTCAGGTACCATCTTCCACCATTTGAAG
pGEX6P2_GST-3C-hPADI6	Bb-F	GCGGCCGCATCGTGACTGAC
	Bb-R	GGATCCCAGGGGCCCCTGGAAC
	I-F	GGGGCCCCTGGGATCCATGGTCAGCGTGGAGGG
	I-R	AGTCACGATGCGGCCGCTTAAGGTACCATCTTCCACCATTTGAAGG
pMCSG_His-3C-hPADI6	Bb-F	AACCCAGCTTTCTTGTACAAAGTGG
	Bb-R	ATGGTGATGGTGATGGTGGGA
	I-F	CATCACCATCACCATCTGGAAGTTCTGTTCCAGGGG
	I-R	CAAGAAAGCTGGGTTTTAAGGTACCATCTTCCACCATTTGAAGG
pET28a_His-hPADI6	Bb-F	CAAAGCCCGAAAGGAAGCTGAG
	Bb-R	GCCGCTGCTGTGATGATGA
	I-F	CATCACAGCAGCGGCATGGTCAGCGTGGAGGG
	I-R	TCCTTTCGGGCTTTGTTAAGGTACCATCTTCCACCATTTGAAGG
pcDNA3.1_Strep-Strep-TEV-hPADI6	Bb-F	TAATAATCTAGAGGGCCCGTTTAAACCC
	Bb-R	GCTGCCGCTACCTGACT
	I-F	TCAGGTAGCGGCAGCATGGTCAGCGTGGAG
	I-R	CCCTCTAGATTATTATTAAGGTACCATCTTCCACCATTTGAAG
pMCSG_His-3C-mPADI6	Bb-F	AACCCAGCTTTCTTGTACAAAGTGG
	Bb-R	GGATCCCAGGGGCCCCTGGAAC
	I-F	GGGCCCCTGGGATCCATGTCTTTTCAGAACTCACTCAGCCT
	I-R	CAAGAAAGCTGGGTTTTATGGGGTCATCTTCCACCACTTG
pcDNA3.1_Strep-Strep-TEV-mPADI6	Bb-F	TAATAATAATCTAGAGGGCCCGTTTAAACCC
	Bb-R	GCTGCCGCTACCTGACT
	I-F	TCAGGTAGCGGCAGCATGTCTTTTCAGAACTCACTCAGCC
	I-R	CCTCTAGATTATTATTATGGGGTCATCTTCCACCACT
pcDNA3.1_FLAG-hPADI6	Bb-F	CCAGTCGACTCTAGAGGATCCCG
	Bb-R	TATCGATGAATTCGCGGCCGV
	I-F	GCGAATTCATCGATAATGGTCAGCGTGGAGGG
	I-R	TCTAGAGTCGACTGGCTAAGGTACCATCTTCCACCATTTGAAG
pCMV_HA-TLE6	Bb-F	GCGGCCGCGGGGATCCAGAC
	Bb-R	GAATTCCTCCATGGCCATAAG
	I-F	GGCCATGGAGGAATTCATGACCTCTAGGGACCAGCC

	I-R	TGGATCCCCGCGGCCGCTCAGTAGGTGATCTGGTACACGGAGG
pCMV_HA-OOEP	Bb-F	GCGGCCGCGGGGATC
	Bb-R	GAATTCCTCCATGGCCATAAGAGC
	I-F	GCCATGGAGGAATTCATGGTCGATGATGCTGGTGC
	I-R	GATCCCCGCGGCCGCTTAAGCAACAGGATCCTGGGGAG
pcDNA3.1_V5-KHDC3L	Bb-F	GGTAAGCCTATCCCTAACCCTCT
	Bb-R	TGATATCTGCAGAATTCCACCACACT
	I-F	ATTCTGCAGATATCAATGGACGCTCCCAGGC
	I-R	AGGGATAGGCTTACCTAATCTAGTAACTGGGTCCCGGGC
pET28a_His-NALP5	Bb-F	TAACAAAGCCCGAAAGGAAGCTG
	Bb-R	GCCGCTGCTGTGATGATGA
	I-F	CATCACAGCAGCGGCATGAAGGTTGCAGGAGGACTTG
	I-R	TCCTTTCGGGCTTTGTTAGTTTTTCCACCAGTACCGGTCATCT
pET28a_His-TLE6	Bb-F	TAACAAAGCCCGAAAGGAAGCTG
	Bb-R	GCCGCTGCTGTGATGATGA
	I-F	CATCACAGCAGCGGCATGACCTCTAGGGACCAGCC
	I-R	TTTCGGGCTTTGTTAGTAGGTGATCTGGTACACGG
pET28a_His-OOEP	Bb-F	TAACAAAGCCCGAAAGGAAGCTG
	Bb-R	GCCGCTGCTGTGATGATGA
	I-F	CATCACAGCAGCGGCATGGTCGATGATGCTGGTGC
	I-R	TCCTTTCGGGCTTTGTTAAGCAACAGGATCCTGGGGAGAG
pET28a_His-TLE6_WD40	Bb-F	TAACAAAGCCCGAAAGGAAGCTG
	Bb-R	GCCGCTGCTGTGATGATGA
	I-F	CATCACAGCAGCGGCAGACTCGCCGTCCCGTG
	I-R	TTTCGGGCTTTGTTAGTAGGTGATCTGGTACACGG
pcDNA3.1_Strep-Strep-TEV-NALP5	Bb-F	TAATAATCTAGAGGGCCCGTTTAAACCC
	Bb-R	GCTGCCGCTACCTGACT
	I-F	TCAGGTAGCGGCAGCATGAAGGTTGCAGGAGGACTTG
	I-R	CCCTCTAGATTATTATTAGTTTTTCCACCAGTACCGGTCA
pcDNA3.1_Strep-Strep-TEV-TLE6	Bb-F	TAATAATCTAGAGGGCCCGTTTAAACCC
	Bb-R	GCTGCCGCTACCTGACT
	I-F	TCAGGTAGCGGCAGCATGACCTCTAGGGACCAGCC
	I-R	CCCTCTAGATTATTATCAGTAGGTGATCTGGTACACGG
pcDNA3.1_Strep-Strep-TEV-OOEP	Bb-F	TAATAATCTAGAGGGCCCGTTTAAACCC
	Bb-R	GCTGCCGCTACCTGACT
	I-F	TCAGGTAGCGGCAGCATGGTCGATGATGCTGGTGC
	I-R	CCCTCTAGATTATTATTAAGCAACAGGATCCTGGGG

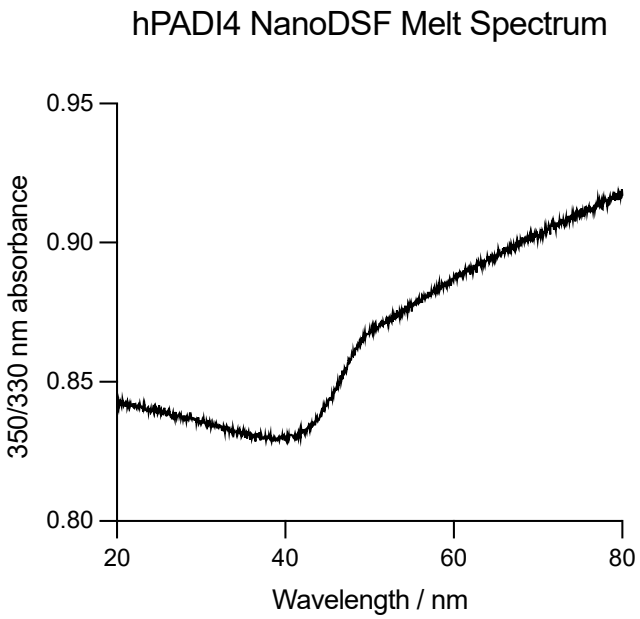
Appendix 4. Mutagenesis primers

Primers used for plasmid mutagenesis. F = Forward, R = Reverse.

Plasmid (Substitution)	Primer	Sequence
pcDNA3.1_Strep-Strep-TEV-mPADI6 (C663A)	F	CATAGGGGACGTCGCTGCCAGTGCCATCATAAAC
	R	GTTTATGATGGCACTGGCAGCGACGTCCCCTATG
pcDNA3.1_FLAG-hPADI6 (H211Q)	F	GAAATATCGGCTAGTCCTCCAGACCTCCAAGGAAGAG
	R	CTCTTCCTTGGAGGTCTGGAGGACTAGCCGATATTTC
pcDNA3.1_FLAG-hPADI6 (P289L)	F	GTCTCAAGACCCGTCAATTCTGGAGACTGTGCTGTACAAAGAC
	R	GTCTTTGTACAGCACAGTCTCCAGAATTGACGGGTCTTGAGAC
pcDNA3.1_FLAG-hPADI6 (W356*)	F	GACCCCAACCGCCTGGGCAGGTGACTCCAGGATGAGATGG
	R	CCATCTCATCCTGGAGTCACCTGCCCAGGCGGTTGGGGTC
pcDNA3.1_FLAG-hPADI6 (L375S)	F	CACAAGACAACGTCCTCGATCCTCGACACACCTC
	R	GAGGTGTGTCGAGGATCGAGGACGTTGTCTTGTG
pcDNA3.1_FLAG-hPADI6 (M477V)	F	CTACTCAGATTGGCTAGTGACTGGCCACGTGG
	R	CCACGTGGCCAGTCACTAGCCAATCTGAGTAG
pcDNA3.1_FLAG-hPADI6 (G540R)	F	GCAGATCAGCTCCTGTCTAATCGGAGGGAAGCCAAAACCATCG
	R	CGATGGTTTTGGCTTCCCTCCGATTAGACAGGAGCTGATCTGC
pcDNA3.1_FLAG-hPADI6 (D547N)	F	GGGAAGCCAAAACCATCAACCAACTTCTGGCTG
	R	CAGCCAGAAGTTGGTTGATGGTTTTGGCTTCCC
pcDNA3.1_FLAG-hPADI6 (E586K)	F	GGTGGAACAGGACATCATCAAGATTCCCCAGCTGTTCTG
	R	CAGAACAGCTGGGGAATCTTGATGATGTCCTGTTCCACC
pcDNA3.1_FLAG-hPADI6 (N598S)	F	GTTCTGCTTGGAGAAGCTGACTAGCATCCCCTCTGACCAGCAG
	R	CTGCTGGTCAGAGGGGATGCTAGTCAGCTTCTCCAAGCAGAAC
pcDNA3.1_FLAG-hPADI6 (P632A)	F	CCTGGGGATCCCCAAGGCTTTTGGGCCCCAAATC
	R	GATTTGGGGCCCAAAAGCCTTGGGGATCCCCAGG
pcDNA3.1_FLAG-hPADI6 (P632L)	F	GAACCTGGGGATCCCCAAGCTGTTTGGGCCCCAAATCAAGGG
	R	CCCTTGATTTGGGGCCCAAACAGCTTGGGGATCCCCAGGTTC
pcDNA3.1_FLAG-hPADI6 (P632T)	F	GAACCTGGGGATCCCCAAGACCTTTGGGCCCCAAATCAAGGG
	R	CCCTTGATTTGGGGCCCAAAGGTCTTGGGGATCCCCAGGTTC

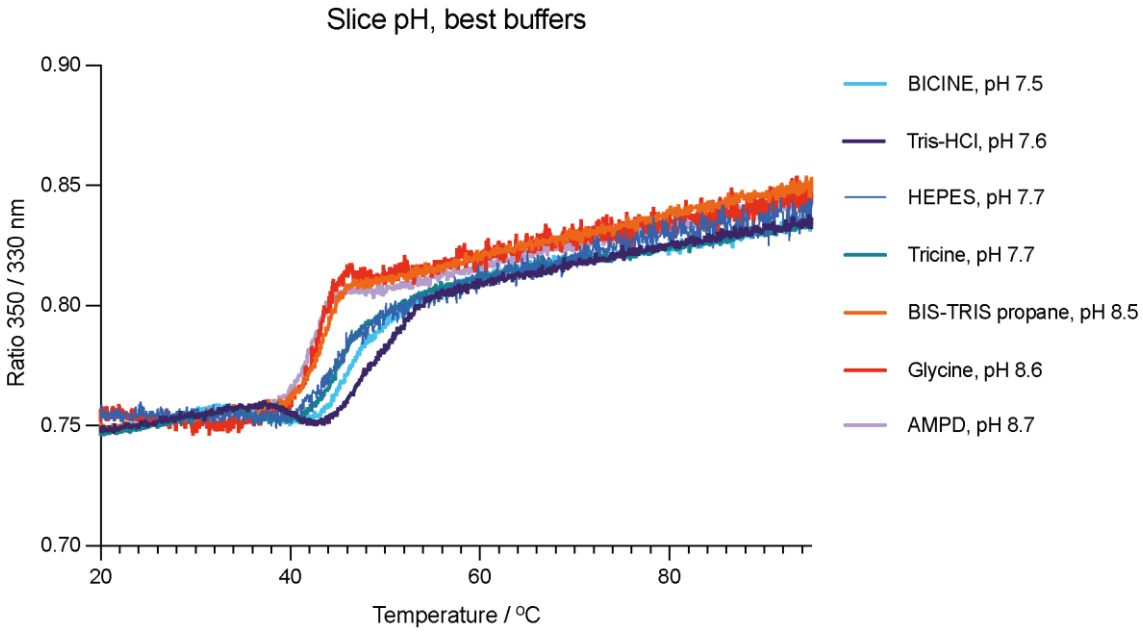
Appendix 5. hPADI4 NanoDSF

NanoDSF melting spectrum of hPADI4 purified from *E. coli* and donated by Dr. Tessa Bertran of the Walport group. Absorbance was measured at 350 and 330 nm and the ratio plotted. T_m = 46.5 ^oC. Experiment was performed in triplicate.



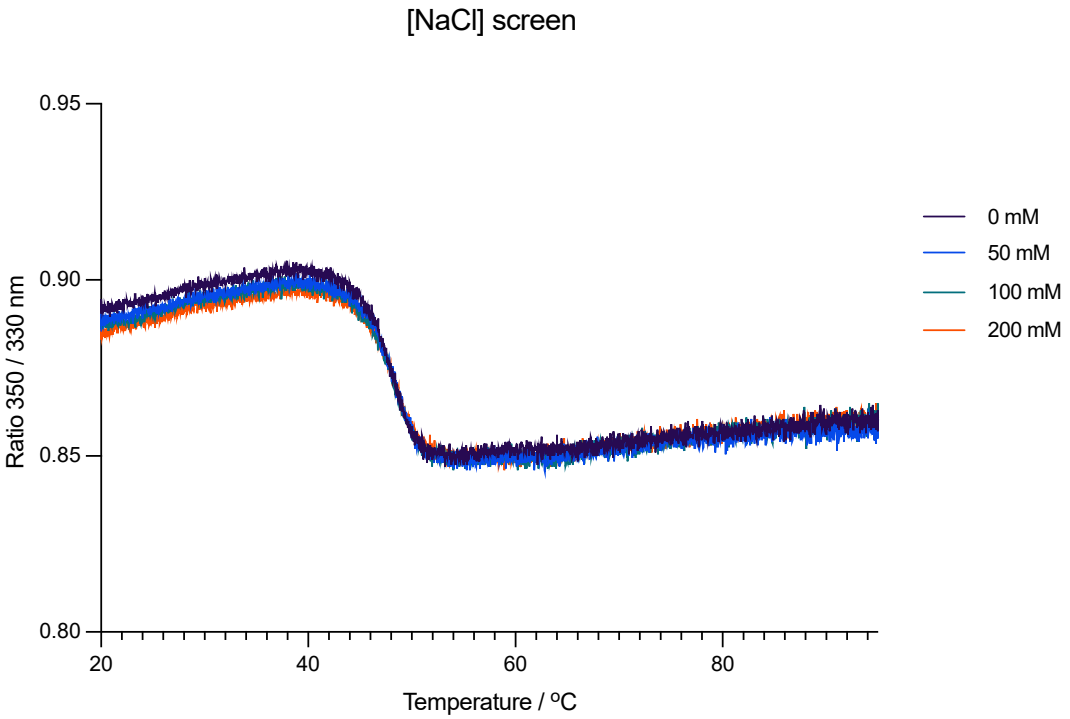
Appendix 6. Slice pH best buffers NanoDSF data

NanoDSF melting spectrum of hPADI6 purified from *E. coli* (pMCSG_His-3C-hPADI6 construct) diluted in the top seven buffers from the Slice pH buffer screen. Absorbance was measured at 350 and 330 nm and the ratio plotted. Tris-HCl, pH 7.6 stabilised the protein the most with a T_m of 49.5 °C.



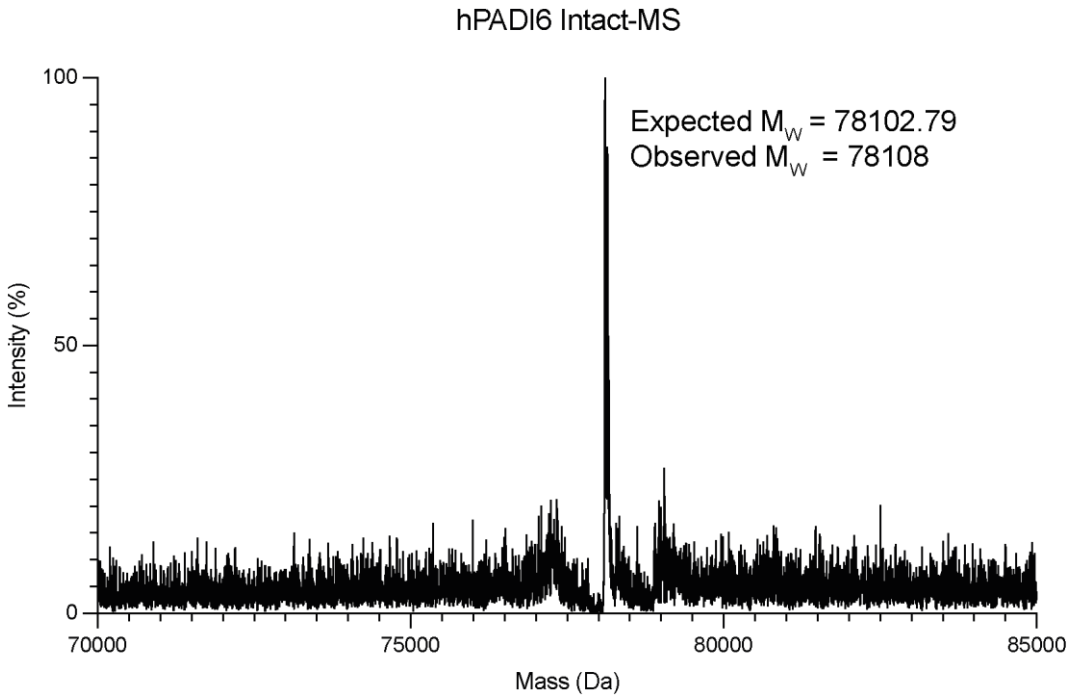
Appendix 7. NaCl concentration NanoDSF data

NanoDSF melting spectrum of hPADI6 purified from *E. coli* (pMCSG_His-3C-hPADI6 construct) in Tris-HCl, pH 7.6 buffer with either 0, 50, 100, or 200 mM NaCl. Absorbance was measured at 350 and 330 nm and the ratio plotted. No difference in melting temperature was observed between NaCl concentrations.



Appendix 8. Expi293 expressed hPADI6 Intact-MS

Intact-MS spectra of hPADI6 purified from Expi293 cells (SS-TEV-hPADI6 construct). Data collected by Sarah Maslen of the Crick Proteomics STP and plotted in GraphPad prism.



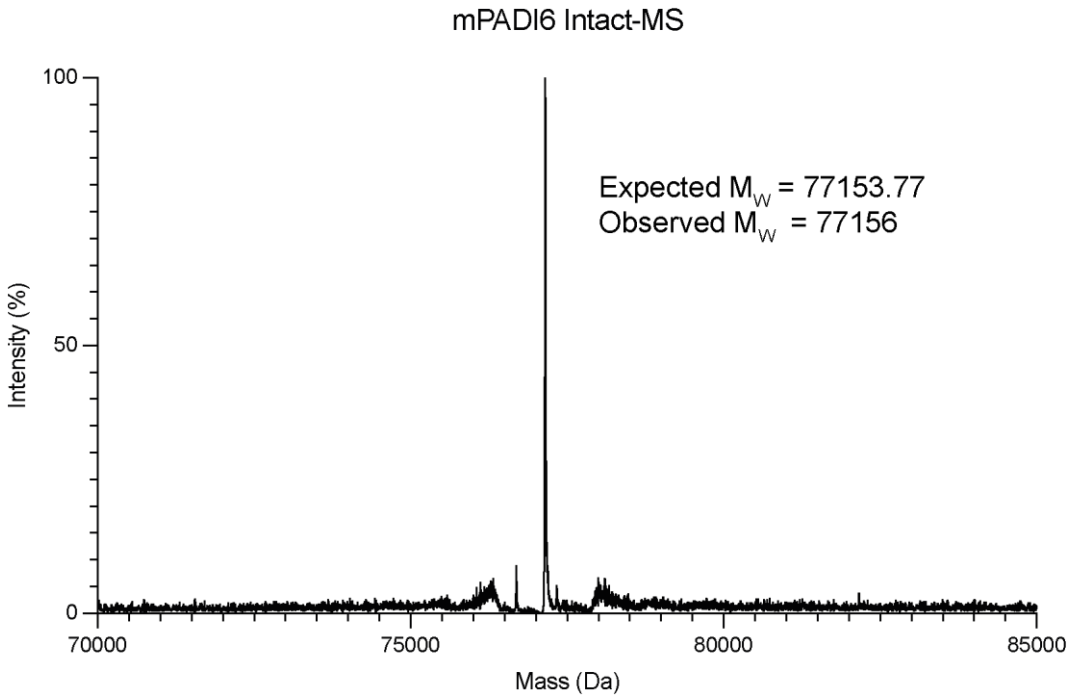
Appendix 9. hPADI6 X-Ray crystal structure statistics

hPADI6 X-Ray crystal structure statistics. Model and table produced by Stephane Mouilleron of the Structural Biology STP.

	hPADI6
Resolution range	54.06 - 2.44 (2.48 - 2.44)
Space group	$P2_12_12_1$
Unit cell a, b, c	104.04 123.33 126.54
α, β, γ	90 90 90
Total reflections	896 262 (36 822)
Unique reflections	65 162 (2 859)
Multiplicity	13.8 (12.9)
Completeness (%)	99.20 (93.53)
Mean I/sigma(I)	4.34 (0.19)
Wilson B-factor	58.14
R-merge	0.38 (7.48)
R-meas	0.39 (7.79)
R-pim	0.10 (2.16)
CC1/2	0.97 (0.26)
Reflections used in refinement	60 733 (2 573)
Reflections used for R-free	3 026 (144)
R_{work}	0.26 (0.47)
R_{free}	0.30 (0.47)
Number of non-hydrogen atoms	10 276
macromolecules	10 276
ligands	0
solvent	0
Protein residues	1 332
RMS(bonds)	0.002
RMS(angles)	0.58
Ramachandran favoured (%)	97.65
Ramachandran allowed (%)	2.35
Ramachandran outliers (%)	0.00
Clash score	1.96
Average B-factor	77.20

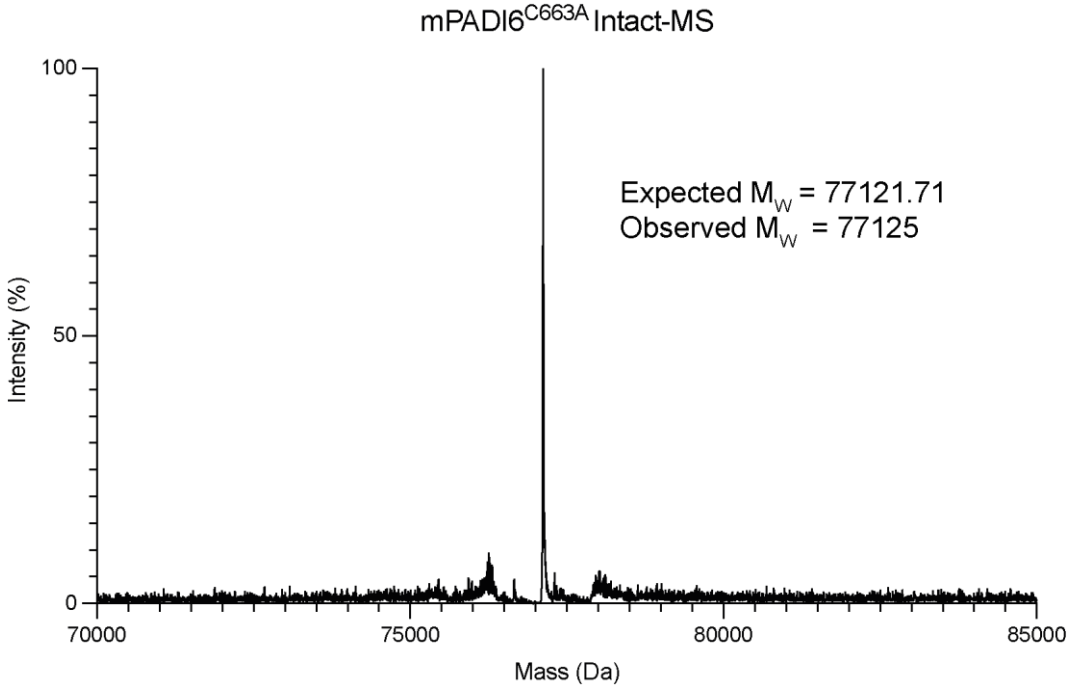
Appendix 10. Expi293 expressed mPADI6 Intact-MS

Intact-MS spectra of mPADI6 purified from Expi293 cells (SS-TEV-mPADI6 construct). Data collected by Sarah Maslen of the Crick Proteomics STP and plotted in GraphPad prism.



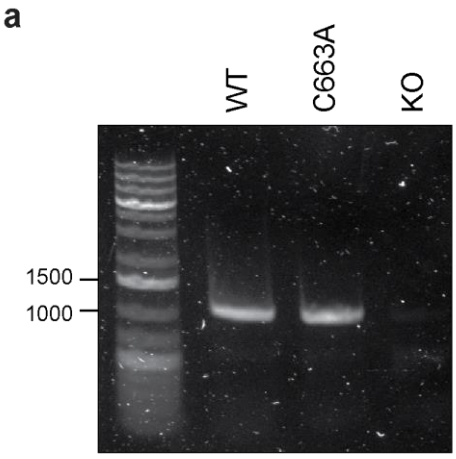
Appendix 11. Expi293 expressed mPADI6^{C663A} Intact-MS

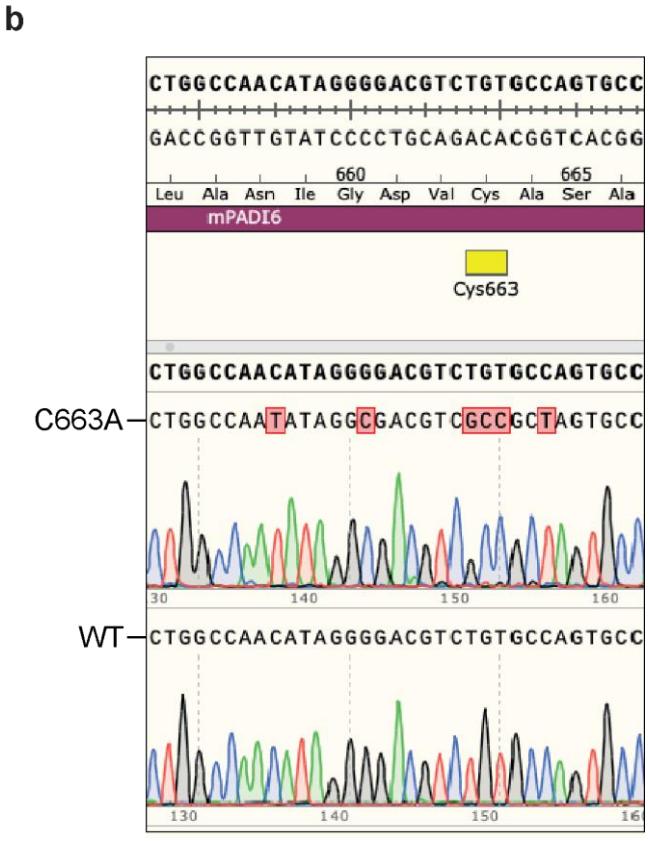
Intact-MS spectra of mPADI6^{C663A} purified from Expi293 cells (SS-TEV-mPADI6^{C663A} construct). Data collected by Sarah Maslen of the Crick Proteomics STP and plotted in GraphPad prism. M_w matches the expected M_w of the cysteine to alanine substitution.



Appendix 12. *Padi6*$^{C663A/C663A}$ ovary QC

a) 1% agarose gel image of Padi6 mRNA amplification from WT, Padi6$^{C663A/C663A}$ (C663A) and Padi6$^{-/-}$ (KO) ovaries. No *Padi6* mRNA was amplified from KO ovaries. **b)** Screenshot of sanger sequencing results of Padi6 mRNA fragment amplified from WT and C663A ovaries. The C663A fragment possessed the desired TGT (cysteine) to GCC (alanine) codon switch, along with other synonymous mutations included in the guide RNA cassette.





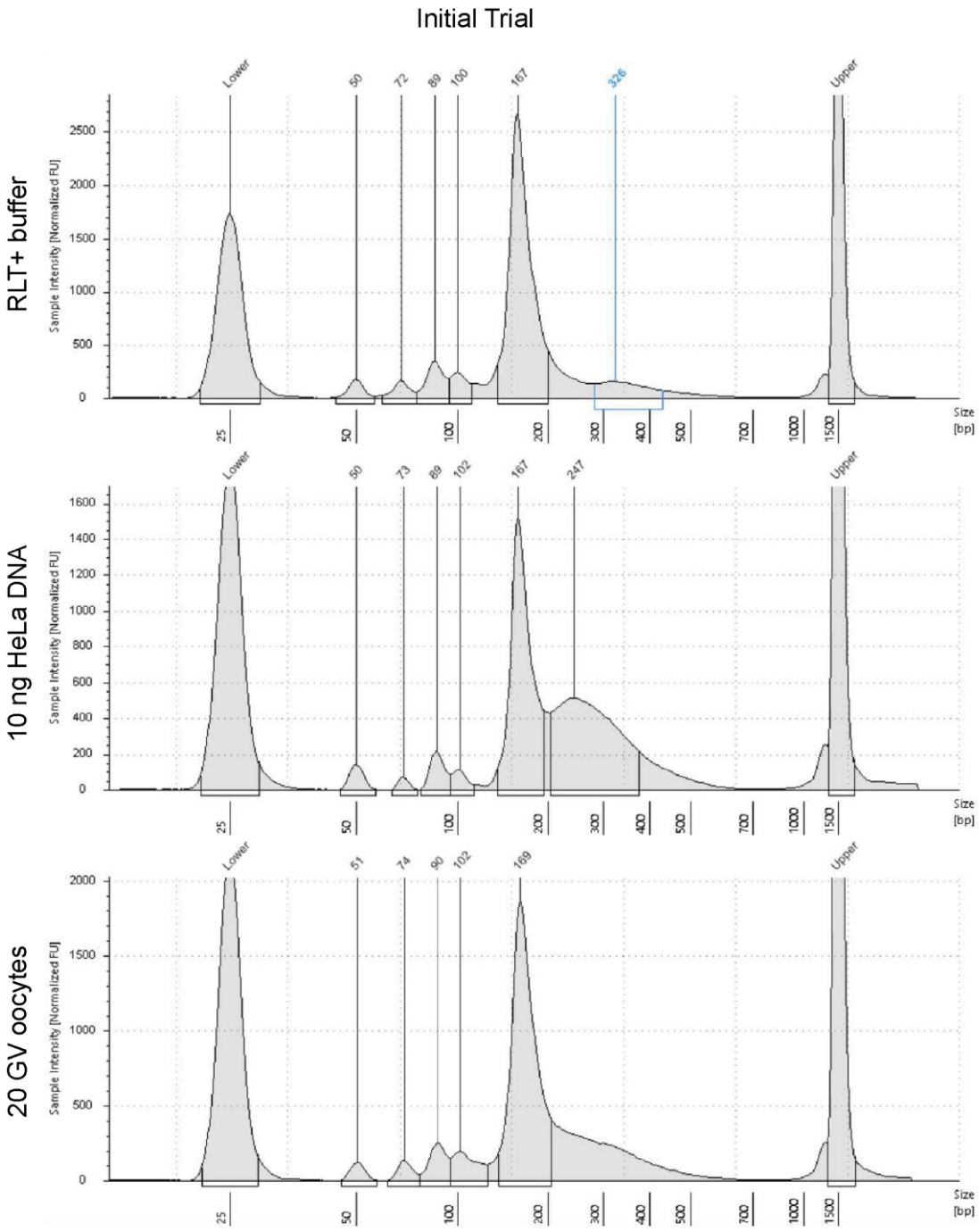
Appendix 13. BS-seq indexing primers

Indexing primers used for BS-seq, sequences synthesised by IDT detailed below.

Primer	Sequence
i7_UDI1	CAAGCAGAAGACGGCATACGAGATAACCGCGGGTGACTGGAGTTCAGACGTGT
i7_UDI2	CAAGCAGAAGACGGCATACGAGATGGTTATAAGTGACTGGAGTTCAGACGTGT
i7_UD3	CAAGCAGAAGACGGCATACGAGATCCAAGTCCGTGACTGGAGTTCAGACGTGT
i7_UDI4	CAAGCAGAAGACGGCATACGAGATTTGGACTTGTGACTGGAGTTCAGACGTGT
i7_UDI5	CAAGCAGAAGACGGCATACGAGATCAGTGGATGTGACTGGAGTTCAGACGTGT
i7_UDI6	CAAGCAGAAGACGGCATACGAGATTGACAAGCGTGACTGGAGTTCAGACGTGT
i7_UDI7	CAAGCAGAAGACGGCATACGAGATCTAGCTTGGTGACTGGAGTTCAGACGTGT
i7_UDI8	CAAGCAGAAGACGGCATACGAGATTCGATCCAGTGACTGGAGTTCAGACGTGT
i7_UDI9	CAAGCAGAAGACGGCATACGAGATCCTGAACTGTGACTGGAGTTCAGACGTGT
i7_UDI10	CAAGCAGAAGACGGCATACGAGATTTCAGGTCGTGACTGGAGTTCAGACGTGT
i7_UDI11	CAAGCAGAAGACGGCATACGAGATAGTAGAGAGTGACTGGAGTTCAGACGTGT
i7_UDI12	CAAGCAGAAGACGGCATACGAGATGACGAGAGGTGACTGGAGTTCAGACGTGT
i7_UDI13	CAAGCAGAAGACGGCATACGAGATAGACTTGGGTGACTGGAGTTCAGACGTGT
i5_UDI1	AATGATACGGCGACCACCGAGATCTACACAGCGCTAGACACTCTTTCCCTACACGAC
i5_UDI2	AATGATACGGCGACCACCGAGATCTACACGATATCGAACACTCTTTCCCTACACGAC
i5_UDI3	AATGATACGGCGACCACCGAGATCTACACCGCAGACGACACTCTTTCCCTACACGAC
i5_UDI4	AATGATACGGCGACCACCGAGATCTACACTATGAGTAACACTCTTTCCCTACACGAC
i5_UDI5	AATGATACGGCGACCACCGAGATCTACACAGGTGCGTACACTCTTTCCCTACACGAC
i5_UDI6	AATGATACGGCGACCACCGAGATCTACACGAACATACACACTCTTTCCCTACACGAC
i5_UDI7	AATGATACGGCGACCACCGAGATCTACACACATAGCGACACTCTTTCCCTACACGAC
i5_UDI8	AATGATACGGCGACCACCGAGATCTACACGTGCGATAACACTCTTTCCCTACACGAC
i5_UDI9	AATGATACGGCGACCACCGAGATCTACACCCAACAGAACACTCTTTCCCTACACGAC
i5_UDI10	AATGATACGGCGACCACCGAGATCTACACTTGGTGAGACACTCTTTCCCTACACGAC
i5_UDI11	AATGATACGGCGACCACCGAGATCTACACCGCGGTTCACACTCTTTCCCTACACGAC
i5_UDI12	AATGATACGGCGACCACCGAGATCTACACTATAACCTACACTCTTTCCCTACACGAC
i5_UDI13	AATGATACGGCGACCACCGAGATCTACACAAGGATGAACACTCTTTCCCTACACGAC

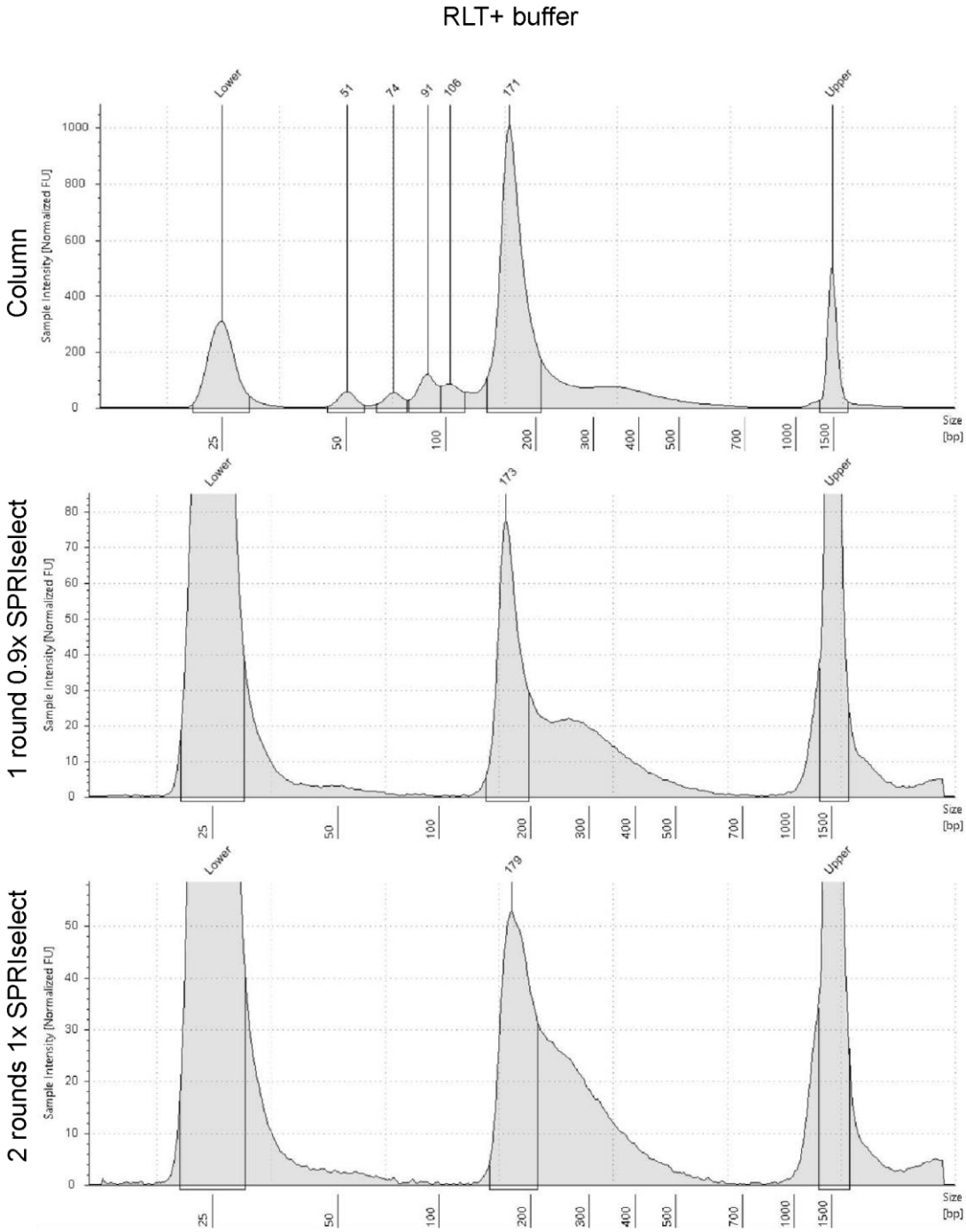
Appendix 14. Initial prep TapeStation chromatograms

High Sensitivity TapeStation chromatograms from low input BS-seq initial trial of RLT+ buffer, 10 ng HeLa DNA, and 20 WT GV oocytes.

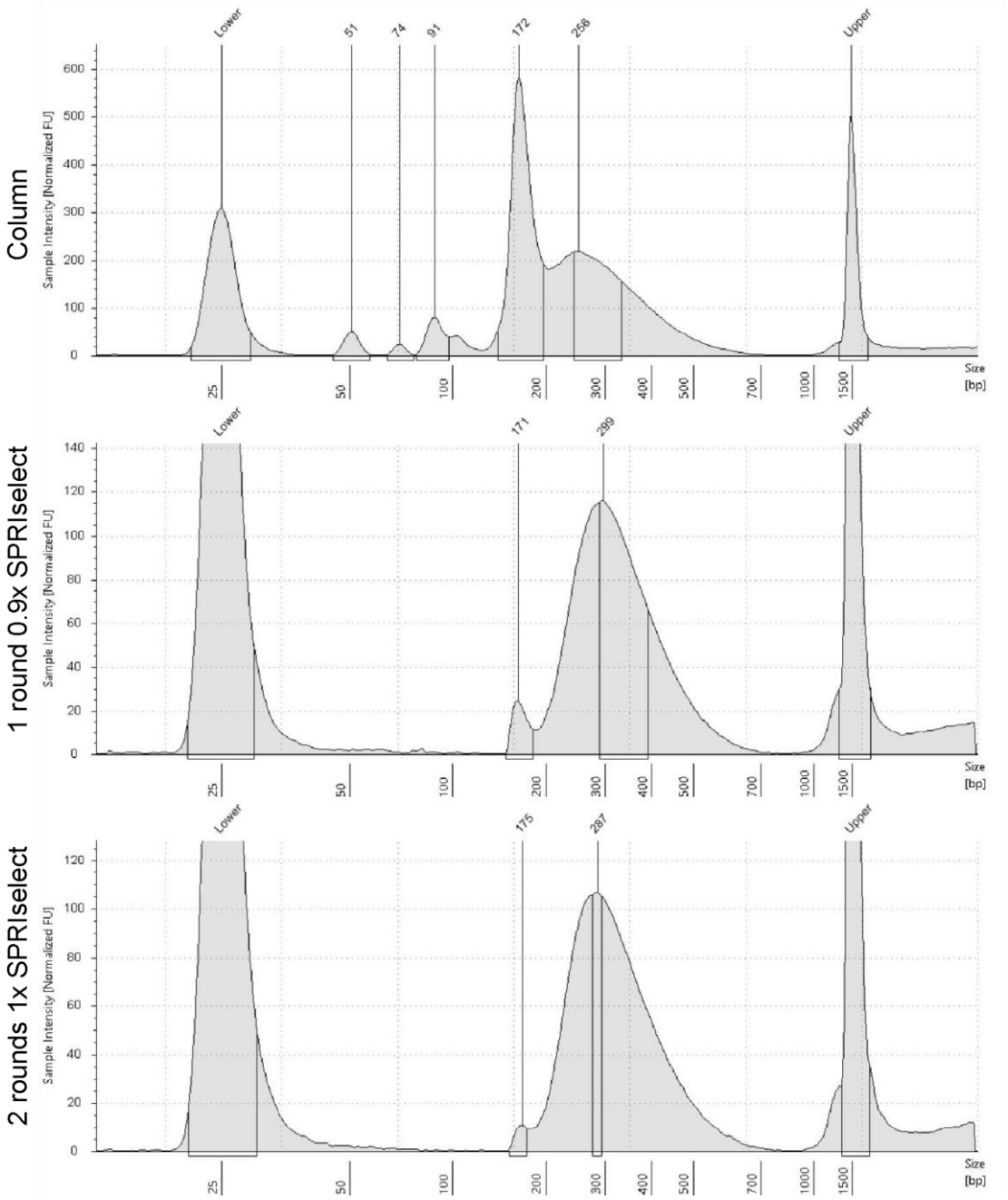


Appendix 15. Low input BS-seq bead clean up TapeStation chromatograms

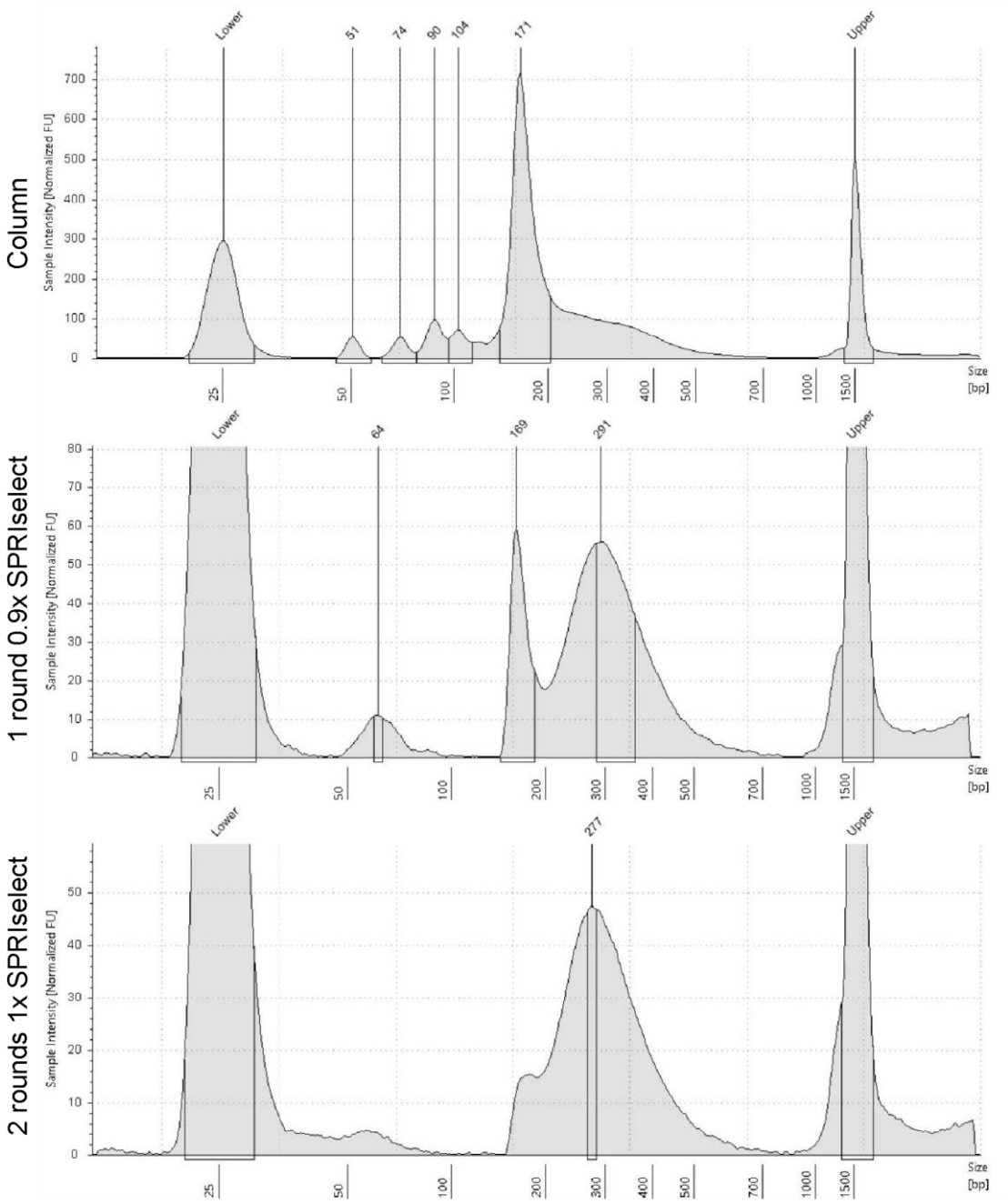
High Sensitivity TapeStation chromatograms from low input BS-seq bead clean-up trial of libraries prepared from RLT+ buffer, 10 ng HeLa DNA, and 20 WT GV oocytes. Top panel for each library shows chromatogram without bead clean-up, middle panel shows chromatogram after 1 round of 0.9x SPRI select clean-up, bottom panel shows chromatogram after 2 rounds of 1x SPRI select clean-up.



10 ng HeLa DNA

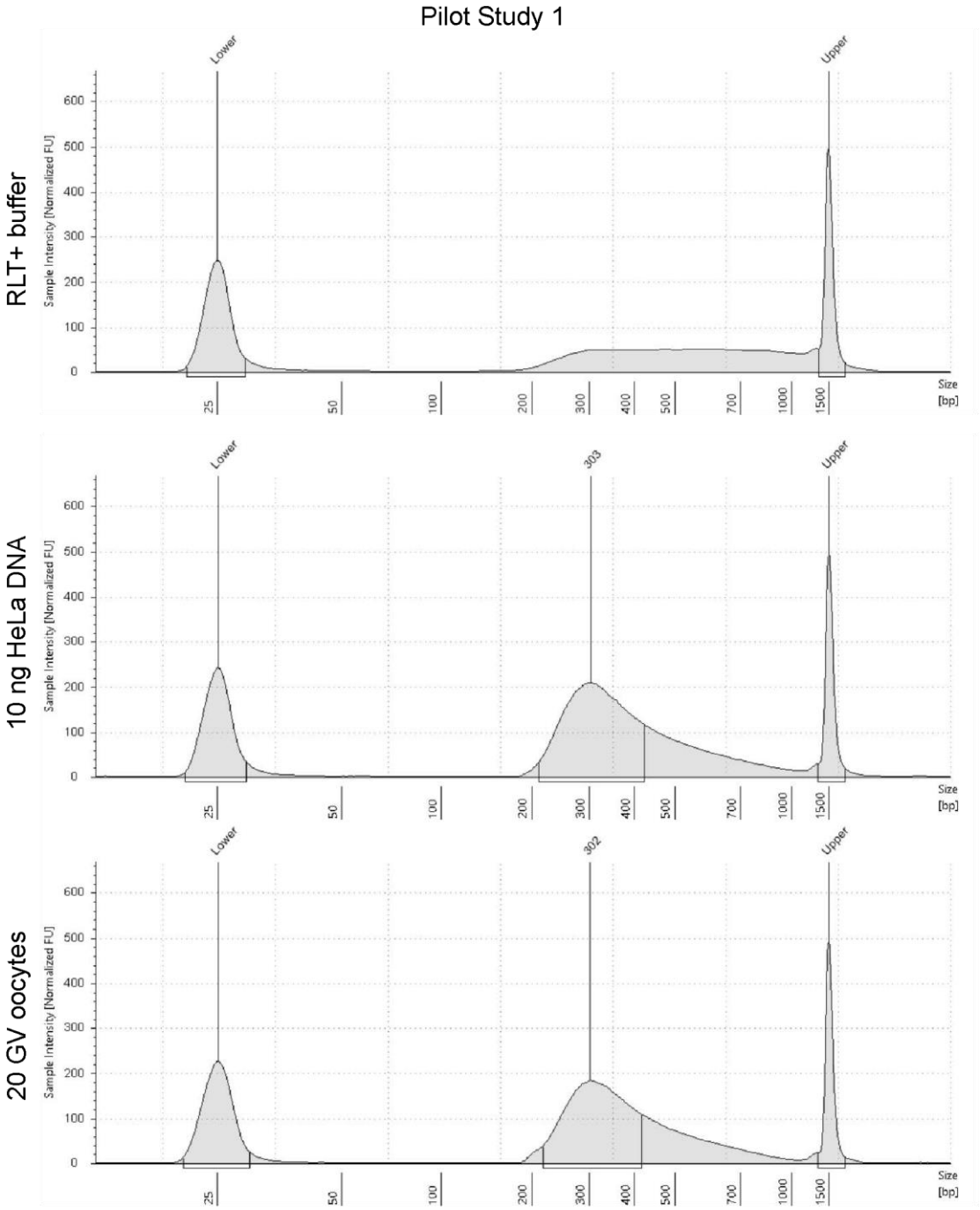


20 GV oocytes



Appendix 16. Low input BS-seq Pilot Study 1 TapeStation chromatograms

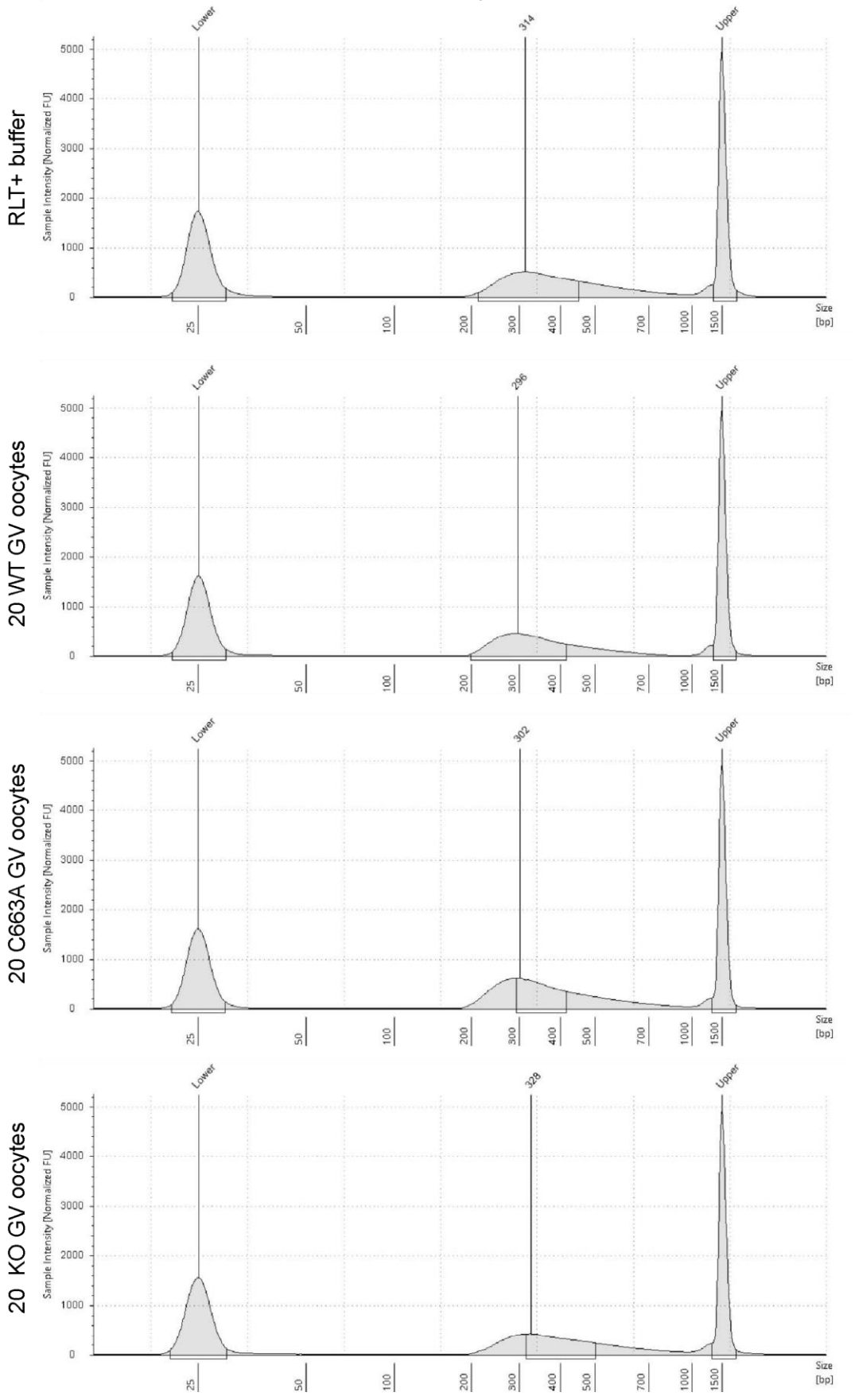
High Sensitivity TapeStation chromatograms from low input BS-seq Pilot Study 1 of RLT+ buffer, 10 ng HeLa DNA, and 20 WT GV oocytes.



Appendix 17. Low input BS-seq Pilot Study 2 TapeStation chromatograms

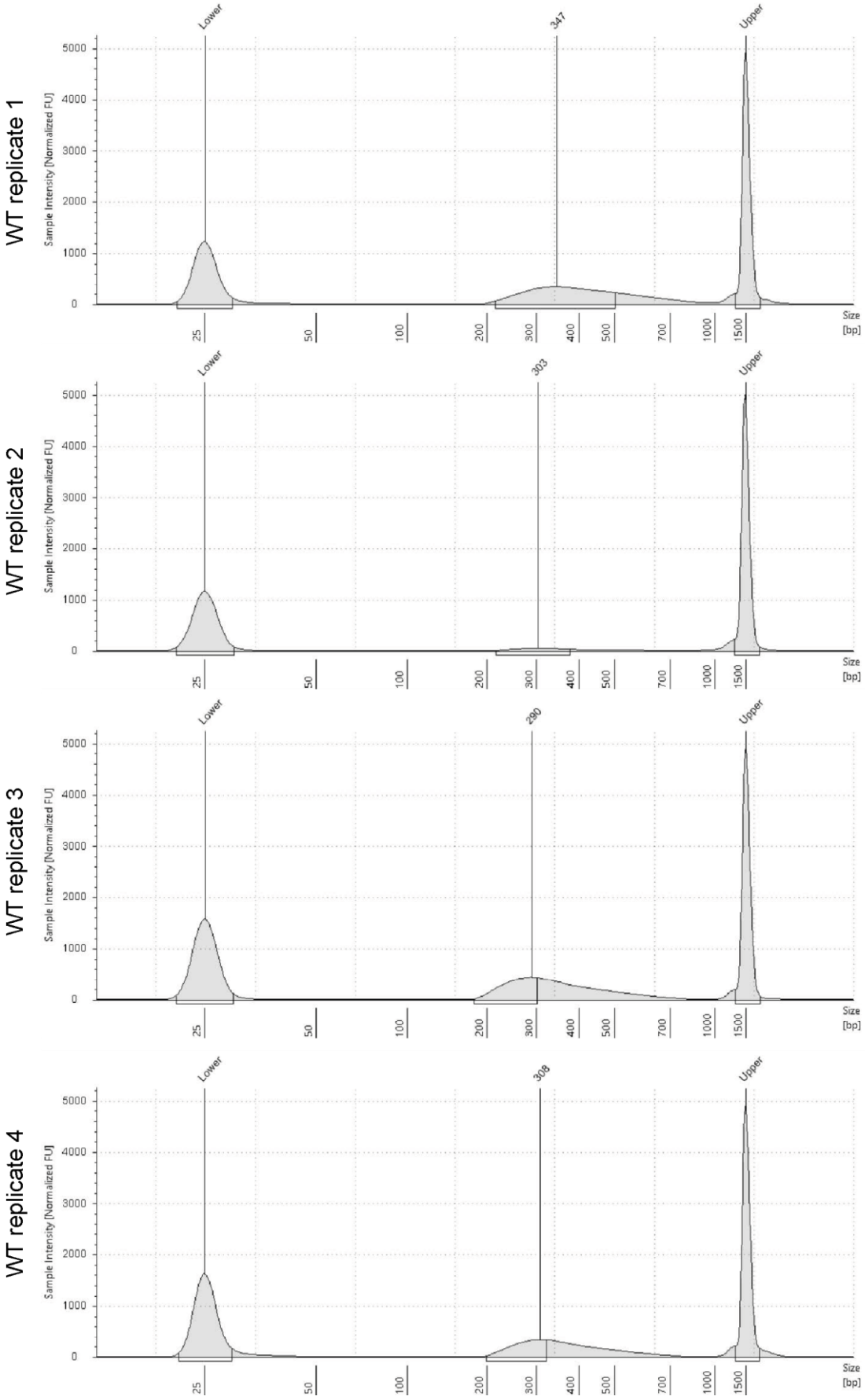
High Sensitivity TapeStation chromatograms from low input BS-seq initial trial of RLT+ buffer, 20 WT GV oocytes, 20 *Padi6*$^{C663A/C663A}$ (C663A) GV oocytes, and 20 *Padi6*$^{-/-}$ (KO) GV oocytes.

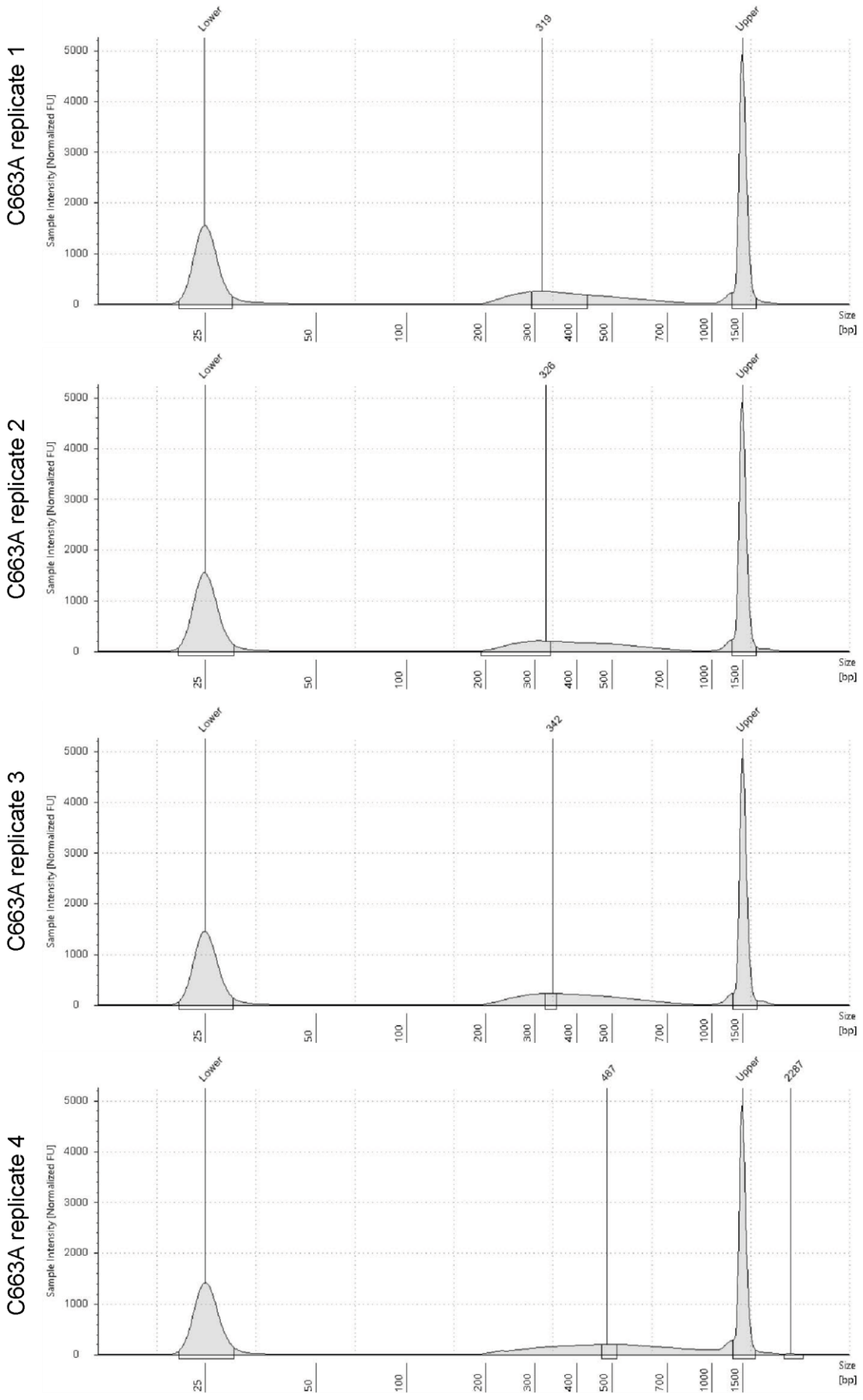
Pilot Study 2



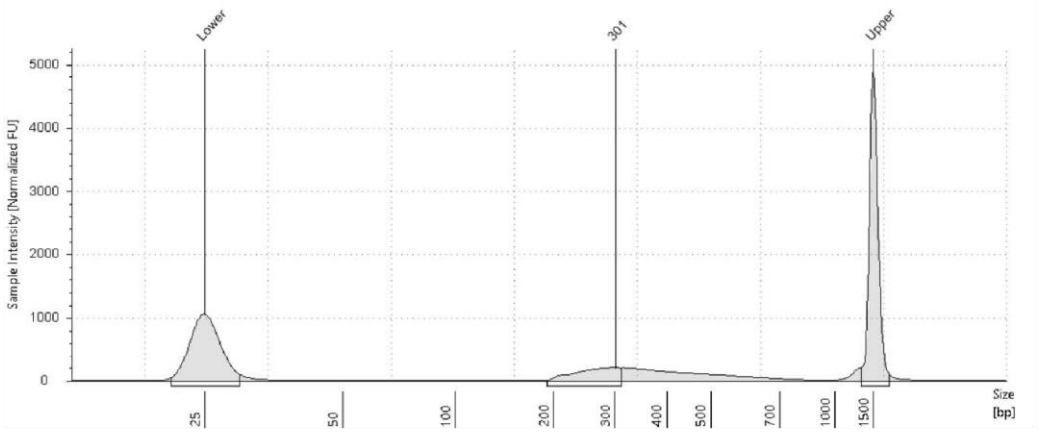
Appendix 18. Low input BS-seq Full experiment TapeStation chromatograms

High Sensitivity TapeStation chromatograms from low input BS-seq final experiment. C663A = *Padi6*$^{C663A/C663A}$, KO = *Padi6*$^{-/-}$. Four replicate libraries were prepared from 20 GV oocytes from each genetic strain, as well as a negative control library prepared from RLT+ buffer.

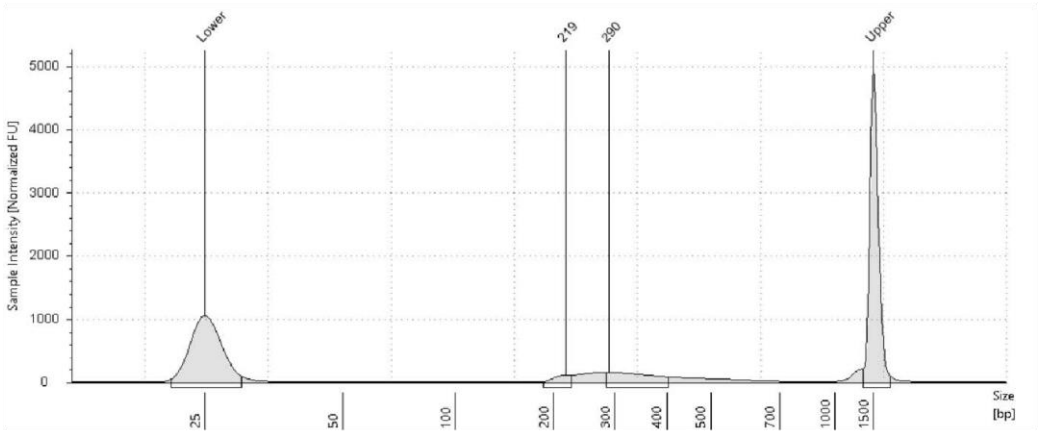

WT replicate 1
Lower
347
Upper
Sample Intensity [Normalized FU]
5000
4000
3000
2000
1000
0
25
50
100
200
300
400
500
700
1000
1500
Size [bp]
WT replicate 2
Lower
303
Upper
Sample Intensity [Normalized FU]
Size [bp]
WT replicate 3
Lower
290
Upper
Sample Intensity [Normalized FU]
Size [bp]
WT replicate 4
Lower
308
Upper
Sample Intensity [Normalized FU]
Size [bp]


C663A replicate 1
Lower
319
Upper
Sample Intensity [Normalized FU]
5000
4000
3000
2000
1000
0
25
50
100
200
300
400
500
700
1000
1500
Size
[bp]
C663A replicate 2
Lower
328
Upper
Sample Intensity [Normalized FU]
5000
4000
3000
2000
1000
0
25
50
100
200
300
400
500
700
1000
1500
Size
[bp]
C663A replicate 3
Lower
342
Upper
Sample Intensity [Normalized FU]
5000
4000
3000
2000
1000
0
25
50
100
200
300
400
500
700
1000
1500
Size
[bp]
C663A replicate 4
Lower
487
Upper
2287
Sample Intensity [Normalized FU]
5000
4000
3000
2000
1000
0
25
50
100
200
300
400
500
700
1000
1500
Size
[bp]

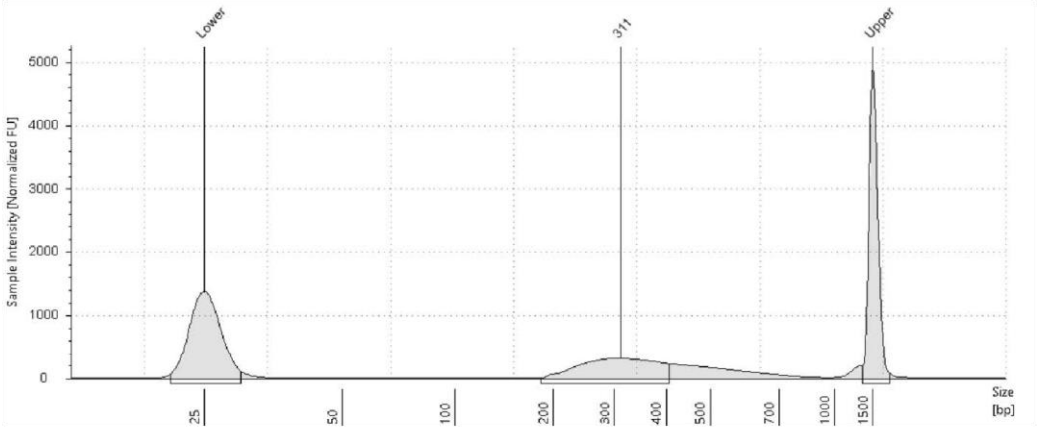
KO replicate 1



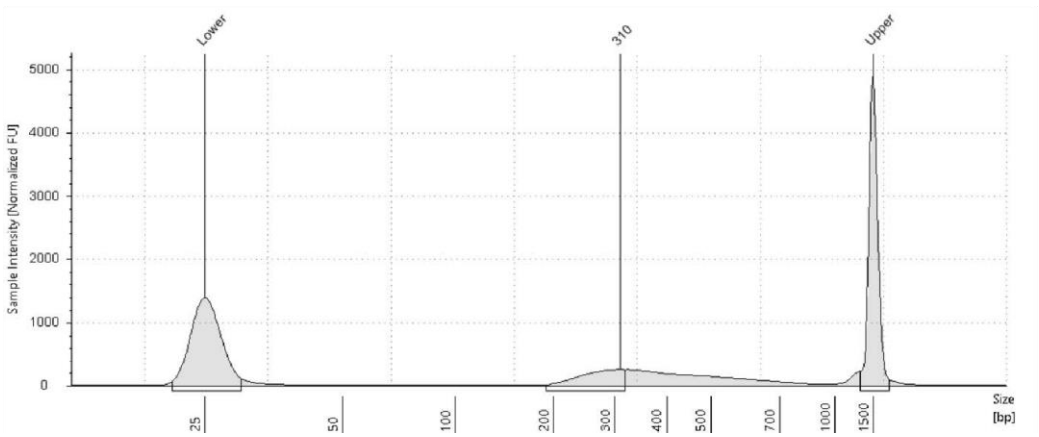
KO replicate 2



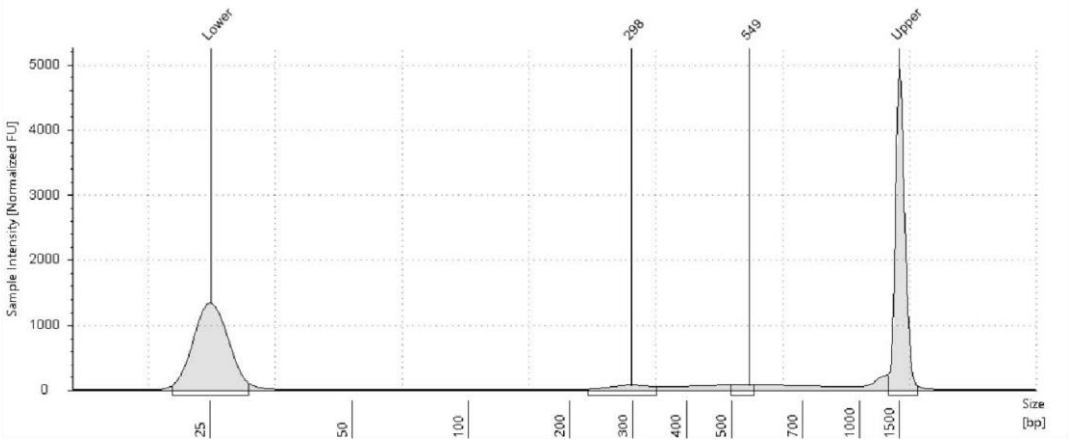
KO replicate 3



KO replicate 4



RLT+


Lower
298
549
Upper
Sample Intensity [Normalized FU]
5000
4000
3000
2000
1000
0
25
50
100
200
300
400
500
700
1000
1500
Size
[bp]

Appendix 19. Low input BS-seq Full Experiment statistics table

Full Experiment sequencing and data processing statistics. C663A = *Padi6*$^{C663A/C663A}$, KO = *Padi6*$^{-/-}$. WT replicate 2 and C663A replicate 4 that did not cluster after PCA and had poor library size distributions and sequencing statistics are coloured red.

Condition	**WT**				**C663A**				**KO**				**RLT+**
Replicate	1	2	3	4	1	2	3	4	1	2	3	4	-
Total sequences (million)	192.5	203.4	216.7	219.1	180.2	189.7	196.8	197.7	201.7	209.8	183.3	266.4	203.8
Duplicated reads (%)	61	86	54.5	57.1	65.9	60.6	60.8	74.3	60.2	64.2	61.3	62.5	88.1
GC content (%)	29	27	29	29	28	29	29	26	29	30	28	20	32
Mapping to mouse genome													
Mapping efficiency (%)	44.2	52.2	42.4	42.6	49.6	41.6	42.4	30	40.3	39	46.6	43.2	0
Reads removed in deduplication (%)	77.5	96.4	70.9	72.9	81.7	75.9	76.2	90.5	75.6	78.7	78.3	77.9	-
Sequences left after deduplication (million)	19.7	3.8	23.6	23.4	19.6	21.8	18.1	5.3	19.2	16.3	20.3	19.9	-
CpG methylation (%)	38.9	38.9	39.3	39.3	39.1	39.1	38.8	38.4	40.4	40.4	39.4	40.2	-
CHG methylation (%)	4	3.8	4	4	4.1	4.1	4.1	4	4.4	4.4	4.3	4.4	-
CHH methylation (%)	3.8	4	3.9	3.9	3.9	3.9	3.9	4.1	4.2	4.2	4	4.1	-

Appendix 20. Permissions

Figure 1.11:

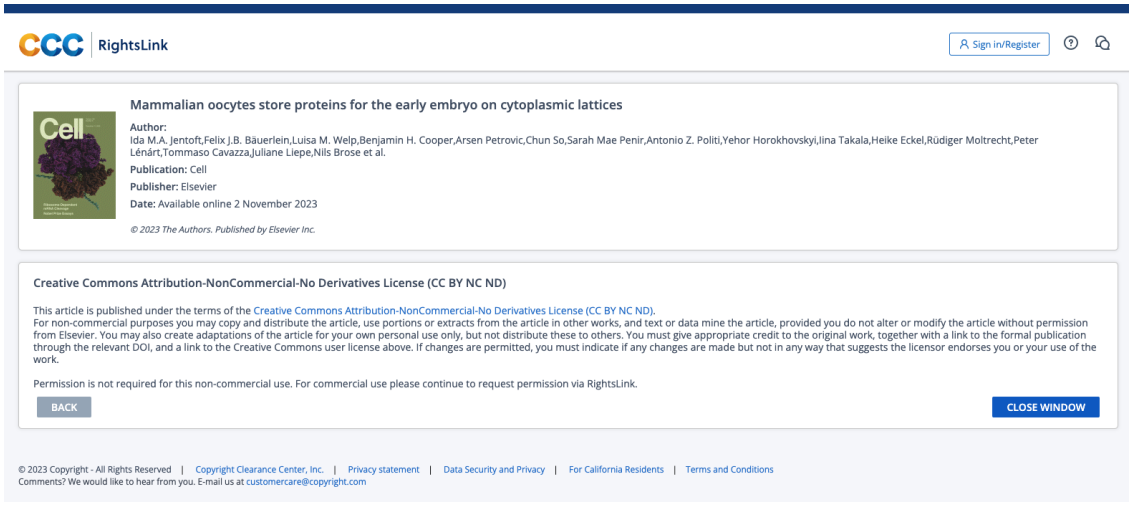


Figure 1.17c:



This is a License Agreement between Jack Williams ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	28-Nov-2023	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1421018-1		
ISSN	1545-4509	**Publisher**	ANNUAL REVIEWS
		Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Annual review of biochemistry	**Publication Type**	e-Journal
		Start Page	135
Article Title	Role of Mammalian DNA Methyltransferases in Development.	**End Page**	158
		Issue	1
Date	01/01/1932	**Volume**	89
Language	English	**URL**	http://arjournals.annualreviews.org/loi/biochem
Country	United States of America		
Rightsholder	Annual Reviews, Inc.		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	**Distribution**	U.K. and Commonwealth (excluding Canada)
Number of Charts / Graphs / Tables / Figures Requested	1	**Translation**	Original language of publication
Format (select all that apply)	Electronic	**Copies for the Disabled?**	Yes
		Minor Editing Privileges?	Yes
Who Will Republish the Content?	Academic institution	**Incidental Promotional Use?**	No
Duration of Use	Life of current edition	**Currency**	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Investigation into the function of PADI6 in early embryonic development	**Institution Name**	Imperial College London
		Expected Presentation Date	2023-12-05
Instructor Name	Jack Williams		

ADDITIONAL DETAILS

Order Reference Number	N/A	**The Requesting Person / Organization to Appear on the License**	Jack Williams

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Figure 1d, Figure 1e	**Title of the Article / Chapter the Portion Is From**	Role of Mammalian DNA Methyltransferases in Development.
Editor of Portion(s)	Chen, Zhiyuan; Zhang, Yi	**Author of Portion(s)**	Chen, Zhiyuan; Zhang, Yi
Volume / Edition	89	**Issue, if Republishing an Article From a Serial**	1
Page or Page Range of Portion	135-158		
		Publication Date of Portion	2020-06-20