IDENTIFICATION AND CHARACTERISATION OF ALTERNATIVE FORMS OF SETD2/HYPB

(SET domain-containing protein 2 / Huntingtin yeast partner B)



Benjamin Mark Lee

Department of Biochemistry and Wolfson College
University of Oxford

Supervisor:	Prof Louis C Mahadevan
Co-supervisor:	Dr Erika J Mancini

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Trinity term 2011

Word count: 47,143

Identification and characterisation of alternative forms of SETD2/HYPB

(SET domain–containing protein 2 / Huntingtin yeast partner B)

Benjamin Mark Lee

A thesis for the degree of Doctor of Philosophy

Trinity term, 2011 : Wolfson College, University of Oxford

Abstract

SETD2/HYPB (SET domain–containing protein 2 / Huntingtin yeast partner B) is the predominant lysine methyltransferase in mammals that mediates histone H3 lysine-36 (H3K36) trimethylation, which is associated with transcription elongation and RNA splicing. SETD2 is further implicated in p53 function, vascular development, cancer progression and, through Huntingtin-interaction, Huntington's disease. Although different transcripts and putative protein isoforms have been detected previously, their identity, function and significance have not been rigorously investigated.

This thesis aims to identify and characterise endogenous transcripts and protein isoforms of SETD2 in mouse fibroblasts. Affinity-purified N- and C-terminal antibodies specifically detected the ≈290 kDa methyltransferase ($p290^{SETD2}$), verified by RNAi, in addition to N-terminal-specific ≈120 kDa protein, and C-terminal-specific forms at ≈140 and ≈66 kDa (p66), which all appeared too stable to deplete by transient siRNA transfection. Conserved in human and mouse cells, immunodetection of p66 exhibited unusual requirement for denaturation with urea at 95°C. Subcellular fractionation revealed distinct extraction properties of putative isoforms and facilitated partial purification of p66 for proteomic analysis. Co-fractionation and co-migration by two-dimensional gel electrophoresis of p66 detected by two independent C-terminal antibodies suggested it represents a novel C-terminal-specific isoform. Reverse transcription–PCR and DNA-sequencing demonstrated the existence of multiple, alternatively-spliced *Setd2* transcripts that plausibly generate truncated proteins. A transcript variant containing a novel complete open-reading-frame, consistent for p66 generation, was identified. Its ectopic expression in mouse fibroblasts produced a distinct SETD2 isoform, whose physical and extraction characteristics were studied in comparison with endogenous immunoforms.

In summary, this thesis demonstrates that multiple alternatively-spliced transcripts arise from the *Setd2* gene, consistent with immunodetection of several C- and N-terminal-specific putative SETD2 isoforms, additional to the H3K36 methyltransferase. Verification of these isoforms by independent methods would have implications for proposed interactions and function of SETD2 in transcription, epigenetics, cancer development and Huntington's disease.

Acknowledgments

This thesis would never have materialised were it not for the help and support of many people.

First and foremost I would like to thank Prof Louis Mahadevan, my supervisor, for not only giving me the opportunity to study in such a fantastic lab, but also the intellectual freedom to conceive and develop this project along its current lines. His unwavering support, advice encouragement and criticisms are second to none, and will inspire me for many years to come.

I would also like to thank Dr Erika Mancini, my co-supervisor, for her continued support and interest in this project (despite my inability to keep in contact and the structural biology never coming to light!). I'd also like to thank Dr Mary Gregoriou for her advice and support in the early decisions of my DPhil.

None of this work would have been possible without the help of everyone in the Nuclear Signalling Laboratory over the last few years: Dr Catherine Hazzalin, Dr Alison Clayton, Dr Duen-Wei Hsu, Dr Nicholas Crump, Edgar Pogna, Ya-Ting Han, Dr Guan-Hei Lee, Jonathan Neve and Caroline Schneider. It's been fantastic to work in such an intellectually stimulating environment, but more so, I feel very lucky to have been surrounded by people always happy to share their time to discuss experimental and intellectual issues. Thank you all for your support, advice and encouragement—and awesome lab-meeting cakes.

I am grateful to Jonathan Neve for performing some of the RT-PCR analysis, which are included in Chapter 5. I would like to thank Dr Mukram Mackeen and Prof Benedikt Kessler (IMM), Dr Benjamin Thomas and Gabriela Ridlova (Dunn School), who performed the mass spectrometry. I would like to thank several people for specific reagents: Maria Hoellerer for the p3×FLAG-CMV-10 vector, Nikolay Kukushkin for the pcDNA3−NHK−HA construct, and Nuno Nunes for the 293T cells.

I am very grateful to the Wellcome Trust for their generous financial support over the last four years.

I also wouldn't be where I am today without the continued support of my parents, John and Pauline, and brother, Tom. Thanks for always being there for me.

Thanks to all the people I've met during my time here in Oxford for making the last several years the best I've ever had. Especially, I want to thank Vicki Patterson, Matt Dicks, Kimberley Bryon, Mike Dodd and Sarah Beal—thanks for sharing all the ups and downs!

Finally, Sean: this thesis has been as much a part of your life here in Oxford as it has mine over the last four years. Thank you for your constant patience, understanding and encouragement; and for taking care of all of the plans over the last year: I'm not sure I could have made it without you.

Publications

Morra R, **Lee BM**, Shaw H, Tuma R, Mancini EJ (2011) Concerted action of the PHD, chromo and motor domains regulates the human chromatin remodelling ATPase CHD4 (*submitted*).

Lee BM & Mahadevan LC (2009) Stability of histone modifications across mammalian genomes: Implications for 'epigenetic' marking. *J Cell Biochem*, **108** (1), 22–34.

Abbreviations

Abbreviation	Definition
4IPBA	4-iodophenylboronic acid
APS	Ammonium persulphate
AS	Alternative splicing
AWS	Associates with SET
AU	Acetic acid–urea
bp	Base-pair
CC	Coiled-coil
ccRCC	Clear cell renal cell carcinoma
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation–sequencing
CMV	Cytomegalovirus
CTD	Carboxy-terminal domain
Da	Dalton
DDC	Differential detergent extraction and centrifugation
DMEM	Dulbecco's modified Eagle's medium
DMPC	Dimethylpyrocarbonate
DNase	Deoxyribonuclease
dNTPs	Deoxyribonucleotides
DOC	Deoxycholate
DRB	5,6-dichloro-1-β-D-ribofuranosylbenzimidazole
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ESC	Embryonic stem cell
EST	Expressed sequence tag
EtBr	Ethidium bromide
FCS	Foetal calf serum
FT	Flow-through
HA	Haemagglutinin
HD	Huntington's disease
HDAC	Histone deacetylase
HEAT	Huntingtin/Elongation factor 3/Protein phosphatase 2A/TOR1
HEK	Human embryonic kidney
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulphonic acid
hnRNP	Heterogeneous nuclear ribonucleoprotein
HP1	Heterochromatin protein 1
HRP	Horseradish peroxidase
HYPB	Huntingtin yeast partner B
Hypo	Hypotonic extract
IWS1	Interacts with Spt6 protein 1
kb	Kilobase

Abbreviation	Definition
kDa	Kilodalton
KDM	Lysine demethylase
KMT	Lysine methyltransferase
LB	Lysogeny broth
MALDI	Matrix-assisted laser desorption ionisation
miRNA	MicroRNA
MMLV	Moloney murine leukaemia virus
MOPS	3-(N-morpholino)propanesulphonic acid
mRNP	Messenger ribonucleoprotein
MS	Mass spectrometry
NAse	Nuclease digest extract
NCS	New-born calf serum
NMD	Nonsense-mediated decay
NMR	Nuclear magnetic resonance
NSD	Nuclear receptor–binding SET domain
ORF	Open reading frame
P	Pellet
PAGE	Polyacrylamide gel electrophoresis
PAS	Polyadenylation site
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
pCTD	Phosphorylated carboxy-terminal domain
PEG	Polyethylene glycol
PHD	Plant homeodomain
PRC2	Polycomb repressive complex 2
PTB	Polypyrimidine tract–binding protein
pTEFb	Positive transcription elongation B
PTM	Post-translational modification
PVDF	Polyvinylidene fluoride
QP-[0]	Glutamine and proline–rich, uncharged region
RACE	Rapid amplification of cDNA ends
RNAi	RNA interference
RNAPII	RNA polymerase II
RNA-seq	Ribonucleic acid–sequencing
RNase	Ribonuclease
rRNA	Ribosomal RNA
RS	Arginine–serine
RT	Reverse transcription/transcriptase
RT-PCR	Reverse transcription–polymerase chain reaction
SA-PMP	Streptavidin–paramagnetic particle
SCA1	Spinocerebellar ataxia type 1
SET	Suppressor of variegation 3–9, Enhancer of *Zeste*, *Trithorax*
SETD2	SET domain–containing protein 2
SDS	Sodium dodecyl sulphate
SILAC	Stable isotope labelling with amino acids in cell culture
siRNA	Short interfering ribonucleic acid

Abbreviation	Definition
snRNA	Small nuclear ribonucleic acid
SN / S	Supernatant
SRI	Set2–Rpb1 interacting
SSC	Saline–sodium citrate
TAE	Tris-acetate-EDTA
TBP	TATA box-binding protein
TCA	Trichloroacetic acid
TE	Tris–EDTA
TEMED	*N, N, N', N'*-tetramethyl-ethane-1,2-diamine
TOF	Time-of-flight
Tris	Tris(hydroxymethyl)aminoethane
TSS	Transcriptional start site
TD	Triton X-100–Sodium deoxycholate extract
TX	Triton X-100 extract
TX–DOC	Triton X-100–Sodium deoxycholate buffer or extract
UPLC	Ultra performance liquid chromatography

Histone modifications are defined by Brno nomenclature (Turner, 2005), *i.e.* H3K36me3 represents trimethylation of lysine-36 of histone H3.

The following abbreviations are used to refer to immunoreactive bands:

p*X*$^{A/B/C}$ protein of approximate apparent molecular mass *X* (kDa), detected with *α-A/B/C* antibodies.

p290^{SETD2} confirmed SETD2 isoform of approximately 290 kDa.

The following notation abbreviates splicing patterns within transcripts:

12–15 indicates a transcript where exon 12 splices directly to exon 15 (exons 13 and 14 are excluded).

12»15 abbreviates 12–13–14–15, a transcript where exon 12 splices to exon 13, which splices to exon 14, and finally to exon 15.

8–9.5 signifies a transcript where exon 8 splices to exon 9 with retention of intronic sequence following exon 9.

Contents overview

CHAPTERS

1 Introduction

2 Materials and methods

3 Immunodetection of p290^{SETD2} and putative SETD2 isoforms

4 Generation of a p66-enriched fraction for two-dimensional gel electrophoresis and proteomic analysis

5 Identification of alternative forms of murine *Setd2* transcripts

6 Ectopic expression of a SETD2 isoform encoded by a variant transcript and comparison with p66$^{C/CT21}$

7 Final discussion and future directions

APPENDICES

A Supplementary results for Chapter 3
Immunodetection of p290^{SETD2} and putative SETD2 isoforms

B Supplementary results for Chapter 4
Generation of a p66-enriched fraction

C Supplementary results for Chapter 5
Identification of alternative Setd2 *transcripts*

REFERENCES

Table of contents

Abstract ii
Acknowledgments iii
Publications iv
Abbreviations v
Contents overview viii
Table of contents ix
List of figures xviii
List of tables xxi

CHAPTER 1 INTRODUCTION

1.1. The complexity of eukaryotic gene regulation 1
1.1.1. Chromatin and modification of nucleosome structure 1
1.1.2. Histone modification 2

1.2. Overview of SET domain–containing protein 2 (SETD2) 6
1.2.1. Identification of SETD2 6
1.2.2. Overview of SETD2 protein structure and interactions 7

1.3. SETD2 as lysine methyltransferase 12
1.3.1. SETD2 mediates trimethylation of histone H3K36 13
1.3.1.1. SETD2 as lysine methyltransferase in vitro 13
1.3.1.2. SETD2 mediates H3K36me3 in vivo 14
1.3.2. Genome-wide localisation of H3K36 methylation 15
1.3.2.1. H3K36me3 marks chromatin transcribed by RNA polymerase II 15
1.3.2.2. H3K36me3 differentiates exonic versus intronic chromatin 16
1.3.2.3. SETD2-mediated H3K36me3 is associated with RNAPII elongation 17
1.3.3. Mechanism of SETD2-dependent generation of H3K36me3 20
1.3.3.1. SETD2 interacts directly with the carboxy-terminal domain of RNA polymerase II 21
1.3.3.2. The relationship between CTD phosphorylation and SETD2 function is contentious 23
1.3.3.3. The IWS1–SPT6 complex mediates SETD2-dependent H3K36me3 24
1.3.3.4. H3K36 trimethylation in vivo requires heterogeneous nuclear ribonucleoprotein L 25
1.3.3.5. Influence of RNA splicing on H3K36me3 generation 25

1.4. Proposed functions of SETD2-mediated H3K36me3 29
1.4.1. Roles of H3K36 methylation 29
1.4.2. Roles in transcription elongation: suppression of 'cryptic' transcripts? 32

1.4.2.1. Set2–Rpd3(S) pathway and regulation of intragenic histone acetylation 32
1.4.2.2. SETD2–KDM5A pathway suppresses intragenic H3K4me3 33
1.4.3. Regulation of RNA processing 34
1.4.4. H3K36 methylation inhibits H3K27 methylation 36
1.4.5. Relationship with DNA methylation and heterochromatin 37
1.5. Additional functions of SETD2 38
1.5.1. SETD2 as a putative transcription factor 39
1.5.2. SETD2 regulates the p53 transcription network 40
1.5.3. SETD2 is required for vasculogenesis and embryonic development 41
1.5.4. *SETD2* as a putative tumour suppressor gene 41
1.5.5. SETD2 interacts with Huntingtin 42
1.6. Diversity of transcriptional and translation products of the *SETD2* gene 45
1.6.1. Understanding biological diversity between transcriptome and proteome 45
1.6.2. Evidence for multiple forms of SETD2 47
1.7. Aims and overview of thesis 50

CHAPTER 2 MATERIALS AND METHODS

2.1. Materials and general procedures 51
2.1.1. PCR primers 51
2.1.2. Peptides and antibodies 51
2.2. Peptide affinity–purification of antisera 51
2.2.1. Generation of peptide-coupled affinity columns 51
2.2.2. Affinity purification of α-SETD2 antisera 52
2.3. Cloning of epitope-tagged *Setd2* expression constructs 53
2.3.1. Plasmids and PCR primers 53
2.3.2. PCR amplification of *Setd2*(13»21) and cloning into pJET1.2/blunt vector 53
2.3.3. Sub-cloning of *Setd2*(13»21) into epitope-tagging vectors 55
2.4. Cell culture and treatment 55
2.5. Transient transfection of 10T½ cells with small-interfering RNA (siRNA) 56
2.6. Transient transfection of 10T½ cells by electroporation 56

2.7. Preparation of cell lysates and subcellular fractions 58

2.7.1. Phosphatase and protease inhibitors 58

2.7.2. TX–DOC lysis and extraction of TX-DOC-resistant material 58

2.7.3. Subcellular fractionation by nuclease digestion and high-salt extraction 59

2.7.4. Differential detergent fractionation to produce a $p66^{C/CT21}$-enriched fraction 60

2.7.5. Preparation of acid extracts 61

2.7.6. Trichloroacetic acid (TCA) precipitation 61

2.8. Protein electrophoresis 62

2.8.1. Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) 62

2.8.2. Acetic acid–urea polyacrylamide gel electrophoresis (AU-PAGE) 63

2.8.3. Two-dimensional AU/SDS–PAGE 64

2.9. Coomassie Brilliant Blue staining 65

2.10. Western transfer 65

2.10.1. Western transfer after SDS-PAGE 66

2.10.2. Western transfer after AU-PAGE 66

2.10.3. Western transfer after Coomassie staining 66

2.10.4. Ponceau S staining 66

2.11. Immunoblotting 67

2.11.1. Peptide–dot blotting 67

2.11.2. Immunoblotting 67

2.11.3. Enhanced chemiluminescence 69

2.11.4. Peptide competition 69

2.11.5. Quantification and mass estimation of immunoreactive proteins 69

2.12. In-gel digestion and mass spectrometry 70

2.12.1. In-gel trypsin digestion and C18 purification 70

2.12.2. In-gel digestion with chymotrypsin 71

2.12.3. Mass spectrometry 72

2.13. RNA extraction 72

2.13.1. Total RNA extraction 72

2.13.2. Poly(A) RNA selection 73

2.14. RNA gels and electrophoresis 73

2.15. Northern blotting 74

2.15.1. Northern transfer to nylon membranes 74

2.15.2. Preparation of radiolabelled probes 74

2.15.3. Hybridisation of radiolabelled probes 75

2.16. Reverse transcription–PCR (RT-PCR) 75

2.16.1. DNase treatment 75

2.16.2. Reverse transcription 76

2.16.3. PCR 76

2.17. DNA extraction, electrophoresis and sequencing 78

2.17.1. Plasmid purification 78

2.17.2. DNA electrophoresis and gel purification 78

2.17.3. DNA sequencing 78

CHAPTER 3 IMMUNODETECTION OF P290^{SETD2} AND PUTATIVE SETD2 ISOFORMS

3.1. Introduction 79

3.1.1. Antibodies as tools to investigate the identity and function of endogenous proteins 79

3.1.2. RNA interference to investigate protein identity and function 80

3.2. Objectives 81

3.3. Results 82

3.3.1. Affinity-purified rabbit antibodies to an N-terminal region of SETD2 detect p290^{N} and p120^{N} proteins in murine fibroblasts 83

3.3.1.1. Affinity-purification of N-terminal-specific SETD2 antibodies 83

3.3.1.2. Transfer conditions influence the relative intensities of p290^{N} and p120^{N} 87

3.3.2. Affinity-purified sheep antibodies to a C-terminal region of SETD2 specifically detect multiple proteins in murine fibroblasts 89

3.3.3. p66^{C} is detected by C-terminal α-SETD2 antibody—when extracts are heated in urea 93

3.3.4. Characterisation of the minimal conditions required for p66^{C} detection 99

3.3.5. Conservation of α-SETD2-reactive proteins in multiple murine and human cell types 105

3.3.6. Transient siRNA transfection reduces p290$^{N/C}$ expression exclusively 110

3.3.7. p290^{SETD2} is relatively labile compared to additional immunoreactive proteins 115

3.4. Discussion 118

3.4.1. Summary of results 119

3.4.2. Detection of putative SETD2 isoforms in addition to p290^{SETD2} 120

3.4.3. Feasibility of transient siRNA transfection to investigate identity of α-SETD2 immunoforms 122

3.4.4. Stability of p290^{SETD2} 123

3.4.5. Detection of p66^{C} after extensive thermal denaturation with urea 124

3.4.5.1. Urea mediates physical and chemical changes to protein structure 124

3.4.5.2. Apparent rarity of proteins exhibiting detection characteristics of p66^{C} 129

3.4.5.3. Consequences of the denaturation requirement for identification of p66^{C} 130

3.5. Conclusions 131

CHAPTER 4 GENERATION OF A P66-ENRICHED FRACTION FOR TWO-DIMENSIONAL GEL ELECTROPHORESIS AND PROTEOMIC ANALYSIS

4.1. Introduction 132

4.2. Objectives 133

4.3. Results 133

4.3.1. Differential extraction of α-SETD2-reactive proteins by nuclease digestion and high-salt extraction 133

4.3.2. Generation of a p66^{C}-enriched fraction by differential detergent extraction and centrifugation 140

4.3.3. p66 is detected by two C-terminal region α-SETD2 antibodies 149

4.3.4. AU-PAGE permits identification of p66^{C} by mass spectrometry in the absence of thermo-urea denaturation 150

4.3.5. Two-dimensional AU/SDS-PAGE confirms p66^{C} detection in the absence of thermo-urea denaturation 154

4.3.6. Mass spectrometry identifies CKAP4 as the predominant constituent of the p66$^{C/CT21}$-enriched spot 159

4.4. Discussion 167

4.4.1. Summary of results 167

4.4.2. Subcellular fractionation of p290^{SETD2} and putative SETD2 isoforms 167

4.4.3. The complexed nature of p66$^{C/CT21}$ 170

4.4.4. Proteomic identification of $p66^{C/CT21}$ 171

4.5. Conclusion 174

CHAPTER 5 IDENTIFICATION OF ALTERNATIVE FORMS OF MURINE *SETD2* TRANSCRIPTS

5.1. Introduction 176

5.1.1. Challenges associated with defining transcript variants within complex transcriptomes 176

5.1.2. Functional consequences of alternative transcript generation 177

5.1.3. Current understanding of mammalian *Setd2* transcripts 178

5.2. Objectives 178

5.3. Terminology 178

5.4. Results 181

5.4.1. Multiple *Setd2* transcripts are detected by northern blotting using 3' region probes 181

5.4.2. Identification of a highly conserved genomic sequence between exons 12 and 13 184

5.4.3. Identification of alternative *Setd2* transcripts that contain exon 12b spliced downstream to exon 13 188

5.4.4. Identification of alternative *Setd2* transcripts that contain exon 12b spliced upstream to exon 12 191

5.4.5. Major transcripts across the 12–13 region exclude exon 12b 191

5.4.6. Discovery of exon 12c confirms existence of mutually-exclusive splicing events involving exon 12b 194

5.4.7. Alternative 3' splice sites define two 5' boundaries of exon 14 196

5.4.8. Isolation of 'retained-intron' transcripts 197

5.4.9. Identification of a mouse brain cDNA containing a complete novel ORF 201

5.5. Discussion 207

5.5.1. Summary of results 207

5.5.2. How many alternative *Setd2* transcripts exist? 208

5.5.2.1. Existence of the canonical 1»21 transcript 208

5.5.2.2. Evidence for additional transcripts in previously published data 208

5.5.2.3. The extent of transcript variation from RT-PCR data 209

5.5.3. The biological significance of alternative *Setd2* transcripts 211

5.5.3.1. The fate of transcripts: translation and turnover 211
5.5.3.2. Retained introns predict premature termination 212
5.5.3.3. Exon 12b and 12c 213
5.5.3.4. Novel ORF within 12−12b−13»21 cDNA 214

5.6. Conclusion 216

CHAPTER 6 ECTOPIC EXPRESSION OF A SETD2 ISOFORM ENCODED BY A VARIANT TRANSCRIPT AND COMPARISON WITH P66$^{C/CT21}$

6.1. Introduction 219

6.2. Results 220

6.2.1. Design of constructs to express epitope-tagged SETD2(13»21) proteins 220

6.2.2. SETD2(13»21) and p66$^{C/CT21}$ co-fractionate by nuclease–high-salt extraction but exhibit different electrophoretic migration by SDS-PAGE 223

6.2.3. SETD2(13»21) does not recapitulate detergent-extractability of p66$^{C/CT21}$ 230

6.3. Discussion 235

6.3.1. Summary of results 236

6.3.2. Anomalous migration of SETD2(13»21) 236

6.3.3. Subcellular fractionation of SETD2(13»21) 238

6.3.4. Implications for the identity of p66$^{C/CT21}$ 239

6.3.5. Implications for p290^{SETD2} 240

6.3.6. The nature of the endogenous 12−12b−13»21-containing transcript variant and the role of exon 12b in *Setd2* expression 243

6.4. Conclusion 244

CHAPTER 7 FINAL DISCUSSION AND FUTURE DIRECTIONS

7.1. Identification of putative SETD2 isoforms 245

7.2. Verified alternative transcripts indicate potential protein diversity 246

7.3. Functional significance of SETD2 diversity 247

7.3.1. General comments on putative isoforms 247

7.3.2. The SETD2(13»21) variant: C-terminal interaction domains independent of KMT activity ? 248

7.3.3. Implications for p290^{SETD2} structure: the SRI domain and mechanism of H3K36me3 generation 249

7.3.4. p66$^{C/CT21}$ as a putative C-terminal SETD2 isoform 249

7.3.5. What regulates alternative splicing of *Setd2*? Is splicing altered during development or disease? 250

7.3.6. Significance for SETD2 function in chromatin modification, tumourigenesis and Huntington's disease 251

7.4. Conclusion 252

APPENDIX A SUPPLEMENTARY RESULTS FOR CHAPTER 3: *IMMUNODETECTION OF P290^{SETD2} AND PUTATIVE SETD2 ISOFORMS*

A-1 Estimation of apparent molecular masses of immunoreactive proteins ... 253

A-2 Characterisation of minimal conditions for p66^{C} detection 254

A-2.1. Thermo-urea denaturation causes loss of p120^{N} detection 254

A-2.2. Immunodetection of p66^{C} requires concomitant heating with urea 254

A-3 Distribution of p66^{C} after ultracentrifugation of TX–DOC extracts 258

A-4 Effect of emetine treatment on SETD2 immunoforms 261

APPENDIX B SUPPLEMENTARY RESULTS FOR CHAPTER 4: *GENERATION OF A P66-ENRICHED FRACTION*

B-1 p66^{C} and p66^{CT21} co-fractionate after extensive extraction of nuclease-resistant material 265

B-2 α-SETD2 CT21 antibody does not recognise p290^{SETD2} 266

APPENDIX C SUPPLEMENTARY RESULTS FOR CHAPTER 5: *IDENTIFICATION OF ALTERNATIVE* SETD2 *TRANSCRIPTS*

C-1 Exon–intron structure of alternatively spliced murine *Setd2* transcripts ... 271

C-1.1. Exon 12–Intron–Exon 12b 271

C-1.2. Exon 12b–Intron–Exon 13 271

C-1.3. Exon 12b–Intron–Exon 12c 272

C-2 *Setd2* RT-PCR sequencing data 273

C-2.1. Sequence of 12b–13 273

C-2.2. Sequence of 12b–13–14 273

C-2.3. Sequence of 12b–13–14–15 273

C-2.4. Sequence of 12–12b 274

C-2.5. Sequence of 11–12–12b 275
C-2.6. Sequence of 12–12b–12c 276
C-2.7. Sequence of 6–7–8–9–10.5 277
C-2.8. Sequence of 8–9.5 278
C-2.9. Sequence of 17–18.5 279

REFERENCES 280

List of figures

CHAPTER 1

Figure 1.1 Genome-wide distribution of histone modifications in mammals 3

Figure 1.2 Domain structure and interactions of SETD2 8

Figure 1.3 SETD2 protein sequence: domain organisation and antibody epitopes 10

Figure 1.4 H3K36me3 correlates with exon–intron organisation 18

Figure 1.5 Model for coupling SETD2-dependent H3K36me3 to transcription elongation by RNA polymerase II 27

CHAPTER 3

Figure 3.1 Generation of affinity-purified antibodies specific to an N-terminal region of SETD2 84

Figure 3.2 Transfer conditions influence detection of $p290^{N}$ and $p120^{N}$ proteins by affinity-purified α-SETD2 N-terminal antibody 88

Figure 3.3 Generation of affinity-purified antibodies specific to a C-terminal region of SETD2 91

Figure 3.4 Affinity-purified α-SETD2 C-terminal antibody detects $p66^{C}$ in TX–DOC extracts after thermal denaturation 95

Figure 3.5 Immunodetection of $p66^{C}$ is eliminated by pre-adsorption with C-terminal SETD2 peptide 98

Figure 3.6 Immunodetection of $p66^{C}$ is influenced by [urea], temperature and duration of heating 101

Figure 3.6 (continued) 103

Figure 3.7 Conservation of SETD2 immunoforms in multiple mammalian cell types 107

Figure 3.7 (continued) 109

Figure 3.8 Transient transfection of siRNAs targeting 5' and 3' regions of *Setd2* only depletes $p290^{N/C}$ 112

Figure 3.8 (continued) 114

Figure 3.9 $p290^{SETD2}$ is relatively unstable compared to $p120^{N}$, $p140^{C}$ and $p66^{C}$ 117

Figure 3.10 The decomposition of urea causes carbamylation of proteins 125

CHAPTER 4

Figure 4.1 Subcellular fractionation by hypotonic lysis, nuclease digestion and high salt extraction 134

Figure 4.2 Subcellular fractionation of 10T½ cells by sequential nuclease digestion and high-salt extraction reveals distinct fractionation of SETD2 immunoforms 136

Figure 4.3 Virtually all $p66^{C}$ that is extracted by high-salt buffer exists in a complexed form that sediments by ultracentrifugation 137

Figure 4.4 Characterisation of the detergent extraction and sedimentation properties of C-terminal SETD2 immunoforms 141

Figure 4.5 Subcellular fractionation by differential detergent extraction and centrifugation (DDC) 144

Figure 4.6 Generation of a $p66^{C}$-enriched fraction by differential detergent extraction and centrifugation 145

Figure 4.7 Fractionation of SETD2 immunoforms and other proteins in DDC-derived extracts 146

Figure 4.8 Two C-terminal-region SETD2 antibodies detect similar proteins in $p66^{C/CT21}$-enriched fractions after separation by AU-PAGE 151

Figure 4.9 Two-dimensional AU/SDS-PAGE confirms detection of $p66^{C}$ and $p66^{CT21}$ in the absence of thermo-urea denaturation 155

Figure 4.9 (continued) 157

Figure 4.10 Localisation of $p66^{C/CT21}$ on a Coomassie-stained 2D-AU/SDS-gel 161

Figure 4.11 Mass spectrometry–based identification of select proteins in the $p66^{C/CT21}$-enriched TD-S^{100} fraction resolved by two dimensional electrophoresis 164

CHAPTER 5

Figure 5.1 Structure of the murine *Setd2* gene 179

Figure 5.2 Multiple transcripts in poly(A)+ RNA are detected with probes against the 3' region of *Setd2* 182

Figure 5.3 Identification of a highly conserved element between exons 12 and 13 185

Figure 5.4 Exon 12b splices to upstream and downstream exons 189

Figure 5.5 Major splice products across exons 12 and 14 exclude exon 12b 192

Figure 5.6 Exon 12b splices to a novel downstream exon (12c) 195

Figure 5.7 Alternative 3' splice sites define two 5’ boundaries of exon 14 198

Figure 5.8 Identification of *Setd2* transcripts with retained introns 199

Figure 5.9 The 12–12b–13»21 cDNA contains a novel ORF encoding a ≈61 kDa protein containing C-terminal and CT21 epitopes 203

Figure 5.9 (continued) 205

Figure 5.10 Summary of alternatively spliced transcripts of murine *Setd2* in 10T½ fibroblasts 217

CHAPTER 6

Figure 6.1 Epitope-tagged SETD2(13»21) constructs 221

Figure 6.2 FLAG-tagged SETD2(13»21) co-fractionates with $p66^{C/CT21}$ after nuclease–salt extraction but migrates anomalously at ≈88 kDa 224

Figure 6.3 HA-tagged SETD2(13»21) co-fractionates with $p66^{C/CT21}$ after nuclease–salt extraction but migrates anomalously at ≈85 kDa 228

Figure 6.4 FLAG-tagged SETD2(13»21) does not recapitulate the DDC fractionation properties of $p66^{C/CT21}$ 231

Figure 6.5 HA-tagged SETD2(13»21) does not recapitulate the DDC fractionation properties of $p66^{C/CT21}$ 234

APPENDICES

Figure A.1 Loss of $p120^{N}$ when extracts are heated in urea 255

Figure A.2 $p66^{C}$ detection requires concomitant thermo-urea denaturation 256

Figure A.3 $p66^{C}$ is virtually absent from complexed material in TX–DOC extracts 260

Figure A.4 Effect of emetine on $p290^{SETD2}$ 263

Figure B.1 Co-fractionation of $p66^{C}$ and $p66^{CT21}$ in extracts derived from high-salt and detergent solubilisation of nuclease-resistant material 267

Figure B.2 α-SETD2 CT21 antibody does not recognise $p290^{SETD2}$ 268

List of tables

Table 1.1 – H3K36-methyl-binding proteins 30
Table 1.2 – SETD2 homolog/H3K36 methylation function in non-mammalian systems..... 31
Table 1.3 – Proteomic diversity of chromatin modifiers 48
Table 1.4 – Multiple forms reported for mammalian SETD2 proteins 49
Table 2.1 – α-SETD2 antibodies 52
Table 2.2 – Plasmids and clones 54
Table 2.3 – PCR primers for generation of epitope-tagged constructs 54
Table 2.4 – Small-interfering RNAs (siRNAs) 57
Table 2.5 – Antibodies used in immunoblotting 68
Table 2.6 – DNA probes for northern blotting 74
Table 2.7 – RT-PCR conditions 77
Table 2.8 – PCR primers for RT-PCR and DNA sequencing 77
Table 3.1 – α-SETD2 antibodies 82
Table 3.2 – Summary of mammalian α-SETD2 immunoforms detected 121
Table 4.1 – Proteins identified in TD-S^{100} fraction by mass spectrometry 166
Table 5.1 – Predicted proteins within 12b−13»21 cDNA 215
Table 7.1 – Estimation of molecular mass of α-SETD2 immunoforms 253

Chapter 1 Introduction

1.1. The complexity of eukaryotic gene regulation

Appropriate regulation of gene expression underlies all forms of unicellular and multi-cellular life. Its importance and complexity are demonstrated by dynamic, spatial and temporal changes in gene regulation that occur during the development of a multi-cellular organism from a single zygote (Hamatani *et al.*, 2004; Diez-Roux *et al.*, 2011), and the consequences of dysregulation in producing disease states (Villard, 2004). Eukaryotic gene expression is regulated at many levels, involving chromatin structure (Li *et al.*, 2007a); initiation, elongation and termination of transcription (Venters & Pugh, 2009); RNA processing, stability and subcellular localisation (Moore & Proudfoot, 2009); and finally, translation into proteins, which themselves may exist as multiple, functionally distinct, populations within a cell (let alone a whole organism) and be subject to diverse post-translational modifications (PTMs) (Orphanides & Reinberg, 2002; Mann & Jensen, 2003). Understanding the regulation and function of genes and their products within their natural context is critically important to fully comprehend molecular mechanisms that underlie the complexity of life, and differences between health and disease.

This thesis focuses on the expression of the mammalian *SETD2* gene, which encodes SET domain−containing protein 2 (SETD2). This introduction will describe our current understanding of its role in mammalian cell biology, and the rationale for its investigation.

1.1.1. Chromatin and modification of nucleosome structure

The nuclear eukaryotic genome is compacted and regulated by association of DNA with histones into nucleosome-based chromatin (Luger *et al.*, 1997), and modulated by ATP-dependent remodelers (Clapier & Cairns, 2009), histone chaperones (Avvakumov *et al.*, 2011), histone variant incorporation (Talbert & Henikoff, 2010), cytosine methylation (Klose & Bird, 2006), and numerous post-translational histone modifications (Kouzarides, 2007). Although general discussion of chromatin modification is beyond the scope of this thesis,

histone modification and trimethylation of lysine-36 of histone H3, in particular will be considered in depth due to their relevance to SETD2 function.

1.1.2. Histone modification

Histones are subject to a plethora of post-translational modifications, including: lysine acetylation; methylation of lysine and arginine; phosphorylation of serine, threonine and tyrosine; lysine ubiquitinylation; ADP-ribosylation of lysine and glutamic acid; lysine sumoylation; and proline isomerisation (Kouzarides, 2007). In most cases, histone modification is known to be dynamic and reversible by the enzymatic action of acetyltransferases and deacetylases, methyltransferases and demethylases, and kinases and phosphatases (Kouzarides, 2007). Chromatin immunoprecipitation (ChIP) coupled with microarrays (ChIP-chip) or high-throughput DNA sequencing (ChIP-seq) has revolutionised our understanding of the genomic locations of histone modifications (Wu *et al.*, 2006; Barski & Zhao, 2009). A schematic of the distribution of selective histone modifications across a mammalian chromomosome is depicted in **Figure 1.1A**. The quantification and interpretation of the physical manifestation of ChIP signals, however, remain highly contentious areas which can compromise our understanding of chromatin modification (Clayton *et al.*, 2006; Lee & Mahadevan, 2009).

In general, histone modifications may alter chromatin structure and function in several ways. Firstly, they can directly alter the biophysical properties of an individual nucleosome. For example, H3K79me2 modifies a hydrophobic patch on the nucleosome surface (Lu *et al.*, 2008) whilst acetylation alters nucleosome shape (Bauer *et al.*, 1994). Secondly, post-translational modifications may disrupt inter-nucleosome interactions and thus higher-order chromatin structure. For example, H4K16ac prevents interaction between the positively charged H4 tail and an acidic patch on H2A, thus promoting chromatin decompaction (Shogren-Knaak *et al.*, 2006). Thirdly, histones modifications may generate a binding site for the recruitment of effector proteins (Taverna *et al.*, 2007; Yun *et al.*, 2011). A 'histone code' is supported by examples of combinatorial readout of histone

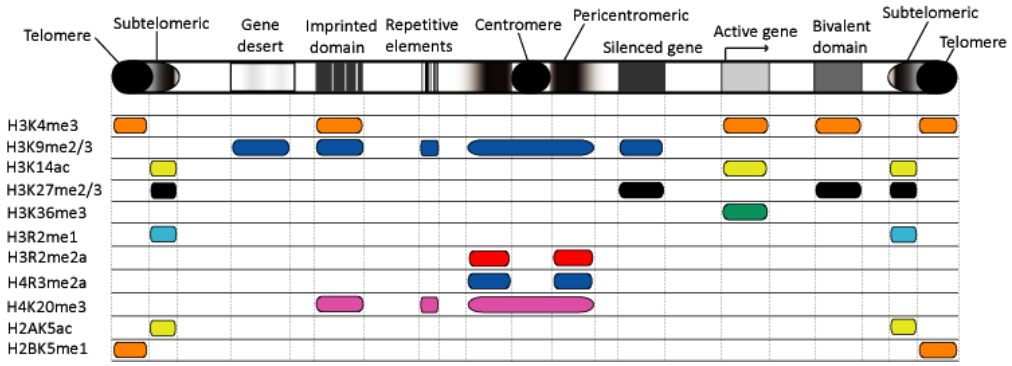
Figure 1.1 Genome-wide distribution of histone modifications in mammals

(A) Specific histone modifications are enriched over distinct regions of a schematic mammalian chromosome (not to scale).

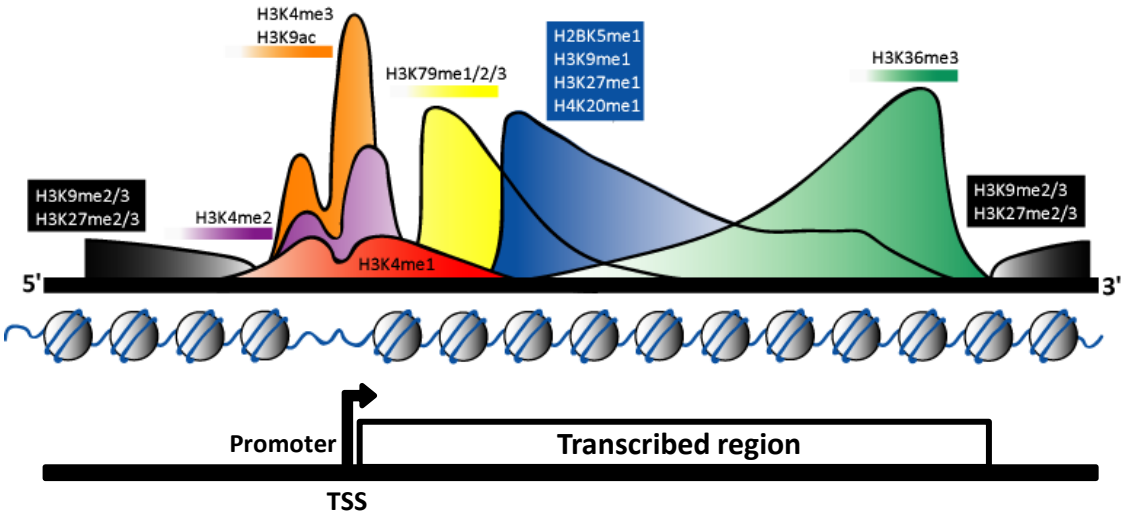
(B) Generalised profile of histone modifications associated with actively transcribed genes. TSS (transcriptional start site).

Adapted from (Lee & Mahadevan, 2009).

(A)


Telomere
Subtelomeric
Gene desert
Imprinted domain
Repetitive elements
Centromere
Pericentromeric
Silenced gene
Active gene
Bivalent domain
Subtelomeric
Telomere
H3K4me3
H3K9me2/3
H3K14ac
H3K27me2/3
H3K36me3
H3R2me1
H3R2me2a
H4R3me2a
H4K20me3
H2AK5ac
H2BK5me1

(B)


H3K4me3
H3K9ac
H3K79me1/2/3
H2BK5me1
H3K9me1
H3K27me1
H4K20me1
H3K36me3
H3K9me2/3
H3K27me2/3
H3K4me2
H3K4me1
H3K9me2/3
H3K27me2/3
5'
3'
Promoter
Transcribed region
TSS

modifications but remains contentious (Strahl & Allis, 2000; Lee *et al.*, 2010; Gardner *et al.*, 2011). Overall, the molecular consequences of histone modification are complex and context-dependent.

Histone acetylation is traditionally associated with the expression of poised and actively transcribed genes (Clayton *et al.*, 2006). This was originally supported by evidence from chromatin immunoprecipitation experiments (Hebbes *et al.*, 1988), and the correlation of acetylation with transcription-associated nucleosomes retained on organomercurial columns (Allegra *et al.*, 1987; Chen & Allfrey, 1987). Numerous ChIP–chip/seq experiments have elaborated this correlation on a genome-wide scale: acetylation peaks with H3K4me3 at promoters and transcriptional start sites (TSSs), and can also be found at enhancer elements (Roh *et al.*, 2005; Guenther *et al.*, 2007; Koch *et al.*, 2007; Wang *et al.*, 2008b; Heintzman *et al.*, 2009; Xie *et al.*, 2009). However, exceptions to this correlation exist; for example, H3K4ac has recently been reported to be involved in heterochromatin formation in *Schizosaccharomyces pombe* (Xhemalce & Kouzarides, 2010), while H4K12 acetylation is associated with heterochromatin formation in *Drosophila* (Turner *et al.*, 1992). Acetylation not only neutralises the positive charge of lysine, which may disrupt electrostatic interactions between histones and DNA (Hansen *et al.*, 1998a) that occur within and between nucleosomes (Luger *et al.*, 1997; Shogren-Knaak *et al.*, 2006), but may also recruit bromodomain-containing proteins, often found in chromatin remodelers and transcription factors (Sanchez & Zhou, 2009).

Trimethylation of H3K4 is almost ubiquitously associated with promoters and TSSs of poised and active genes throughout eukaryotes (Santos-Rosa *et al.*, 2002; Bell *et al.*, 2007; Mikkelsen *et al.*, 2007; Kolasinska-Zwierz *et al.*, 2009; Roudier *et al.*, 2011). H3K4me3 primarily functions through recruitment of proteins containing plant homeodomain (PHD) motifs or chromodomains, such as TAF3 (a component of TFIID) and the chromatin remodelling enzyme CHD1, which have roles in transcription initiation and RNA splicing, respectively (Sims *et al.*, 2007; Vermeulen *et al.*, 2007).

H3K9 trimethylation, on the other hand, is associated with constitutive heterochromatin, gene repression and inactive imprinted genes (Richards & Elgin, 2002; Schotta *et al.*, 2003; Lee & Mahadevan, 2009). Recognition of H3K9me3 by heterochromatin protein 1 (HP1) generates a repressive chromatin structure, hypoacetylated and self-sustaining through HP1-mediated recruitment of K9-methyltransferases and deacetylases (Lachner *et al.*, 2001; Nakayama *et al.*, 2001). However, H3K9me3 and HP1γ have also been reported to be associated with actively transcribed genes (Vakoc *et al.*, 2005; Blahnik *et al.*, 2011) and contribute towards regulation of alternative splicing (Saint-Andre *et al.*, 2011).

1.2. Overview of SET domain–containing protein 2 (SETD2)

1.2.1. Identification of SETD2

A truncated human protein encoded by the 3′ region of the *SETD2* gene was first identified by a yeast two-hybrid assay as an interactor with Huntingtin, and named Huntingtin yeast partner B (HYPB) (Faber *et al.*, 1998). The presence of a SET domain, revealed after full-length complementary DNA (cDNA) assembly, along with the ability to bind DNA provided the first evidence for a putative role of SETD2 in transcriptional regulation (Rega *et al.*, 2001). SETD2 was later shown to possess methyltransferase activity towards lysine-36 of histone H3 *in vitro* (Sun *et al.*, 2005), and required specifically for global trimethylation of this residue *in vivo* (Edmunds *et al.*, 2008; Yoh *et al.*, 2008; Hu *et al.*, 2010). From various identification methods and nomenclature, SETD2/HYPB is also known as: HIF-1 (Huntingtin-interacting factor 1), $p231^{HBP}$ (Huntingtin-binding protein), KMT3A (lysine methyltransferase 3A), KIAA1732 and HSPC069.

SETD2 homologs appear widespread throughout eukaryotes and have been characterised in zebrafish (Sun *et al.*, 2008), *Drosophila* (Bell et al., 2007; Stabell et al., 2007), *S. pombe* (Morris *et al.*, 2005), *Neurospora crassa* (Adhvaryu *et al.*, 2005) and *Saccharomyces cerevisiae* (Strahl *et al.*, 2002; Landry *et al.*, 2003; Xiao *et al.*, 2003; Kizer *et al.*, 2005; Du *et al.*, 2008; Youdell *et al.*, 2008; Psathas *et al.*, 2009; Lin *et al.*, 2010). The budding yeast homolog, Set2, is most thoroughly characterised in parallel to mammalian SETD2; however the function of yeast Set2 and the mechanism of generating H3K36 methylation derivatives should be

cautiously extrapolated to mammals. Yeast Set2 is the sole lysine methyltransferase (KMT) responsible for generating all three H3K36 methylation states *in vivo* (Strahl *et al.*, 2002) whereas the mammalian SET2-family includes SETD2, NSD1, NSD2, NSD3 and ASH1L (Volkel & Angrand, 2007). Although SETD2 most closely resembles the domain organisation of Set2, and certain interactions appear to be conserved (discussed below), it is unclear to what extent all of the characteristics of the yeast homolog are retained in SETD2 and what features of H3K36 generation by Set2 are conferred by other KMTs in mammals. For example, in *Drosophila*, zebrafish and mammals, the NSD-family methyltransferases appear to be predominantly responsible for H3K36me2 rather than H3K36me3 (Rayasam *et al.*, 2003; Bell *et al.*, 2007; Li *et al.*, 2009b; Yamada-Okabe *et al.*, 2010). Other methyltransferases, such as SETMAR and SMYD-family enzymes have also been reported with K36-methylation activity *in vitro* but their specificities *in vivo* are unclear (Lee *et al.*, 2005; Brown *et al.*, 2006).

1.2.2. Overview of SETD2 protein structure and interactions

The most up-to-date information on murine SETD2 on current databases predicts a ≈286 kDa protein of 2537 residues, which is similarly conserved for human SETD2 (≈287 kDa; 2564 residues). SETD2 is a large, multi-domain protein that interacts with various factors as indicated in **Figure 1.2**. The C-terminal region of ≈1200 residues is characterised by several functionally characterised domains, which will be discussed below. The function of the N-terminal region (≈1400 residues), however, is unclear: predicted to be structurally disordered (Yang *et al.*, 2005), it contains no recognisable domains but comprises regions of compositional bias, of unclear function. The polypeptide sequence of SETD2 is provided in **Figure 1.3**, highlighting in more detail domains and antibody epitopes investigated in this thesis.

It should be noted that understanding of mammalian SETD2 comes from both murine and human sources. Certain interactions have only been characterised by ectopic expression of defined truncated constructs; others by *in vitro* techniques, and some confirmed by *in vivo* techniques. However, the number and nature of endogenous proteins that are derived from the *SETD2* gene is unclear (discussed in section 1.6.2). Therefore, **Figure 1.2B**

Figure 1.2 Domain structure and interactions of SETD2

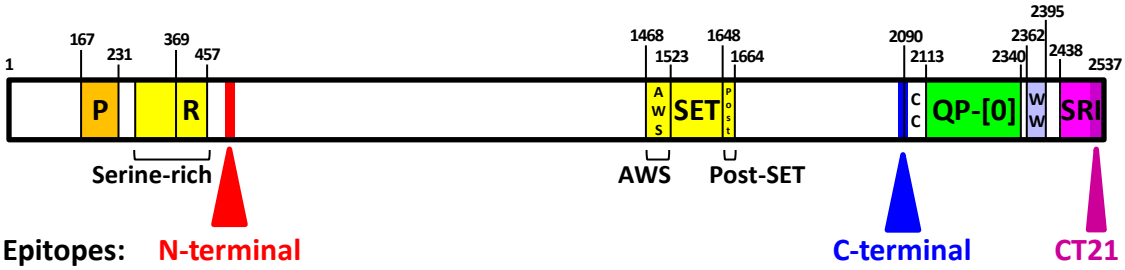
(A) SETD2 domain organisation: Schematic representation (to scale) of murine SETD2 (residue numbers correspond to UniProt KB entry E9Q5F9). Abbreviations: AWS (associates with SET); CC (coiled-coil, predicted); P (proline-rich region); Post (Post-SET); QP-[0] (glutamine and proline–rich, uncharged region); R (arginine-rich region); SET (Suppressor of variegation 3–9, Enhancer of *Zeste*, *Trithorax*); SRI (Set2–Rpb1 interaction); WW (WW domain).

(B) SETD2 interactions: Indicated proteins and DNA bind to C-terminal regions of SETD2. Different coloured segments designate multiple sites of interaction. Abbreviations: hnRNP-L (heterogeneous nuclear ribonucleoprotein L); IWS1 (interacts with Spt6 protein 1).

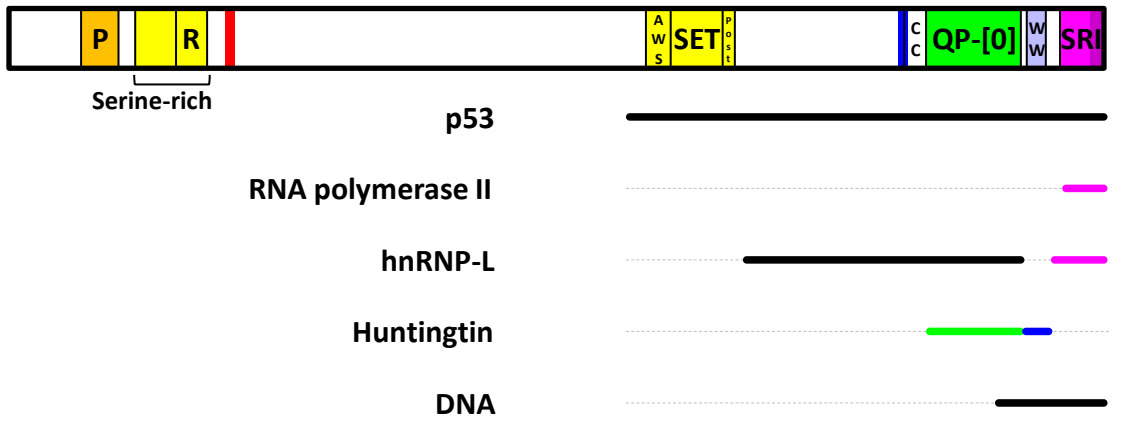
(A)

Murine SETD2

286 kDa
2537 amino acids



(B) SETD2 interactors



IWS1

CIC (Capicua homolog 1)

Ataxin-1

Retinoic acid receptor α complex

Interaction properties unclear

Figure 1.3 SETD2 protein sequence: domain organisation and antibody epitopes

Predicted amino acid sequence of murine SETD2 protein of ≈286 kDa (UniProt KB entry E9Q5F9). Antibody epitopes are highlighted, along with the domains. Domains were identified using Prosite, Pfam/SMART, and SAPS tools.

```
   1 MKPLPSQQPPPKMGDFYDPEHPTPEEEENEAKIENVQKTGFIKGPVFKGV
  51 ASSRFLPKGTKTKVNLEEQGRQKVSFSFSFTKKTLQNRFLTALSNEKQSD
 101 SPNSPAPPLQVDSNPKVKMDAGDTFPATEESSPPKSRVELGRIHFKKHLL
 151 HVTSRPQLAASTTAASPLPPTTQLPAVLAESMIDSPPSSPPPPPPPPQAS
 201 SPSPPAQISEPVALPQPPATALMTSPPGPLPGDVAVRAQKESPVKSGPEV
 251 LEVDTKQDIVSNSLEEHTVQTLKEQADHLLQKEDSHIGKEEEVSDGSKIS
 301 LSSKKASSKKKSSQFEGTFLGSESDEDSVRTSSSQRSHDLKSSTSIDKER
 351 DFKKSSAPSKSEDLGKSSRSKTERDDRYCSYSKLERDTRYVSSRCRSERD
 401 RRRSRSRSRSDRASRTSLSYSRSERSHYYDSERRYHRSSPYRERTRYSRP
 451 YTDNRARESSDSEDEYKKTYPRRTSAHSYRDLRTSSSYSKFDRDCKTETS
 501 YLEMERRGKYTSKLERESKRTSEHETIKRCCSPPNELGFRRGSSYSKHDN
 551 STSRYKSALSKSISKNDKFKNSFCCTELNEENKQSHSFSLQTPCSKGSEL
 601 RTINKISEREKTGSPTPSNQLNDSPTFKKLDESPVLKPEFIGHDGRESIK
 651 ELELSKVKNDQLRNFCSIELNVNGSPETEADVATFCTSKTDAISMTSDDS
 701 VTGSEVSPLIKACMLSSNGFQNVGRCRERDSDDTCRQHNTSKSPFREMEP
 751 LLSPHHDKLMSLPVKTIDYPKTLIKEPVDKRHSCCKTKDSDIYCSPNENP
 801 EAENAEPSAMTISSHSFVNVHLESKTVICDNREPTDRHSENTCDEYKQSI
 851 GSTSSASHNHFDGLYEPIGSSGISSLQSPPSGIRCEENTSPTLDAVESKK
 901 GIDFLKYARKETDVGSALPDSGKGFSWENRHNNVLSGQSLQEAQEEGNSI
 951 LHERRGRPEIPLDEEQRGHTHISDDSEVVFPYDLNLTMEDSDGITYTLKC
1001 DSSGNAPEIVSTVHEDYSGSSASSSDESDSEDTESDDSSIPRNRLQSVVV
1051 VPKNSTLPMEETSPCSSRSSQSYKHYSDRWEDGLETRRHAYEEEYESKGC
1101 SQTEKYFLHKGTERSAESCYSQFGRKADNHLPDIAHAQSDGVDSTSQTDS
1151 RSDHLGHLNPEDTLRAKTSRPQELPVYSDDFEDLPNKSRQQMIFSNRPDS
1201 SRLGKTELSFSSSCDISRMDGLHSSEELRNLGWDFSQQERPTTTYQQPDS
1251 SYGTCGTHKYQQSTEHYGGTHNYWQGNGYWDPRSAGRPPGTGLAYDRIQG
1301 QVPDSLTDDREEEEHWDQRSGSHFSSPSNKFFFHQKDKGSVQAPEISSNS
1351 IKDALVMNERKDFSKNFEKNDIKERGPPKKRRQELESDSESDGELQARKK
1401 VRVEMEQGESSVPQHSELMGPSCAMDDFRDPQRWKEFAKLGKMPCYFDLI
1451 EENVYLTERKKNKSHRDIKRMQCECTPLSKDERAQGEVACGEDCLNRLLM
1501 IECSSRCPNGDYCSNRRFQRKQHADVEVILTEKKGWGLRAAKDLPSNTFV
1511 LEYCGEVLDHKEFKARVKEYARNKNIHYYFMALKNDEIIDATQKGNCSRF
1601 MNHSCEPNCETQKWTVNGQLRVGFFTTKLVPSGSELTFDYQFQRYGKEAQ
1651 KCFCGSANCRGYLGGENRVSIRAAGGKMKKERSRKKDSVDGELEALMENG
1701 EGLSDKNQVLSLSRLMVRIETLEQKLTCLKLIQNTHSQSCLKSFLERHGL
1751 SLLWIWMAELGDGRESNQKLQEEIIKTLEHLPIPTKNMLEESKVLPIIQR
1801 WSQTKTAVPQLSEGDGYSSENTSRAHTPLNTPDPSAKPSTEMDTDTPKKL
1851 IFRRLKIISENSMDSAVSDVTSELECKDGKEDLDQLETVTVEEDEELQSQ
1901 QLLPQQLCESKVESEATIEVSKLPTSEPEADTETEPKDSNGTKLEETIAE
1951 ETPSQDEEEGVSDVESERSQEPPDKTVDISDLATKLLDSWKDLKEVYRIP
2001 KKSQTEKESTVAERGRDAAAFRDQTAPKTPNRSRERDPDKQSQNKEKRKR
2051 RGSLSPPSSAYERGTKRPDDRYDTPTSKKKVRIKDRNKLSTEERRKLFEQ
2101 EVAQREAQKQQQQMQNLGMTSPLPFDSLGYNASHHPFAGYPPGYPMQAYV
2151 DPSNPNAGKVLLPTPSMDPVCSPAPYDHAQPLVGHSTESLAAPPSVPVVP
2201 HVAASVEVSSSQYVAQNESVVHQDSNVPVMPVQAPGPVQGQNYNVWESNQ
2251 QSVSVQQQYSPAQSQTTIYYQGQTCSTVYSVTSPYSQTTPPIVQSYAQPS
2301 LQYIQGQQIFTAHPQGVVVQPTAAVTSIVAPGQPQSLQPPEMVVTNNLLD
2351 LPPPSPPKPKTIVLPPNWKTARDPEGKIYYYHVITRQTQWDPPTWESPGD
2401 DASLEHEAEMDLGTPTYDENPMKTSKKPKTAEADTSSELAKKSKEVFRKE
2451 MSQFIVQCLNPYRKPDCKVGRITTTEDFKHLARKLTHGVMNKELKYCKNP
2501 EDLECNENVKHKTKEYIKKYMQKFGAVYKPKEDTELE
```

Polyproline
Proline-rich

Serine-rich

RS-like

N-term Ab

AWS domain
SET domain

Post-SET domain

C-term Ab
Coiled-coil
Uncharged
Proline-rich
Gln-rich

WW domain
SRI domain

CT21 Ab

provides a representative overview of all reported interactions, but does not convey the confidence by which these interactions can be assumed to occur *in vivo* for a single protein.

Various genetic and biochemical approaches have characterised mammalian SETD2; these indicate that SETD2 functions in numerous processes, the most prominent being trimethylation of lysine-36 of histone H3. Its KMT function is arguably best-characterised, especially from studies of the Set2 homolog in *S. cerevisiae*. The role of SETD2 in generation of this modification and its proposed functions will be discussed first, given greater understanding of this aspect of SETD2. Following this, I will discuss additional roles of SETD2 that appear distinct from K36me3 generation. It is important to note, however, that often there is no clear evidence why a SETD2-related function may be dependent—or equally, independent—of KMT activity. Whilst genetic studies in yeast can employ histone modification substitutions to probe the effect of H3K36 methylation, functional studies of mammalian SETD2 have relied on depletion of the entire protein or overexpression of truncated constructs. Thus without specific KMT inhibition, it is currently unclear whether functions are dependent on KMT activity, H3K36me3 generation specifically, or the action of other domains within SETD2.

1.3. SETD2 as lysine methyltransferase

The major focus of this thesis is the endogenous diversity of forms of mammalian SETD2. As an H3K36-methyltransferase, mechanistic information gained from studying yeast Set2 will be considered relevant. However, because the downstream effects of H3K36me3 are not the major focus of this thesis, they will only be considered in detail for the mammalian system. Indeed, many functions for H3K36me3 have been described for other organisms, which may or may not be relevant to SETD2 function in mammalian cells. These are briefly summarised later in **Table 1.2** for completeness.

1.3.1. SETD2 mediates trimethylation of histone H3K36

1.3.1.1. *SETD2 as lysine methyltransferase in vitro*

Lysine methyltransferases catalyse the addition of up to three methyl groups, each derived from *S*-adenosylmethionine, to the ε-amino group of lysine, producing mono/di/tri-methyl-lysine and *S*-homocysteine (Dillon *et al.*, 2005). Except for the H3K79-methyltransferase DOT1 (Ng *et al.*, 2002), KMTs derive methylation activity from the SET (Suppressor of variegation(3–9)—Enhancer of *Zeste*—*Trithorax*) domain, identified originally as a component shared between transcriptional regulators with KMT activity in *Drosophila* (Dillon *et al.*, 2005). Bioinformatic and structural analysis demonstrates that SET domains often function in combination with various flanking domains that contribute towards substrate binding and regulation of KMT activity (Qian & Zhou, 2006). Additionally, SET domains have also been described as interaction domains; for example, binding proteins with similarity to dual-specificity phosphatases (Cui *et al.*, 1998; De Vivo *et al.*, 1998), or binding single-stranded DNA, which may disrupt nucleosome patterning (Krajewski *et al.*, 2005).

SETD2 contains a tripartite SET domain organisation (**Figure 1.2A**): a canonical SET domain, flanked by a cysteine-rich AWS (associates with SET) domain, and a Post-SET domain (Sun *et al.*, 2005). Initial predictions of KMT activity (Rega *et al.*, 2001) were confirmed by demonstration that the isolated tripartite region catalyses methylation of lysine-36 of histone H3 *in vitro* (Sun *et al.*, 2005). SETD2 exhibits preference for methylation of H3K36 within nucleosomes over histone octamers and isolated histones (Sun *et al.*, 2005; Li *et al.*, 2009b), similar to homologs in unicellular eukaryotes (Strahl *et al.*, 2002; Adhvaryu *et al.*, 2005; Morris *et al.*, 2005; Du *et al.*, 2008). SETD2 can catalyse all three H3K36 methylation states *in vitro* (Li *et al.*, 2009b; Yuan *et al.*, 2009) and can utilise nucleosomal substrates containing unmethylated H3K36 or H3K36me1/2 (Jia *et al.*, 2009; Li *et al.*, 2009b), suggesting that SETD2 activity does not require a pre-methylated substrate.

Nucleosomal substrate preference may be rationalised from genetic studies in yeast and structural studies of SETD2 paralogs in humans. Yeast Set2 requires a nucleosomal surface

comprising H4K44 and several H3 and H2A residues within its vicinity: residue substitution prevents Set2 from interaction with chromatin *in vivo*, and leads to reductions in global H3K36me2 and H3K36me3 (although H3K36me1 is unaffected) (Du *et al.*, 2008; Du & Briggs, 2010). In humans, structures of the catalytic domains of SETD2 and K36-related KMT paralogs, NSD1 and ASH1L, reveal that the Post-SET domain contributes a critical regulatory loop, which appears to inhibit substrate access to the active site, in the absence of nucleosome interactions (An *et al.*, 2011; Qiao *et al.*, 2011).

Curiously, the isolated SET domain from SETD2 also exhibited auto-methylation activity *in vitro* independent of the presence of nucleosomes or histone substrates (Sun *et al.*, 2005). Although of unclear significance for SETD2, such auto-modification is not uncommon; for example, auto-acetylation of the histone acetyltranfserases P/CAF and p300/CBP increases their acetyltransferase activity (Santos-Rosa *et al.*, 2003; Thompson *et al.*, 2004); similarly, kinase regulation by auto-phosphorylation is widespread in signal transduction (Smith *et al.*, 1993). The observation of auto-methylation opens up the possibility that SETD2 may methylate non-histone substrates *in vivo*, similar to many KMTs that were originally identified as histone methyltransferases; for example, SMYD2 methylates p53 and Rb (Saddic *et al.*, 2010); NSD1 acts on the p65 subunit of NFκB (Lu *et al.*, 2010); and SET7/9 methylates diverse non-histone substrates, including STAT3, TAF10, E2F1 and Rb (Kouskouti *et al.*, 2004; Kontaki & Talianidis, 2010; Munro *et al.*, 2010; Yang *et al.*, 2010).

1.3.1.2. *SETD2 mediates H3K36me3 in vivo*

SETD2 acts as a trimethyltransferase towards H3K36 *in vitro*, but how does it function *in vivo*? Knockdown by RNA interference (RNAi) in murine and human cells reveals virtually complete elimination of global H3K36me3, specifically—with no discernible effects on global H3K36me1/2 (Edmunds *et al.*, 2008; Yoh *et al.*, 2008). Homologous disruption of *Setd2* in mice is embryonic lethal, but *Setd2*$^{-/-}$ embryos at E10.5 exhibit almost complete loss of H3K36me3 (Hu *et al.*, 2010), consistent with the aforementioned RNAi studies.

Although these interpretations appear conclusive, two points are worth considering: firstly, SETD2 may produce these results indirectly, if it functions upstream of, or in parallel with,

other methyltransferases. Secondly, although SETD2 does not influence global H3K36me1/2, a contribution towards a minor proportion of H3K36me1/2 *in vivo* cannot be formally excluded. However, the majority of H3K36me2 is likely catalysed by SETD2-paralogs, such as NSD2, which contribute substantially towards H3K36me2 *in vivo* (Li *et al.*, 2009b).

The involvement of SETD2 in the production of H3K36me3 immediately poses mechanistic questions: How does SETD2 generate H3K36me3? What is the function of H3K36me3? Before addressing these questions in sections 1.3.3 and 0, the genomic location of H3K36me3 will now be discussed.

1.3.2. Genome-wide localisation of H3K36 methylation

1.3.2.1. *H3K36me3 marks chromatin transcribed by RNA polymerase II*

The distribution of H3K36me3 (**Figure 1.1A-B**) has been investigated using ChIP coupled with RT-PCR to probe individual loci (Bannister *et al.*, 2005; Vakoc *et al.*, 2006; Edmunds *et al.*, 2008), or by ChIP-chip and ChIP-seq to reveal its genome-wide localisation (Barski *et al.*, 2007; Guenther *et al.*, 2007; Mikkelsen *et al.*, 2007; Hon *et al.*, 2009). One of the earliest analyses of K36me3 distribution demonstrated its association across transcribed chromatin, with enrichment at 3′ regions, in active chicken genes (Bannister *et al.*, 2005). Similar distributions of H3K36me3 were observed across the transcribed regions of highly expressed mammalian genes (Vakoc *et al.*, 2006; Kim *et al.*, 2007; Edmunds *et al.*, 2008). ChIP-chip and ChIP-seq have consistently corroborated the 3′-skewed distribution across transcribed regions of active RNA polymerase II (RNAPII)–dependent genes, on a genome-wide scale in a variety of eukaryotes: mammals (Barski *et al.*, 2007; Guenther *et al.*, 2007; Mikkelsen *et al.*, 2007; Wang *et al.*, 2008b; Schwartz *et al.*, 2009; Huff *et al.*, 2010), *Drosophila* (Bell *et al.*, 2007; Roy *et al.*, 2010), *Caenorhabditis elegans* (Kolasinska-Zwierz *et al.*, 2009; Gerstein *et al.*, 2010) and yeast (Pokholok *et al.*, 2005). The association of H3K36me3 with active genes and peaks of H3K4me3 at promoters has revealed H3K4me3–K36me3 domains as predictors of novel genes, which has been especially informative for discovery

of large intervening non-coding RNAs and microRNAs (Barski *et al.*, 2009; Guttman *et al.*, 2009; Zhu *et al.*, 2011).

Although the majority of H3K36me3 appears to associate with RNAPII-transcribed genes, H3K36me3 antibodies strongly stain nucleoli, which do not colocalise with SETD2 (Muller *et al.*, 2007). Furthermore, recent reports have described H3K36me3 associated with constitutive heterochromatin, which also appears independent of *SETD2* knockdown (Chantalat *et al.*, 2011).

Genes transcribed by RNA polymerases I and III (RNAPI and RNAPIII, respectively) appear to lack H3K36me3 (Barski *et al.*, 2010; Oler *et al.*, 2010; Zentner *et al.*, 2011), consistent with studies in yeast (Rao *et al.*, 2005). Although this may be due to lack of recruitment of SETD2 (since RNAPI and RNAPIII lack the carboxy-terminal domain of RNAPII, which interacts with SETD2—discussed in section 1.3.3), the recent observation that RNAPII can be localised at RNAPIII-transcribed regions, along with positive transcription elongation factor B (pTEFb), and RNAPII phosphorylated on serine-2 (phosphoserine-2) and on serine-5 (phosphoserine-5) poses challenges for models of SETD2 recruitment discussed below (Barski *et al.*, 2010; Oler *et al.*, 2010; Raha *et al.*, 2010). Given RNAPIII-transcribed genes are rarely spliced, this may contribute towards decreased SETD2 recruitment or activity, as discussed in section 1.3.3.5 (de Almeida *et al.*, 2011; Kim *et al.*, 2011).

1.3.2.2. *H3K36me3 differentiates exonic versus intronic chromatin*

High resolution ChIP-seq studies have revealed differential marking of exons and introns with a variety of histone modifications: H3K36me3 being the most prominent mark enriched on exons (Andersson *et al.*, 2009; Kolasinska-Zwierz *et al.*, 2009; Schwartz *et al.*, 2009; Spies *et al.*, 2009; Dhami *et al.*, 2010; Huff *et al.*, 2010). The first indication of this differential distribution came from observation of its genome-wide enrichment on exons compared to introns in *C. elegans*, mice and humans (Kolasinska-Zwierz et al., 2009). Independent studies corroborated these observations in general, but disagree on several issues: firstly, whether H3K36me3 skews towards the 5′ of exons (Andersson et al., 2009;

Hon et al., 2009; Schwartz et al., 2009) or the 3′ side (Kolasinska-Zwierz et al., 2009; Nahkuri et al., 2009); secondly, whether nucleosome density predominantly accounts for exonic H3K36me3 enrichment (Nahkuri et al., 2009; Schwartz et al., 2009; Tilgner et al., 2009) or insufficiently explains it (Andersson *et al.*, 2009; Kolasinska-Zwierz *et al.*, 2009; Spies *et al.*, 2009; Dhami *et al.*, 2010; Huff *et al.*, 2010); and finally, differences in co-enrichment with other modifications (Andersson *et al.*, 2009; Nahkuri *et al.*, 2009; Schwartz *et al.*, 2009; Tilgner *et al.*, 2009; Dhami *et al.*, 2010; Huff *et al.*, 2010).

Genes with large first introns highlight that H3K36me3 is largely excluded from the first intron, which is enriched with H3K79me2 (Huff *et al.*, 2010). As schematically represented in **Figure 1.4**, H3K36me3 then progressively increases over internal exons (Andersson *et al.*, 2009; Hon *et al.*, 2009; Schwartz *et al.*, 2009; Huff *et al.*, 2010; de Almeida *et al.*, 2011) until decreasing sharply near the polyadenylation site (Barski et al., 2007). The significance of these observations will be discussed further in relation to RNA splicing later (sections 1.3.3.5 and 1.4.3).

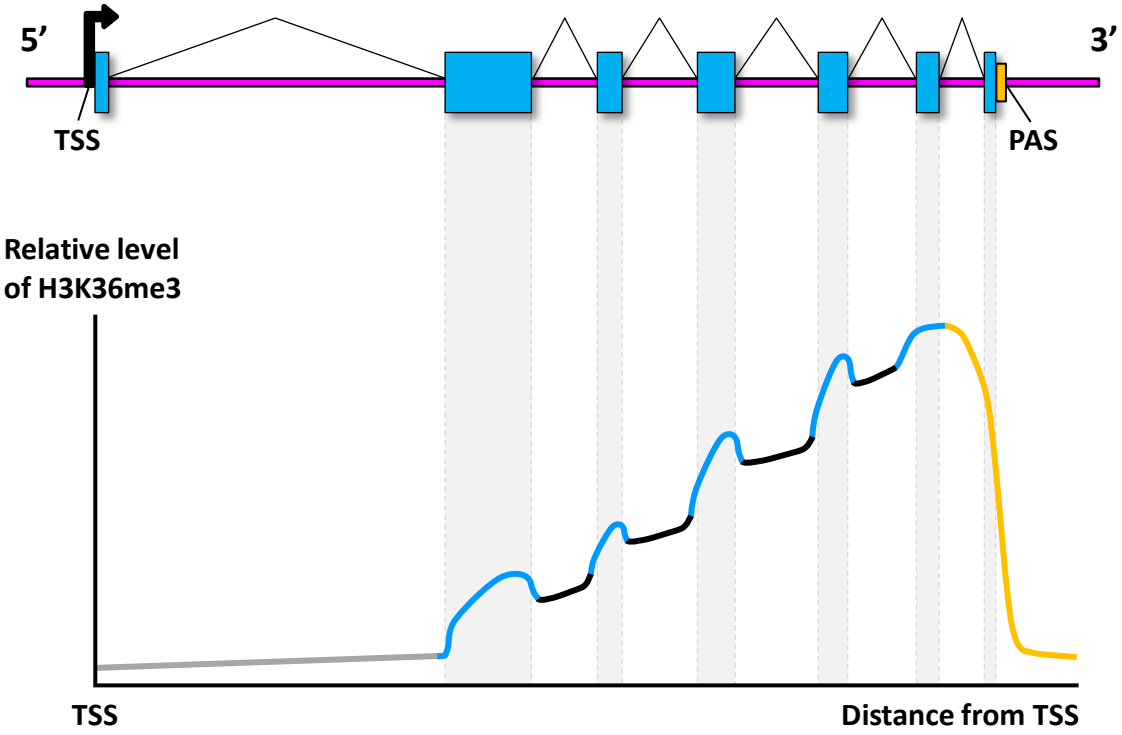
1.3.2.3. *SETD2-mediated H3K36me3 is associated with RNAPII elongation*

Although genome-wide studies clearly demonstrate an association of H3K36me3 with transcribed genes and a correlation between H3K36me3 levels and expression level, they provide limited temporal information on H3K36me3 dynamics.

The mammalian immediate-early genes, c-*fos* and c-*jun*, which are rapidly and transiently induced by mitogens (Hazzalin & Mahadevan, 2002), provide a powerful model for analysis of H3K36me3 dynamics during gene induction: from a poised state to transcriptional activation, and finally shut off back to an inactive state (Hazzalin & Mahadevan, 2002; Edmunds *et al.*, 2008). In the poised state, low basal levels of H3K36me3 occur across transcribed regions. During transcriptional induction, H3K36me3 increases throughout transcribed regions, peaking at the 3′ end, and is maintained for at least two hours after induction (Edmunds *et al.*, 2008). Although both basal and induced H3K36me3 are dependent on SETD2, only the latter is abrogated by inhibition by transcription elongation using the CDK9 inhibitor, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole

Figure 1.4 H3K36me3 correlates with exon–intron organisation

Schematic representation of how H3K36me3 levels vary with gene structure. H3K36me3 remains relatively low at the transcriptional start site (TSS) and over the first intron (grey) then progressively increases over the transcribed region with greater increases in internal exons (blue) compared to introns (black). H3K36me3 levels over final exons are less clear (orange), but often decrease dramatically near the polyadenylation site (PAS) by an unclear mechanism.


5’
3’
TSS
PAS
Relative level
of H3K36me3
TSS
Distance from TSS

(DRB) (Edmunds *et al.*, 2008), suggesting that multiple mechanisms likely exist for SETD2-mediated H3K36me3 deposition across mammalian genes.

Induction of H3K36me3 during gene activation has also been observed for interferon-inducible genes in mammals (Tamura *et al.*, 2009; Buro *et al.*, 2010). Whilst in yeast, increases in H3K36me3 biased towards the 3′ end of transcribed regions are similarly observed during rapid induction of *MET16* (Morillon *et al.*, 2005).

However, there are some anomalous results that conflict with the general correlation between H3K36me3 and transcription elongation. Firstly, the U2 snRNA gene and several large (>100 kb), tumour necrosis factor-α–inducible genes do not acquire H3K36me3 upon activation (Egloff *et al.*, 2009; Wada *et al.*, 2009). Secondly, genome-wide studies in *Drosophila* have also identified two classes of active chromatin, differentiated by H3K36me3 association (Filion *et al.*, 2010). Thirdly, during terminal erythroid differentiation, erythoid-specific genes poised for expression appear pre-marked with higher levels of H3K36me3 compared to repressed genes in progenitor cells, but no increase is observed when the former are induced during differentiation (Wong *et al.*, 2011). Fourthly, the *MET16* induction study in yeast reported transient peaks of H3K36me2/3 at the promoter, during the pre-elongation stage of transcription (Morillon *et al.*, 2005). Finally, H3K36me3 has also been found enriched at promoters of inactive developmental genes in zebrafish sperm (Wu *et al.*, 2011b). The physiological significance of this difference is currently unclear, though repressive functions of SET2-homologs and H3K36me3 have been reported outside of transcribed regions (**Table 1.2**).

1.3.3. Mechanism of SETD2-dependent generation of H3K36me3

Current models of mammalian SETD2-dependent K36-trimethylation have been strongly influenced by genetic and biochemical studies of yeast Set2. These reveal that Set2 recruitment and catalytic activity depend on a myriad of factors: kinase action on transcription elongation factors and the carboxy-terminal domain (CTD) of RNAPII (Li et al., 2002; Li et al., 2003; Xiao et al., 2003; Kizer et al., 2005; Chu et al., 2006; Zhou et al., 2009); histone chaperones Spt6 and Asf1 (Chu et al., 2006; Youdell et al., 2008; Lin et al.,

2010); specific components of the RNAPII-associated PAF complex (Krogan et al., 2003; Chu et al., 2007); Fpr4-mediated prolyl isomerisation of H3P38 (Nelson et al., 2006), and specific histone residues that affect overall H3 charge (Psathas et al., 2009) or engagement of Set2 with an appropriate nucleosomal surface contributed by residues in histones H3, H4 and H2A (Du et al., 2008; Du & Briggs, 2010).

Altogether, these studies are significant in highlighting the complex regulation of a single methyltransferase, and that the factors above are differentially required by Set2 to produce mono-, di- and tri-methylation of H3K36. Combined with the existence of demethylases that target different H3K36 methylation states (Tsukada *et al.*, 2006; Whetstine *et al.*, 2006; Klose *et al.*, 2007), these observations suggest biologically important regulation and functional significance of distinct methylation states of H3K36. However, given the diversification of SET2-homologs in mammals (section 1.2.1), and the specific function of SETD2 as a trimethyltransferase towards H3K36 *in vivo*, how far the regulatory mechanisms pertaining to Set2 in budding yeast can be extrapolated to mammalian SETD2 is unclear.

1.3.3.1. *SETD2 interacts directly with the carboxy-terminal domain of RNA polymerase II*

Different stages of transcription are associated with differential phosphorylation of the CTD of RNAPII, which coordinates co-transcriptional recruitment of histone modifiers and RNA processing factors (Komarnitsky *et al.*, 2000; Venters & Pugh, 2009). Understanding the temporal dynamics and spatial distribution of CTD phosphorylation (pCTD) is challenging due to the 52 consensus repeats (YSPTSPS) present in the CTD, population heterogeneity, and the potential adverse effects on antibody binding by epitope occlusion by closely-spaced phosphorylation (or additional modifications, such as methylation) or conformational changes by prolyl isomerisation (Phatnani & Greenleaf, 2006). Nevertheless, in general terms, phosphorylation at serine-5 by CDK7 in TFIIH is associated with transcription initiation, whereas processive elongation and production of processed transcripts is characterised by phosphorylation of serine-2 by CDK9 in pTEFb (Phatnani & Greenleaf, 2006).

The C-terminal region of human SETD2 was found to interact *in vivo* with phosphorylated forms of RNAPII. Whilst a partial N-terminal construct (residues 503−1498, lacking RS-like and proline-rich regions) did not associate with any form of RNAPII, a construct encompassing the C-terminal half (residues 1418–2564) co-immunoprecipitated with both RNAPII bearing phosphoserine-2 and RNAPII molecules with phosphoserine-5 (Sun *et al.*, 2005). The final ≈140 amino acids, comprising the Set2−Rpb1 interaction (SRI) domain, were found to be sufficient for the interaction with both phosphorylated forms (Sun *et al.*, 2005). Constructs minimally containing the QP-[0] domain (glutamine/proline-rich, low-charged domain—discussed in section 1.5.1) were also sufficient to co-immunoprecipitate (albeit less strongly) both phosphorylated forms of RNAPII (Sun *et al.*, 2005). Although no construct co-immunoprecipitated with non-phosphorylated RNAPII, this conclusion should be interpreted with caution in light of epitope occlusion problems with the 8WG16 (*α*-non-phospho-RNAPII) monoclonal antibody (discussed in (Brookes & Pombo, 2009)).

Although it is unclear whether SETD2 interacts *in vivo* with RNAPII simultaneously phosphorylated on serine-2 and serine-5 residues, or independent populations carrying exclusive phospho-epitopes, biophysical and structural characterisation of the SRI domain *in vitro* indicates that it binds preferentially to doubly phosphorylated CTD peptides spanning at least two repeats (Kizer *et al.*, 2005; Li *et al.*, 2005; Vojnic *et al.*, 2006). Phosphoserine-2 and phosphoserine-5 are likely to coincide on the same CTD repeats during the transition from transcription initiation to elongation. Additional contribution of phosphorylation of serine-7 of the CTD to SETD2−RNAPII interaction has yet to be investigated (Akhtar *et al.*, 2009).

The pCTD−SRI domain interaction therefore provides a potential mechanism to recruit SETD2 to elongating RNAPII for co-transcriptional H3K36me3 deposition. Indeed, DRB treatment reduces SETD2 cross-linking and H3K36me3 across *CD44* in HeLa cells, suggesting co-transcriptional recruitment of SETD2 (de Almeida *et al.*, 2011). However, no evidence exists for the direct importance of specific pCTD−SRI interactions for H3K36me3 generation in mammalian systems. Rather, the evidence is indirect, largely extrapolated from yeast (Li *et al.*, 2002; Krogan *et al.*, 2003; Li *et al.*, 2003; Schaft *et al.*, 2003; Xiao *et al.*,

2003; Kizer *et al.*, 2005), and supported by a conserved interaction of SET2-homologs with phosphoserine-5 in fission yeast and *Drosophila* (Morris *et al.*, 2005; Stabell *et al.*, 2007).

Like human SETD2, the yeast homolog associates with RNAPII that is predominantly phosphorylated on serine-2 and serine-5 of the CTD (Li *et al.*, 2002; Krogan *et al.*, 2003; Li *et al.*, 2003; Schaft *et al.*, 2003). The SRI domain was originally identified in yeast Set2 as a domain sufficient for direct interaction with hyperphosphorylated RNAPII *in vitro* and, critically, *in vivo* (Kizer *et al.*, 2005; Vojnic *et al.*, 2006). Ctk1, homologous to mammalian CDK9, mediates serine-2 phosphorylation during transcription elongation (Sterner *et al.*, 1995; Cho *et al.*, 2001). Since loss of global H3K36me3 is observed in *ctk1Δ* strains and in strains expressing Set2 without the SRI domain (Krogan *et al.*, 2003; Li *et al.*, 2003; Xiao *et al.*, 2003; Kizer *et al.*, 2005), a model was proposed whereby Ctk1-mediated CTD phosphorylation during transcription elongation recruits Set2 via the SRI domain, for co-transcriptional H3K36 methylation across coding regions.

1.3.3.2. *The relationship between CTD phosphorylation and SETD2 function is contentious*

Recent studies indicate that the pCTD–SRI domain recruitment model in yeast is too simplistic: neither Ctk1 nor the SRI domain is absolutely necessary for recruitment of Set2 to transcribed regions (Du *et al.*, 2008; Youdell *et al.*, 2008). Rather, the SRI domain, Ctk1 and CTD phosphorylation are required to maintain normal levels of Set2 protein and specifically for generation of H3K36me3 above H3K36me2 (Youdell *et al.*, 2008).

How might Set2 be recruited in the absence of a functional SRI domain? Both Set2 and SETD2 can interact (albeit weakly) with phosphorylated RNAPII in the absence of the SRI domain: mammalian SETD2 can weakly interact through its QP-[0] domain (Sun *et al.*, 2005), whereas Set2 can accomplish this through a WW domain–containing region (Phatnani *et al.*, 2004). Additionally, the PAF complex, which associates with elongating RNAPII, also contributes to Set2 function (Krogan *et al.*, 2003; Chu *et al.*, 2007), and a similar requirement is observed in human cells (Pirngruber *et al.*, 2009). The PAF complex appears to act as a scaffold for elongation factors and other chromatin modifiers during

transcription (Jaehning, 2010). Although the PAF-subunit CTR9, which is required for global H3K36me3 (Pirngruber *et al.*, 2009), distributes throughout the transcribed regions of active genes (Rahl *et al.*, 2010), no physical interaction between PAF components and SETD2 have been identified in yeast or mammals.

The pCTD–SRI coupling model is also inconsistent with quantitative distributions of H3K36me3 compared to phosphoserine-2. On a genome-wide scale, the progressive increase in the level of H3K36me3 across transcribed regions does not correlate with the extent of increase in phosphoserine-2, which appears to increase at a fixed distance from the transcriptional start site and remains relatively constant across the transcribed region (Kim *et al.*, 2010; de Almeida *et al.*, 2011). Furthermore, RNAPII can elongate through several kilobases of initial intronic sequence with little increase in H3K36me3, despite predicted co-incidence of phosphoserine-2 and phosphoserine-5 (Huff *et al.*, 2010; de Almeida *et al.*, 2011). In fact, the highest peaks of phosphoserine-2 generally occur downstream of the polyadenylation site, where H3K36me3 has already begun to decrease sharply (Barski *et al.*, 2007; Lian *et al.*, 2008; Rahl *et al.*, 2010). Furthermore, a mechanism based solely on the interaction of the SRI domain with the doubly phosphorylated CTD fails to explain increases in promoter-associated H3K36me2 and H3K36me3 during *MET16* induction, which appear before phosphoserine-2 and elongation (Morillon *et al.*, 2005).

1.3.3.3. *The IWS1–SPT6 complex mediates SETD2-dependent H3K36me3*

Human SETD2 is further linked to the elongating hyperphosphorylated RNAPII complex via interactions with the elongation factor IWS1 (Yoh *et al.*, 2008), which couples SETD2 indirectly to SPT6, another pCTD-binding protein, which preferentially interacts with phosphoserine-2 via its SH2 domain, as illustrated in **Figure 1.5** (Yoh *et al.*, 2007). SPT6 is a multi-functional protein that has histone H3–H4 chaperone activity (Kaplan *et al.*, 2003; Endoh *et al.*, 2004; Yoh *et al.*, 2007), which has promoted speculation that after passage of elongating RNAPII, SPT6/IWS1 may couple nucleosome reassembly with SETD2-mediated H3K36me3 (Yoh *et al.*, 2008). IWS1 appears to have a role in mediating SETD2 recruitment at the human *PABPC1* gene, and knockdown of *IWS1* in HeLa cells causes global decreases in H3K36me3 but not H3K36me2 (Yoh *et al.*, 2008). Intriguingly, disruption of any

component of the SETD2–SPT6/IWS1 complex causes nuclear accumulation of polyadenylated RNA, suggestive of defects in export, degradation or release of RNA from transcription sites (Yoh *et al.*, 2008).

1.3.3.4. *H3K36 trimethylation in vivo requires heterogeneous nuclear ribonucleoprotein L*

Despite all previously described interactions of SETD2 (**Figure 1.2**), heterogeneous nuclear ribonucleoprotein L (hnRNP-L) was the sole protein that co-purified with an N-terminally truncated FLAG-tagged SETD2 construct over-expressed in HeLa cells (Yuan *et al.*, 2009). The interaction was confirmed by co-immunoprecipitation of endogenous proteins, and recombinant binding assays defined interactions between multiple SETD2 C-terminal regions (**Figure 1.2B**) and redundant RNA recognition motifs in hnRNP-L (Yuan *et al.*, 2009). RNAi demonstrated that hnRNP-L is required *in vivo* to maintain global H3K36me3 levels, but its effect on SETD2 levels was not investigated (Yuan *et al.*, 2009). The mechanism by which hnRNP-L regulates SETD2 function *in vivo* is unknown. Although hnRNP-L does not discernably affect the K36-trimethyltransferase activity of SETD2 *in vitro* (Yuan *et al.*, 2009), it could plausibly contribute towards SETD2 recruitment or catalytic regulation *in vivo*.

1.3.3.5. *Influence of RNA splicing on H3K36me3 generation*

Transcription and pre-mRNA processing are intimately linked and in certain cases are physically coupled in a co-transcriptional manner (reviewed in (Hagiwara & Nojima, 2007)). RNA splicing has recently emerged as a process that may also influence SETD2-mediated generation of H3K36me3.

For analysis of three intron-containing human genes, inhibition of splicing caused a subtle reduction in SETD2 recruitment and H3K36me3 over internal exons (de Almeida *et al.*, 2011). Conversely, signal-inducible exon inclusion at the human *CD44* gene was associated with small increases in SETD2 and H3K36me3 levels over alternative, but not constitutive, exons (de Almeida *et al.*, 2011). However, since increases in SETD2 recruitment over alternative exons correlated with increased RNAPII levels (de Almeida *et al.*, 2011), it is

unclear whether splicing directly increases SETD2 recruitment or whether the enrichment is due to increased polymerase pausing, which may facilitate more efficient cross-linking of SETD2 to chromatin.

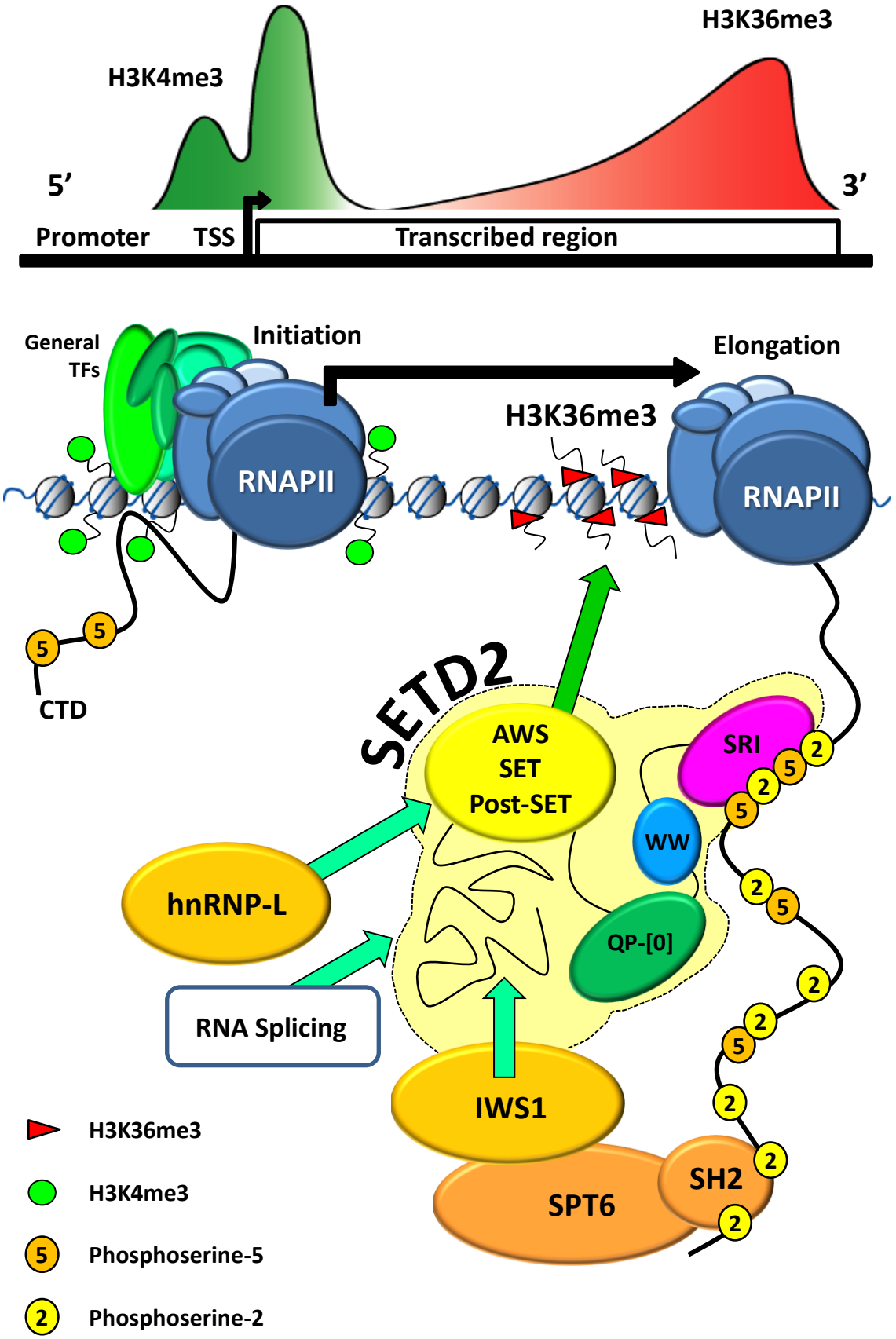
An independent study assessed the contribution of *cis*-regulatory sequences within *β-globin* to H3K36me3 distribution. Mutation of 3′ splice sites produced increases in H3K36me3 across a ≈3 kb region downstream of the polyadenylation site, which did not correlate with changes in RNAPII distribution (Kim *et al.*, 2011). Strikingly, mutation of the polyadenylation signal did not cause increases in these 3′ flanking regions, despite increases in RNAPII density downstream of the polyadenylation site (Kim *et al.*, 2011). The latter result is particularly interesting, because it suggests factors other than the polyadenylation signal must maintain and regulate the 3′ decrease near this region (**Figure 1.4**), despite transcription elongation continuing further downstream than normal. Additionally, this study showed that chemical inhibition of splicing causes redistribution of H3K36me3 to downstream regions on a genome-wide scale (Kim *et al.*, 2011), suggesting that appropriate spliceosome regulation is essential for normal SETD2-mediated H3K36me3.

Splicing cannot be an absolute requirement for SETD2 activity genome-wide, however, because intronless genes still exhibit low—but not insignificant—levels of H3K36me3 (de Almeida *et al.*, 2011). In fact, SETD2 is required for both basal levels and transcription-associated increases in H3K36me3 across the intronless murine c-*jun* gene (Edmunds *et al.*, 2008). Furthermore, experimentally induced changes in splicing of human *YPEL5* and *CD45* genes were not associated with changes in H3K36me3, indicating that RNA splicing does not solely determine exonic chromatin signatures and that gene-specific rules may also apply (Huff *et al.*, 2010).

In summary, current evidence suggests that many factors contribute towards SETD2-mediated K36-trimethylation across transcribed chromatin (summarised in **Figure 1.5**). Further investigation will be required, however, to understand more precisely how SETD2

Figure 1.5 Model for coupling SETD2-dependent H3K36me3 to transcription elongation by RNA polymerase II

Schematic representation of how SETD2 may deposit H3K36me3 in a co-transcriptional fashion through interaction with the elongating form of RNA polymerase II (RNAPII), which is phosphorylated on serine-2 and serine-5 of the carboxy-terminal domain (CTD). SETD2 may also interact indirectly with RNAPII via the IWS1–SPT6 complex. Abbreviations: SH2 (Src homology 2), TF (transcription factor), TSS (transcriptional start site).


H3K36me3
H3K4me3
5'
3'
Promoter
TSS
Transcribed region
General TFs
Initiation
Elongation
H3K36me3
RNAPII
RNAPII
5
5
CTD
SETD2
AWS
SET
Post-SET
SRI
WW
QP-[0]
hnRNP-L
RNA Splicing
IWS1
SPT6
SH2
2
H3K36me3
H3K4me3
Phosphoserine-5
Phosphoserine-2

recruitment and enzymatic activity are regulated to achieve appropriate spatial and temporal patterns of H3K36me3.

1.4. Proposed functions of SETD2-mediated H3K36me3

1.4.1. Roles of H3K36 methylation

Although direct effects of H3K36 methylation on nucleosome structure are unknown, various methyl-K36 binding proteins have been reported from yeast to humans (**Table 1.1**). In many cases their specificity towards different degrees of K36-methylation, alone or in combination with other histone modifications, in a nucleosomal context is unclear, along with their interaction and functional significance *in vivo*.

Many functions have been reported for Set2 homologs in yeasts, worms and flies (summarised in **Table 1.2**). Certain functions are organism-specific, unlikely to be relevant to homologous mammalian processes; for example, dosage compensation in *Drosophila* relies on upregulation of the male X-chromosome, via MOF acetyltransferase recruitment to sites in part marked with H3K36me3 (Larschan *et al.*, 2007; Bell *et al.*, 2008). The remainder of this chapter therefore focuses on function of SETD2/H3K36me3, most directly relevant to mammalian cell biology

An outstanding issue that pervades research of mammalian histone modification is the lack of clarity about the significance of modification of any specific residue of a histone. The function of modifiable residues may be probed in yeast by mutation to an amino acid that cannot be methylated (such as strains with H3K36R or H3K36A substitutions), but such genetic analysis is impractical in mammals due to multiple copies of histone genes. Furthermore, if a methyltransferase modifies non-histone proteins in additional to histones, a further challenge is to discriminate between the relative functional significance of modified substrates.

Related to this issue, much of our understanding of mammalian SETD2 relies on depletion by RNAi or from gene disruption. As discussed below in section 1.5, the multi-domain

Table 1.1 – H3K36-methyl-binding proteins

Protein	Domain	Organism	Function	References
BRPF1	PWWP	Human	Incorporated into MOZ acetyltransferase complex, implicated in gene activation	(Laue *et al.*, 2008) (Vezzoli *et al.*, 2010) (Wu *et al.*, 2011a)
CHD3	PHD	Human	Chromatin remodeler, component of NuRD co-repressor with HDAC activity	(Shi *et al.*, 2006) (Tong *et al.*, 1998)
DNMT3a	PWWP	Mouse	*De novo* DNA methylation	(Dhayalan *et al.*, 2010) (Wu *et al.*, 2011a)
Eaf3	Chromodomain	*S. cerevisiae*	Recruits Rpd3(S) deacetylase complex together with Rco1 subunit to suppress hyperacetylation and intragenic initiation	(Carrozza *et al.*, 2005) (Joshi & Struhl, 2005) (Keogh *et al.*, 2005) (Li *et al.*, 2007b)
Ecm5	PHD	*S. cerevisiae*	Contains Jumonij-C domain but no demonstrated KDM activity	(Shi *et al.*, 2007a)
HDGF	PWWP	Human	Transcriptional repressor	(Wu *et al.*, 2011a) (Yang & Everett, 2007)
JADE1	PHD	Human	May recruit HBO1 acetyltransferase complex to transcribed regions	(Saksouk *et al.*, 2009)
MRG15	Chromodomain	Human	Couples recruitment of PTB in alternative splicing Recruitment of Sin3B deacetylase in transcription elongation	(Luco *et al.*, 2010) (Zhang *et al.*, 2006) (Jelinic *et al.*, 2011)
MSH-3	PWWP	Human	Involved in DNA mismatch repair	(Vermeulen *et al.*, 2010)
MSH-6	PWWP	Human	Involved in DNA mismatch repair	(Vermeulen *et al.*, 2010)
MSL3	Chromodomain	*Drosophila*	Recruits MSL complex for dosage compensation of X-chromosome	(Larschan *et al.*, 2007) (Sural *et al.*, 2008)
MUM1	PWWP	Human	Involved in chromatin changes during DNA repair	(Wu *et al.*, 2011a)
NP-60	PWWP	Human	Putative transcription elongation role; required for p38 kinase activation	(Fu *et al.*, 2006) (Vermeulen *et al.*, 2010)
NSD1	PWWP	Human	H3K36 KMT; nuclear receptor co-regulator	(Kurotaki *et al.*, 2001) (Vermeulen *et al.*, 2010)
NSD2 (WHSC1)	PWWP	Human	H3K36 KMT; nuclear receptor co-regulator	(Li *et al.*, 2009b) (Vermeulen *et al.*, 2010) (Wu *et al.*, 2011a)
NSD3 (WHSC1L1)	PWWP	Human	H3K4/K27 KMT; functions unclear	(Vermeulen *et al.*, 2010) (Wu *et al.*, 2011a)
Nto1	PHD	*S. cerevisiae*	Facilitates chromatin association of NuA3 acetyltransferase in conjunction with Yng1–H3K4me3 interactions	(Martin *et al.*, 2006b) (Martin *et al.*, 2006a) (Shi *et al.*, 2007a)

Abbreviations: HDAC (histone deacetylase); KDM (lysine demethylase); KMT (lysine methyltransferase); PHD (plant homeodomain); PTB (polypyrimidine tract–binding protein); PWWP (proline-tryptophan-tryptophan-proline domain).

Table 1.2 – SETD2 homolog/H3K36 methylation function in non-mammalian systems

Process	Organism	Details	Reference
Gene activation	*S. cerevisiae*	H3K36me2/3 or Set2 needed for recruitment of NuA4 for H4K8ac during gene induction	(Morillon *et al.*, 2005)
NuA3 recruitment	*S. cerevisiae*	Recruitment of H3 acetyltransferase to chromatin; possibly via interaction with Nto1 subunit; gene activation	(Martin *et al.*, 2006b)
Ino2-dependent transcription activation	*S. cerevisiae*	Set2 directly binds Ino2 transcription factor through SET domain; required for Ino2-dependent gene activation	(Dettmann *et al.*, 2010)
Anti-silencing	*S. cerevisiae*	Dependent on H3K36me3 but functions independent of the Rpd3(S) complex	(Tompa & Madhani, 2007)
Gene repression	*S. cerevisiae*	Required for Tup1–Ssn6(Cyc8)-mediated repression; independent of KMT activity	(Tripic *et al.*, 2006)
Gene repression / attenuation	*S. cerevisiae*	*GAL10* non-coding RNA recruits Set2 and Rpd3(S) in *cis* to suppress *GAL1/GAL10* transcription under repressive conditions (glucose)	(Houseley et al., 2008)
TBP-binding at promoters	*S. cerevisiae*	Set2 negatively regulates TBP binding at promoters	(Biswas *et al.*, 2006)
Gene repression	*C. elegans*	MET-1 is required for H3K36me3 and gene repression	(Andersen & Horvitz, 2007)
Dosage compensation	*Drosophila*	H3K36me3 influences spreading of MSL complex along male X-chromosome for enhanced H4K16ac and up-regulated transcription; possibly through interaction with chromodomain of MSL3	(Bell *et al.*, 2008) (Larschan *et al.*, 2007) (Sural *et al.*, 2008)
Dosage compensation	*C. elegans*	MES-4 is involved in dosage compensation	(Bender *et al.*, 2006)
***Trans*-generational inheritance**	*C. elegans*	Role for MES-4 and H3K36me3 in epigenetic inheritance	(Furuhashi *et al.*, 2010) (Rechtsteiner *et al.*, 2010)
DNA replication	*S. cerevisiae*	H3K36me1 advances, while H3K36me3 delays, timing of Cdc45 association with replication origins *set2* mutations suppress hydroxyurea-sensitivity and negatively regulate DNA replication	(Pryde et al., 2009) (Biswas *et al.*, 2008)
Mitotic regulation	*S. pombe*	Release of Cdc14 from nucleolar chromatin (mitosis regulation)	(Hwang & Madhani, 2009)
Repression of meiotic recombination	*S. cerevisiae*	Loss of Set2 leads to elevated recombination at meiotic hotspots	(Merker *et al.*, 2008)
Transcription-coupled repair	*S. cerevisiae*	Recruitment of Rad26 to DNA damage in transcribed genes, dependent on RNAPII and enhanced by Set2	(Malik *et al.*, 2010)

Abbreviations: KMT (lysine methyltransferase); RNAPII (RNA polymerase II); TBP (TATA box–binding protein).

SETD2 protein appears to function as more than just a methyltransferase. Thus the phenotypes that result from modulation of *SETD2* expression cannot formally be ascribed solely to disruption of methyltransferase activity. Additionally, indirect versus direct effects are difficult to distinguish—especially for proteins that are intimately involved in regulation of gene expression.

1.4.2. Roles in transcription elongation: suppression of 'cryptic' transcripts?

1.4.2.1. *Set2–Rpd3(S) pathway and regulation of intragenic histone acetylation*

The function of SETD2 and H3K36me3 cannot be discussed without brief mention of the Set2–Rpd3(S) pathway in budding yeast. During transcription elongation, the Rpd3(S) histone deacetylase complex is recruited via phospho-CTD interactions, and deacetylates nucleosomes within transcribed regions (Drouin *et al.*, 2010; Govind *et al.*, 2010). Although H3K36me3 was originally thought to directly recruit Rpd3(S) (Carrozza *et al.*, 2005; Joshi & Struhl, 2005; Keogh *et al.*, 2005), Rpd3(S) recruitment is independent of Set2 function (Drouin *et al.*, 2010; Govind *et al.*, 2010). However, deacetylation *in vivo* requires Set2 and H3K36 methylation (Carrozza *et al.*, 2005; Joshi & Struhl, 2005; Keogh *et al.*, 2005). Interactions between the chromodomain-containing Eaf3 subunit of Rpd3(S) and H3K36me2/3-containing nucleosomes may stabilise the engagement of Rpd3(S) with its chromatin substrate or may putatively stimulate deacetylation by allostery (Li *et al.*, 2007b; Govind *et al.*, 2010). In *set2Δ* strains, Rpd3(S) activity is abrogated, resulting in global increases in H3 and H4 acetylation within transcribed regions (Carrozza *et al.*, 2005; Joshi & Struhl, 2005; Keogh *et al.*, 2005). For certain genes, this causes transcription initiation from cryptic intragenic promoters (Carrozza *et al.*, 2005; Keogh *et al.*, 2005; Cheung *et al.*, 2008; Pattenden *et al.*, 2010), and appears to promote antisense transcription from downstream promoters (Churchman & Weissman, 2011). Specific yeast strains that maintain H3K36me2 without H3K36me3 do not exhibit acetylation defects or intragenic initiation, suggesting that H3K36me2 is sufficient for Rpd3(S) activity (Du *et al.*, 2008; Youdell *et al.*, 2008; Li *et al.*, 2009a; Hainer & Martens, 2011).

How far this model can be extrapolated to higher eukaryotes is unclear. In contrast to the global effects on H3 and H4 acetylation in yeast, loss of H3K36me3 by dSet2 knockdown in *Drosophila* causes specific increases in global and gene-associated H4K16ac, while loss of both di- and tri-methylated K36 by RNAi against dMes-4 causes decreases in H4K16ac(Bell *et al.*, 2007). Although the existence of the Set2−Rpd3(S) pathway is often inferred in mammals, convincing evidence for involvement of SETD2 and H3K36me3 is lacking. Both a SIN3B deacetylase complex homologous to Rpd3(S) and components of acetyltransferase complexes (such as BRPF1 and JADE1) can interact specifically with H3K36me3-containing nucleosomes and associate with transcribed regions (Saksouk *et al.*, 2009; Vezzoli *et al.*, 2010; Jelinic *et al.*, 2011). However, the direct involvement of SETD2-mediated H3K36me3 on the functions of these deacetylase and acetyltransferase complexes has not been investigated; furthermore, knockdown of SIN3B components did not cause intragenic transcription (Jelinic *et al.*, 2011). Indeed, no changes in global or gene-associated histone acetylation are apparent after *SETD2* knockdown in mouse or human cells (Edmunds *et al.*, 2008; Yoh *et al.*, 2008). However, since many domains bind H3K36me2 and H3K36me3 comparably, and H3K36me2 is distributed over transcribed regions (Blackledge *et al.*, 2010), functions of H3K36me3 may arguably be obscured by redundant action of H3K36me2 which remains after *SETD2* knockdown (Edmunds *et al.*, 2008; Yoh *et al.*, 2008). Likewise, although H3K36me2 is sufficient for the deacetylation pathway in yeast, a contribution of H3K36me3 to the process cannot be excluded at present. Given current models, convincing demonstration of conservation of a Set2−Rpd3(S)-like pathway in mammals will require ChIP analysis of acetylated histones and proposed pathway factors, following depletion of H3K36me2, alone or alongside H3K36me3.

1.4.2.2. *SETD2–KDM5A pathway suppresses intragenic H3K4me3*

While Set2−Rpd3(S)-mediated repression of intragenic initiation in mammals is contentious, a role for SETD2-mediated H3K36me3 has recently been described in embryonic stem cells. H3K36me3 interacts with MRG15 to recruit an MRG15-KDM5B demethylase complex to remove intragenic H3K4me3 at KDM5B-target genes (Xie *et al.*, 2011). Knockdown of *SETD2* or *MRG15* causes loss of KDM5B from transcribed regions,

coincident with increases in intragenic H3K4me3; and loss of KDM5B is associated with cryptic transcription (Xie *et al.*, 2011). Considering H3K36me2 is not expected to be disrupted in this pathway, KDM5B recruitment via MRG15 may represent a specific function for H3K36me3.

1.4.3. Regulation of RNA processing

Transcription and chromatin structure are implicated in regulation of alternative splicing of pre-mRNA (reviewed in (Allemand *et al.*, 2008; Luco *et al.*, 2011)). As mentioned in section 1.3.2.2, H3K36me3 preferentially marks internal exons compared to introns in mammals and *C. elegans*. The generation and functional significance of this chromatin signature is an emerging and important topic; many details are still to be clarified however.

In mouse and human cells, H3K36me3 enrichment is lower on intronless genes compared to intron-containing genes: enrichment on the former appears independent of transcription, whereas for the latter, H3K36me3 levels positively correlate with expression level (de Almeida *et al.*, 2011). For example, no discernible differences in H3K36me3 were observed after changes in expression of the intronless *ZRSR1* gene (de Almeida *et al.*, 2011), which has been interpreted to mean that transcriptional induction is insufficient to mediate elevated H3K36me3 without splicing. However, this observation contradicts a previous observation of inducible, transcription-dependent increases in H3K36me3 at the murine intronless c-*jun* gene (Edmunds et al., 2008).

Differences in H3K36me3–exon marking are strongest for highly expressed genes in humans and mouse (Andersson *et al.*, 2009; Hon *et al.*, 2009; Schwartz *et al.*, 2009). Lower levels of H3K36me3 were found on alternative exons compared to single flanking constitutive exons in mice and *C. elegans* (Kolasinska-Zwierz *et al.*, 2009; Schwartz *et al.*, 2009) but little difference was found in humans (Spies *et al.*, 2009). However, when the inclusion/exclusion state of alternative exons in humans is also considered, H3K36me3 levels are higher on included exons compared to excluded ones (Andersson *et al.*, 2009; Hon *et al.*, 2009; Dhami *et al.*, 2010). More specifically, excluded exons contain lower

H3K36me3 than flanking constitutive ones, whereas H3K36me3 progressively increase across exons where the cassette is included (Hon *et al.*, 2009).

The mechanistic relationship between H3K36me3, exon definition and RNA splicing is currently contentious. The first study to establish a causal link between H3K36me3 and splicing demonstrated that *SETD2* knockdown or overexpression of yeast Set2 in human cells influenced cell type−specific exon exclusion within *FGFR2*, dependent on chromodomain-containing MRG51 and the PTB splicing repressor (Luco *et al.*, 2010). The authors proposed a model whereby SETD2-dependent H3K36me3 over an exon that is destined for PTB-dependent exclusion interacts with MRG15, which recruits PTB to promote exon exclusion (Luco *et al.*, 2010). Although association of H3K36me3 with exon exclusion conflicts with genome-wide correlations discussed above, the H3K36me3-MRG15-PTB-dependent pathway only contributes to a minority of splicing within the human genome (Luco *et al.*, 2010).

Inducible exon skipping within the murine *Ncam* gene is also associated with increases in localised H3K36me3 (Schor *et al.*, 2009). Furthermore, exons co-enriched with H3K36me3 and H3K9me3 are preferentially excluded compared to exons marked solely with H3K36me3 (Hon *et al.*, 2009). Although this is hard to reconcile with a recently reported role for H3K9me3 and HP1γ in promoting exon inclusion (Saint-Andre *et al.*, 2011), such observations are consistent with highly combinatorial and context-dependent read-out of histone modifications in other nuclear processes (Lee *et al.*, 2010).

The complexity of chromatin signatures related to exon−intron marking is further highlighted by the identification of groups of reciprocal modifications whose enrichment on exons or introns varies according to gene expression level and to whether exons are constitutive, or alternatively included or excluded (Dhami *et al.*, 2010). This study suggests the intriguing possibility that exons may be defined by histone modification prior to transcription.

As discussed in section 1.3.3.5, the actual mechanism for enrichment of H3K36me3 over exons versus introns is currently controversial. Additionally, the nature of the upstream

signals and mechanism that produce differential K36me3 deposition in specific cell types is unknown: Is SETD2 or a K36-demethylase differentially recruited or is enzymatic activity modulated across different exons? When is the differential exon modification generated? How stable is the distinguishing exonic enrichment—is it epigenetically inherited? Answers to these important questions will require further investigation of how endogenous SETD2 protein functions, in addition to the dynamics of H3K36me3.

Alongside connections to RNA splicing, SETD2 regulates correct polyadenylation of replication-dependent histone mRNA (Pirngruber et al., 2009). Furthermore, SETD2 knockdown in HeLa cells causes nuclear accumulation of poly(A)+ RNA, suggesting a role in RNA export or degradation (Yoh et al., 2008).

Consistent with a role in RNA processing, the serine- and arginine-rich N-terminal region of SETD2 contains clusters of RS dipeptides, akin to RS-domains in SR and SR-related proteins—factors that are ubiquitously associated with regulation of RNA splicing (Graveley, 2000; Boucher et al., 2001). Minimal RS domains from SR-related proteins are sufficient to function as splicing activators (Philipps et al., 2003), which may be relevant to the correlation between SETD2-mediated H3K36me3 and exon inclusion.

1.4.4. H3K36 methylation inhibits H3K27 methylation

Trimethylation of H3K27 is associated with gene repression, exemplified by the Polycomb system of *Drosophila* and the regulation of developmental gene expression in mammalian embryonic stem cells, required for their pluripotency and during their differentiation (Bantignies & Cavalli, 2006; Schwartz & Pirrotta, 2007). H3K27 methylation in mammals is catalysed by EZH2, the KMT component of Polycomb repressive complex 2 (PRC2) (Kuzmichev *et al.*, 2002). H3K27me3 forms a binding site for the recruitment of Polycomb repressive complex 1 (Fischle *et al.*, 2003; Schwartz & Pirrotta, 2007), which contributes to gene repression through mechanisms that likely involve chromatin compaction (Eskeland *et al.*, 2010) or ubiquitination of histone H2A (Stock *et al.*, 2007). Additionally, H3K27me3 is recognised by the C-terminal domain of EED, another constituent of PRC2, which may facilitate spreading and maintenance of H3K27me3 (Margueron *et al.*, 2009).

Trimethylation of H3K36 and H3K27 appear to have distinct associations with chromatin: H3K36me3 and H3K27me3 rarely co-exist on the same mononucleosomes or H3 molecules (Yuan *et al.*, 2011). Reciprocal patterns of K36 and K27 methylation occur on proximal versus distal immunoglobulin genes (Xu *et al.*, 2008a); and active versus inactive promoters in pluripotent versus differentiated cells (Mikkelsen *et al.*, 2007; Barrand & Collas, 2010). This reciprocal distribution may simply reflect the association of H3K27me3 with the absence of productive transcription elongation whereas H3K36me3 appears dependent on transcription elongation across actively transcribed genes, which are depleted of H3K27me3 (Barski *et al.*, 2007; Mikkelsen *et al.*, 2007). However, a more direct antagonistic relationship between methylation at H3K27 and H3K36 has recently been demonstrated by the observation that nucleosomes pre-methylated on K36 inhibit PRC2-dependent K27 methylation *in vitro* (Schmitges *et al.*, 2011; Yuan *et al.*, 2011). RNAi-mediated knockdown of *SETD2* in human cells is associated with subtle increases in global K27me3 (Yoh *et al.*, 2008; Barrand *et al.*, 2010) although variable effects have been described for promoter-associated H3K27me3 at specific genes in different cell types (Yoh *et al.*, 2008; Barrand *et al.*, 2010). Conversely, increased global levels of H3K36me2/3 resulting from NSD2 overexpression in multiple myeloma cells produce a decrease in global levels of K27me2/3 (Martinez-Garcia *et al.*, 2011). H3K36 methylation may thus have a role in protecting euchromatic regions from Polycomb-mediated silencing. However, since H3K36me2 appears more abundant and widespread than H3K36me3 (Garcia *et al.*, 2008; Blackledge *et al.*, 2010) and is sufficient to inhibit H3K27 methylation by PRC2 *in vitro* (Schmitges *et al.*, 2011), the relative contribution of SETD2-dependent H3K36me3 over H3K36me1/2 catalysed by other enzymes towards genome-wide H3K27 methylation is currently unclear.

1.4.5. Relationship with DNA methylation and heterochromatin

In addition to the post-translational modification of histones, chromatin is also subject to methylation of cytosine within the context of CpG dinucleotides. CpG distribution is skewed: largely depleted across the genome, but enriched in CpG islands often associated with promoters and regulatory elements (Cross & Bird, 1995; Caiafa & Zampieri, 2005; Deaton & Bird, 2011). Although methylation of such regions is conventionally associated

with gene silencing, genomic imprinting, and the inactivation of the mammalian X-chromosome (Newell-Price *et al.*, 2000; Paulsen & Ferguson-Smith, 2001), gene induction has also been found to involve dynamic cycles of DNA methylation (Kangaspeska *et al.*, 2008; Metivier *et al.*, 2008). Additionally, DNA methylation in gene bodies appears to regulate intragenic promoters (Maunakea *et al.*, 2010). Levels of cytosine methylation within transcribed regions are enriched over exons compared to introns, suggesting a potential involvement in exon definition or splicing (Chodavarapu *et al.*, 2010; Choi, 2010).

A correlation between the presence of H3K36me3 and cytosine methylation in transcribed regions has been noted in several studies (Chodavarapu *et al.*, 2010; Choi, 2010; Hawkins *et al.*, 2010). Although a plausible coupling of these modifications is suggested by interaction of the PWWP domain of DNMT3a methyltransferase with H3K36me3 (Dhayalan *et al.*, 2010), depletion of SETD2 by RNAi does not affect targeting of gene-body DNA methylation (Hahn *et al.*, 2011).

Despite the predominant association of SETD2-dependent H3K36me3 with transcribed regions, H3K36me3 has also been associated with heterochromatin formation. In *S. pombe*, Set2-dependent H3K36me3 is associated with transcription of—and required to maintain—centromeric heterochromatin, alongside mechanisms involving H3K9me3 and short interfering RNAs (siRNA) (Chen *et al.*, 2008). Furthermore, H3K36me3 has recently been reported at pericentromeric heterochromatin in murine ESCs and fibroblasts (Chantalat *et al.*, 2011) whilst H3K36me2 has also been identified at centromeres (Bergmann *et al.*, 2011). However, the H3K36me3 enriched in murine pericentromeric heterochromatin does not appear to be affected by knockdown of *SETD2* by RNAi (Chantalat *et al.*, 2011). Thus overall, a functional relationship between H3K36me3 catalysed by SETD2, heterochromatin formation and DNA methylation remains unclear at present.

1.5. Additional functions of SETD2

The SETD2 functions described above appear closely linked to downstream effects mediated by H3K36me3. Often, however, it is impossible to rigorously conclude whether phenotypes resulting from modulation of expression of a large multi-domain KMT are due

to methyltransferase activity, or other functions residing in other regions. SETD2 function in which involvement of KMT activity is less clear will now be described.

1.5.1. SETD2 as a putative transcription factor

Independent of RNAPII-interaction and KMT activity, the C-terminal region of SETD2 itself exhibits additional properties that are associated with transcription regulation.

Firstly, the final 225 residues of human SETD2 (**Figure 1.2B**) can specifically bind DNA from the promoter of the adenoviral *E1A* gene (Rega *et al.*, 2001). The initial *in vivo* observation by yeast one-hybrid assays was verified by electrophoretic mobility shift assays *in vitro* with both recombinant C-terminal fragments and endogenous SETD2 protein from HeLa nuclear extracts (Rega *et al.*, 2001).

Secondly, the C-terminal region of SETD2 contains a low-charged region, rich in glutamine and proline (QP-[0] domain; **Figure 1.2A**) that interacts modestly with RNAPII carrying phosphoserine-2 and phosphoserine-5 *in vitro*, and exhibits strong transactivation activity *in vivo* when expressed in isolation (Sun *et al.*, 2005). Intriguingly, this activity is moderately attenuated in constructs containing the additional C-terminal WW and SRI domains, and this is almost completely suppressed in near-full-length SETD2 and in C-terminal constructs that include the SET domain (Sun *et al.*, 2005). The compositional bias of this region is conserved among vertebrates, but is notably absent from unicellular eukaryotes.

Thirdly, a predicted coiled-coil motif lies immediately N-terminal to the QP-[0] domain in vertebrate SETD2 (**Figure 1.2A**); yeast Set2 also contains a coiled-coil motif within its C-terminus (Xiao *et al.*, 2003). Coiled-coils are found in a variety of transcription factors and are a common oligomerisation motif, which may be relevant to the putative homodimerisation of Set2 (Strahl *et al.*, 2002; Schaft *et al.*, 2003) and homotypic SETD2 interactions in yeast two-hybrid assays (Faber *et al.*, 1998). A further role in protein–protein interactions is supported by *in vitro* binding of hnRNP-L to the region downstream of the

SET domain, encompassing both the coiled-coil and QP-[0] domain (**Figure 1.2B**), independently of WW and SRI domains (Yuan *et al.*, 2009).

1.5.2. SETD2 regulates the p53 transcription network

p53 is an important factor regulating transcriptional responses to DNA damage, oncogene activation and oxidative stress (May & May, 1999; Jin & Levine, 2001). Its importance to human health and disease is highlighted by its involvement in Li-Fraumeni syndrome and its frequent mutation in a variety of cancers (Hollstein *et al.*, 1991).

Selective modulation of p53-target gene expression has been observed after overexpression or siRNA-mediated depletion of SETD2 (Xie *et al.*, 2008). For example, SETD2 promotes p53-dependent expression of anti-apoptotic genes *PUMA* and *NOXA*, cell-cycle-arrest gene *p21*, and anti-angiogenesis factor *TSP1*, but has little effect on other p53-targets (Xie *et al.*, 2008). A physical connection between SETD2 and p53 is supported by co-immunoprecipitation of endogenous proteins, and mapping of their interaction to the C-terminal region of SETD2 (**Figure 1.2B**) and the N-terminal transactivation domain of p53 (Xie *et al.*, 2008). The exact mechanism of how SETD2 regulates p53 function are unclear, but may involve its transcription factor–like properties (section 1.5.1); interaction with RNAPII (section 1.3.3.1); or direct methylation of p53, as observed for other KMTs (Chuikov *et al.*, 2004; Huang *et al.*, 2006; Shi *et al.*, 2007b; Huang *et al.*, 2010).

SETD2 also appears to repress transcription of *HDM2* (Xie *et al.*, 2008), which encodes an E3 ubiquitin ligase that regulates p53 stability and function (Jin & Levine, 2001). Consistent with a role in down-regulation of *HDM2*, *Setd2*$^{-/-}$ embryos display high levels of *Mdm2* transcripts (Hu *et al.*, 2010). A repressive role for SETD2 is perhaps surprising given its involvement in active transcription, but similar inhibitory characteristics have been observed in yeast (Strahl *et al.*, 2002; Landry *et al.*, 2003; Tripic *et al.*, 2006).

SETD2 involvement in p53 biology, and thus contribution to regulation of cell cycle checkpoints, DNA repair and angiogenesis, is likely relevant to its emerging role as a putative tumour suppressor gene (section 1.5.4).

1.5.3. SETD2 is required for vasculogenesis and embryonic development

Consistent with the ubiquitous, conserved association of SETD2 and H3K36me3 with the fundamental process of RNAPII-dependent transcription, SETD2 homologs are essential for normal growth and development in a variety of eukaryotes (Adhvaryu *et al.*, 2005; Morris *et al.*, 2005; Stabell *et al.*, 2007; Xu *et al.*, 2008b; Hu *et al.*, 2010; Yamada-Okabe *et al.*, 2010). In humans, mutations in SETD2 paralogs are implicated in syndromes with defects in physiological and neurological development, and cancer susceptibility; for example, NSD1 in Sotos overgrowth syndrome and Beckwith–Wiedemann syndrome; and WHSC1/NSD2 in Wolf–Hirschhorn syndrome (Stec *et al.*, 1998; Kurotaki *et al.*, 2002; Morishita & di Luccio, 2011).

In mammalian development, murine *Setd2*$^{-/-}$ embryos exhibit lethality at embryonic day E9.5-10.5 with defective blood vessel formation throughout the embryo, yolk-sac and placenta (Hu *et al.*, 2010). *Setd2*$^{-/-}$ yolk-sacs exhibit altered transcription of genes encoding secreted factors and membrane proteins implicated in angiogenesis (Hu *et al.*, 2010). Further supporting involvement in vascular remodelling, *Setd2*$^{-/-}$ embryoid bodies fail to form proper vessels *in vitro*, and endothelial cells subject to *SETD2* knockdown by RNAi display defects in cell migration, vasculogenesis and transcriptional regulation of important angiogenic factors (Hu *et al.*, 2010). Understanding whether SETD2 functions directly as a transcriptional regulator in this process requires further investigation. Importantly, appropriate blood vessel development is critical not only for embryogenesis but also for tumour development and metastasis.

1.5.4. *SETD2* as a putative tumour suppressor gene

The clearest example of *SETD2* disruption in human disease is its significant association with clear-cell renal cell carcinoma (ccRCC), the most prevalent form of adult kidney cancer (Dalgliesh *et al.*, 2010; Duns *et al.*, 2010). Exome-sequencing of ccRCC tumours identified frequent nonsense and frameshift mutations, predicted to produce truncated proteins (Dalgliesh *et al.*, 2010). Loss of SETD2 function is further suggested by an independent study of ccRCC-derived cell lines, where aberrant splicing, and frameshift,

nonsense and missense mutations are associated with depletion of global H3K36me3 (Duns *et al.*, 2010).

Although ccRCC only accounts for a minority of cancer, several additional observations suggest a wider contribution of *SETD2* to tumourigenesis. Firstly, lower *SETD2* mRNA and protein levels correlate with the severity of progressive breast cancer (Al Sarakbi *et al.*, 2009; Newbold & Mokbel, 2010). Secondly, *SETD2* was identified as a candidate tumour suppressor gene in an RNAi screen for metastasis suppression (Gobeil *et al.*, 2008). Finally, a high frequency of transposon insertion into *SETD2* has been reported in mouse models of pancreatic cancer (Varela *et al.*, 2011).

SETD2 inactivation suggests a putative tumour suppressor function: although direct mechanisms are currently unclear, SETD2 involvement in vasculogenesis (Hu *et al.*, 2010), p53 biology (Xie *et al.*, 2008) and alternative splicing (Luco *et al.*, 2010) are likely important. Interestingly, missense mutations outside of the SET domain, which are unaccompanied by H3K36me3 depletion, suggest that disruption of KMT-independent function may be sufficient for contribution to ccRCC (Duns *et al.*, 2010).

Relevant to this discussion, SET2-related paralogs with K36-dimethyltranferase activity are implicated in a range of cancers (Morishita & di Luccio, 2011); for example, NUP98 fusions to NSD1 or NSD3 in acute myeloid leukaemia, and IgH–NSD2 fusion in multiple myeloma. Furthermore, modulation of chromatin modifiers and transcription factors is a common strategy of oncogenic viruses (reviewed in (Flanagan, 2007)): adenovirus may utilise SETD2 as a transcriptional regulator (Rega *et al.*, 2001) whilst yeast two-hybrid assays have identified high connectivity of human SETD2 with proteins of Hepatitis C virus (de Chassey *et al.*, 2008).

1.5.5. SETD2 interacts with Huntingtin

Perhaps the most surprising aspect of SETD2 is its original identification—not as a transcription-associated factor, nor methyltransferase—but as 'HYPB': an interactor of Huntingtin (Faber *et al.*, 1998).

Huntingtin is a large (>350 kDa), ubiquitously expressed protein, which is critical for mammalian development and neuron function (Duyao *et al.*, 1995; Nasir *et al.*, 1995). Despite containing no clear functional domains other than polyglutamine/proline-rich sequences and HEAT repeats confined to its N-terminal region, it interacts with over 100 proteins (Goehler *et al.*, 2004) and is implicated in numerous processes: (i) organisation of endomembranes, cytoskeleton and vesicular transport (Gutekunst *et al.*, 1995; Caviston *et al.*, 2007; Smith *et al.*, 2009); (ii) iron transport and homeostasis (Hilditch-Maguire *et al.*, 2000; Lumsden *et al.*, 2007); (iii) modulation of MAPK signalling (Apostol *et al.*, 2006); and (iv) regulation of gene expression (Zuccato *et al.*, 2003; Benn *et al.*, 2008), which may be mediated through interactions with promoters, transcription factors (p53, NCoR, NFκB, CtBP) and RNA processing factors (PRPF40A, Symplekin, SETD2) (Faber *et al.*, 1998; Kegel *et al.*, 2002; Li & Li, 2004).

Huntington's disease (HD) is caused by an expansion of CAG trinucleotide repeats in exon 1 of the *HTT* gene that produces an N-terminal polyglutamine extension beyond 35 repeats (Duyao *et al.*, 1993; Macdonald *et al.*, 1993). HD is a debilitating neurodegenerative disorder involving dementia and severely impaired motor coordination that affects 1 in ≈15,000 individuals (Walker, 2007). Mutant Huntingtin aggregates into amyloid-like structures in neurons undergoing cell death, which likely interferes with normal protein–protein interactions, though the exact pathological mechanisms are contentious (Davies *et al.*, 1997).

SETD2 interaction with N-terminal fragments of Huntingtin was identified through multiple independent and reciprocal yeast two-hybrid experiments. The interaction clones corresponded to C-terminal regions of SETD2 that encompass either the WW domain or QP-[0] region (**Figure 1.2B**) (Faber *et al.*, 1998; Kaltenbach *et al.*, 2007). Recombinant WW domain–containing constructs bind polyproline (Hu *et al.*, 2004; Kato *et al.*, 2004) and interact with endogenous full-length forms of both wild-type and mutant Huntingtin *in vitro*, dependent on Huntingtin's proline-rich region, and promoted by polyglutamine expansion (Faber *et al.*, 1998; Passani *et al.*, 2000). Furthermore, both proteins colocalise *in vivo* with nuclear inclusions and auto-fluorescent clusters in post-mortem striatum of HD

patients (Passani *et al.*, 2000). Further functional association between the proteins is supported by the observation that disruption of murine *Htt* causes partial delocalisation of SETD2 from complete nuclear staining to cytoplasmic puncta observed by immunofluorescence (Hilditch-Maguire *et al.*, 2000).

Relevant to other polyglutamine-associated neurodegenerative diseases, SETD2 also interacts with both Ataxin-1 and Capicua-homolog-1 (Lim *et al.*, 2006)—mutations in which are responsible for spinocerebellar ataxia type 1 (SCA1) (Lam *et al.*, 2006). Furthermore, even though Huntingtin in zebrafish lacks N-terminal polyglutamine (Karlovich *et al.*, 1998), the SETD2 homolog itself contains a curious CAG/CAA-expansion encoding 48 consecutive glutamine residues within its QP-[0] domain.

The exact role of SETD2 in normal Huntingtin function and HD pathology is poorly characterised at present, but there are intriguing functional links between Huntingtin and certain processes in which SETD2 is implicated. Firstly, *HTT* transcription in human cells is up-regulated by p53 and requires the presence of SETD2 (Feng *et al.*, 2006; Xie *et al.*, 2008). Secondly, mutant Huntingtin reciprocally binds to and upregulates p53 and modulates responses to cellular stress (Bae *et al.*, 2005; Illuzzi *et al.*, 2011). Thirdly, nuclear Huntingtin interacts with proteins implicated in RNA splicing and polyadenylation (Faber *et al.*, 1998; Takagaki & Manley, 2000)—processes that are modulated by SETD2 (discussed in section 1.4.3). Finally, Huntingtin is required for normal levels of H3K27me3 during murine development; it interacts with PRC2, and directly stimulates generation of H3K27me3 *in vitro* and *in vivo* (Seong *et al.*, 2010). This latter association is intriguing given the emerging antagonistic roles of methylation of H3K27 and H3K36 (discussed in section 1.4.4). However, the full complement of Huntingtin-containing complexes is currently unclear: it is unknown whether SETD2 is also incorporated into the same Huntingtin-containing complex as PRC2. Furthermore—and critically—the actual identity of the endogenous SETD2 protein that interacts with Huntingtin *in vivo* was never conclusively determined: it is thus unknown whether Huntingtin interacts with the full-length methyltransferase or an alternative protein comprising either or both of the C-terminal QP-[0] or WW domains.

1.6. Diversity of transcriptional and translation products of the *SETD2* gene

1.6.1. Understanding biological diversity between transcriptome and proteome

Genome sequencing has provided unprecedented understanding of DNA sequence, organisation and evolution yet the functional output of the genome is still under intense investigation (Lander, 2011). It was realised early that knowledge of the sequence of a genome does not give us a complete understanding of how genes function: delineating transcript composition by computational methods alone is difficult and a contemporary appreciation of the prevalence of non-coding transcription undermines conventional understanding of the physical nature of a gene (Carninci *et al.*, 2005; van Bakel & Hughes, 2009).

Genome-wide RNA-sequencing (RNA-seq) and microarray studies highlight the complexity and diversity of transcription, resulting from alternative splicing (Pan *et al.*, 2008; Wang *et al.*, 2008a), polyadenylation (Shepard *et al.*, 2011) and RNA editing (Rosenberg *et al.*, 2011). By current estimates, 90–95% of mammalian genes produce at least one alternatively spliced variant (Pan *et al.*, 2008; Wang *et al.*, 2008a). Although RNA-sequencing confirms the prevalence and complexity of alternative splicing, a substantial gap remains in understanding the true nature of the population of full-length transcripts from which RNA-sequencing reads derive: the technique provides detailed sequence information of fragments of transcripts, yet reveals limited information about the complete structure of individual, full-length molecules (Ozsolak & Milos, 2011).

Even less well understood is the functional significance of transcript variants on a genome-wide scale. Transcript diversity may generate protein isoforms with different domain composition and thus function. Alternatively, comparable proteins may be produced from transcripts that only differ in untranslated regions (UTRs), which may regulate protein synthesis or transcript stability (Moore & Proudfoot, 2009). Alternative transcripts may be generated, which themselves, may not contribute towards protein generation, but which

are degraded by nonsense-mediated decay, and are thus implicated in regulation of levels of gene expression (Lareau *et al.*, 2007; Ni *et al.*, 2007; Saltzman *et al.*, 2008).

Unless directly investigated, the existence and primary structure of a protein is largely inferred from transcript models, which are traditionally derived from *in silico* assembly from expressed sequence tags (ESTs) and cDNAs from diverse sources (Heber *et al.*, 2002). At the proteome level, direct investigation by mass spectrometry can occur in two formats: (i) bottom-up approaches provide sequence information from proteolytic fragments, often from heterogeneous complex mixtures of cellular proteins; (ii) top-down analysis provides information on protein structure derived from fragmentation of isolated intact proteins within the mass spectrometer (Guerrera & Kleiner, 2005). The former technique is limited in its ability to ascribe peptides to specific isoforms deriving from alternatively spliced transcripts (Godovac-Zimmermann *et al.*, 2005). Although the top-down approach has the advantage that individual isoforms can be identified by assembling constituent peptide fragments, this technique requires intensive purification to homogeneity, which is often challenging for proteins of low abundance within a complex mammalian proteome (Rabilloud *et al.*, 2010; Chait, 2011).

Despite the high sensitivity, resolution and capacity of modern mass spectrometry systems, proteomics of mammalian cells is still difficult due to proteomic diversity, heterogeneity of post-translational modification, and the wide dynamic range of protein expression (Rabilloud *et al.*, 2010). Akin to the differences in information provided by RNA-sequencing versus northern blotting of transcripts, mass spectrometry provides sequence information of protein fragments with limited information on the full-length structure of—and thus potential diversity of—protein isoforms, whereas western blotting permits analysis of size and provenance of isoforms, but with limited sequence information (Rabilloud *et al.*, 2010).

Fundamental important questions therefore remain in the post-genomic era: What exactly are the primary structures of full-length RNA transcripts? How does transcript diversity relate to potential to encode variant protein isoforms? What exactly are the primary

structures of full-length endogenous proteins? What is the functional significance of protein isoform diversity? Although these are challenging and wide-reaching questions that, in general terms, are beyond the scope of this thesis, it is within this context that SETD2 is here investigated. Many examples exist of functionally distinct isoforms pertaining to transcription factors (Lopez, 1995), and some notable examples for chromatin modifiers are presented in **Table 1.3**. Relevant to this thesis, certain genes that produce proteins with chromatin modifying activity also generate isoforms that appear to lack such activity (Sakamoto *et al.*, 2000; Frescas *et al.*, 2008; Verrier *et al.*, 2011). Up to this point, SETD2 has been discussed as a single ≈290 kDa protein; however, evidence exists for variant forms.

1.6.2. Evidence for multiple forms of SETD2

Although SETD2 is implicated in a variety of fundamental processes as described above, many mechanistic and functional aspects of the mammalian protein require further investigation. A fundamental unresolved issue is the nature of endogenous forms of SETD2 itself. Although the canonical SETD2 protein predicted from reference transcripts is expected to be ≈290 kDa (**Figure 1.2**), alternative proteins differing in molecular mass have been identified by immunoblotting and mass spectrometry (**Table 1.4**). Significantly, C-terminal SETD2 antibodies specifically recognise multiple distinct bands within and between different cell types (Passani *et al.*, 2000). Mass spectrometry has identified a 33 kDa species in human platelets (Garcia *et al.*, 2004) and a ≈140 kDa form in a purified retinoic acid receptor–containing complex involved in granulopoiesis (Fujiki *et al.*, 2009). The presence of SETD2 in enucleate cells, combined with reports of its localisation to cytoplasm and secretory vesicles (Hilditch-Maguire *et al.*, 2000; Passani *et al.*, 2000; Muller *et al.*, 2007; Wegrzyn *et al.*, 2010), suggest potential functions unrelated to chromatin and transcription regulation. Reports of differently sized transcripts (Faber *et al.*, 1998; Rega *et al.*, 2001) and evidence for alternative splicing from ESTs, cDNA and RNA-sequencing (discussed further in Chapter 5) further support SETD2 diversity.

Table 1.3 – Proteomic diversity of chromatin modifiers

Gene	Functional diversity	Reference
KAT5 *(TIP60)* *Histone acetyltransferase*	Alternative exon exclusion generates an isoform lacking a proline-rich segment, which localises to the nucleus and cytoplasm. The canonical form retains the proline-rich segment and is exclusively localised to the nucleus.	(Ran & Pereira-Smith, 2000)
HDAC9 *Histone deacetylase*	Several alternatively spliced forms with differential tissue expression; protein isoforms exhibit different subcellular localisations and interactions with co-repressors.	(Zhou *et al.*, 2001) (Petrie *et al.*, 2003)
HDAC7 *Histone deacetylase*	Alternative intron retention generates isoforms with different subcellular localisations and effects on transcription factors for smooth muscle development.	(Margariti *et al.*, 2009)
PRMT1 *Arginine methyltransferase*	Seven isoforms resulting from alternative splicing, with different substrate specificities, catalytic activities and subcellular localisations.	(Goulet *et al.*, 2007)
EHMT2 *(G9a)* *H3K9 methyltransferase*	Alternative splicing produces isoforms with different subcellular localisation.	(Brown *et al.*, 2001)
MLL5 *H3K4 methyltransferase*	≈200 kDa and ≈75 kDa forms generated by different polyadenylation; only the latter co-activates RARα in granulopoiesis.	(Fujiki *et al.*, 2009)
NSD1 *H3K36 methyltransferase*	Cell type–specific alternative splicing (intron-retention) generates short/long forms—variable N-terminus. Unclear significance.	(Kurotaki *et al.*, 2002) (Lucio-Eterovic *et al.*, 2010)
WHSC1 *(NSD2/MMSET)* *H3K36 methyltransferase*	Three isoforms (MMSET I, MMSET II and RE-IIBP) generated by alternative promoters and splicing. Differ in subcellular localisation and effect on transcription regulation of specific target genes.	(Garlisi *et al.*, 2001) (Kim *et al.*, 2008) (Todoerti *et al.*, 2005)
WHSC1L1 *(NSD3/WHISTLE)* *H3K4/K27 methyltransferase*	WHISTLE as a truncated, alternatively spliced form of NSD3(WHSC1L1). Unclear significance.	(Angrand *et al.*, 2001) (Kim *et al.*, 2006)
SUV420H1 *H4K20 methyltransferase*	Two isoforms differ by C-terminal truncation, exhibiting different subcellular localisation and chromatin targets.	(Tsang *et al.*, 2010)
KDM2A *H3K36me2 demethylase*	Intragenic promoter produces SF-KDM2A, lacking N-terminal demethylase domain and KDM activity, but retains interaction domains.	(Frescas *et al.*, 2008)
KDM4A *H3K9me2/3 demethylase*	Intragenic promoter produces isoform without Jumonji demethylase domain; specifically required for muscle development.	(Verrier *et al.*, 2011)
KDM1A *(LSD1)* *H3K4me1/2 demethylase*	Alternative splicing produces four different isoforms, different tissue expression, similar demethylase activity	(Zibetti *et al.*, 2010)
CHD7 *Chromatin remodeler*	Alternative splicing produces an isoform that contains N- and C-terminal regions, but lacks ≈250 kDa-worth of central domains. Function unclear.	(Colin *et al.*, 2010)
CHD8 *Chromatin remodeler*	N-terminal-specific isoform lacks remodelling domains, but has novel β-catenin-binding activity	(Sakamoto *et al.*, 2000)

KDM (lysine demethylase); MMSET (multiple myeloma SET); NSD (nuclear receptor and SET domain–containing); WHSC (Wolf–Hirschhorn syndrome candidate).

Many SETD2 interactions are supported by over-expression of truncated forms *in vivo* or by analysis of similarly modified, recombinant constructs *in vitro*, which may potentially reflect functions of specific isoforms with different domain composition. Notably, although convincing evidence for *in vivo* SETD2–Huntingtin interaction exists, the nature of the endogenous SETD2 form that specifically interacts with Huntingtin has never been reported: does Huntingtin interact with a full-length 290 kDa methyltransferase, or a truncated C-terminal variant that may lack KMT activity? The potential existence of protein isoforms complicates our understanding of the relationships between domains and functions within SETD2, and its reported interactions with other cellular components. Characterisation of the endogenous forms of SETD2 is therefore immediately important to understand the mechanisms and significance of SETD2 function as described above, and thus directly relevant to clarifying its contribution to human health and disease.

Table 1.4 – Multiple forms reported for mammalian SETD2 proteins

Mass (kDa)*	Organism	Source	Technique	Reference
33	Human	Platelets	Mass spectrometry	(Garcia *et al.*, 2004)
< 55	Human	HeLa cells	Immunoblot (N-terminal serum‡)	(Rega *et al.*, 2001)
≈130	Human	HL60 cells	Mass spectrometry	(Fujiki *et al.*, 2009)
> 150	Human	HEK293 cells	Immunoblot†	(Yuan *et al.*, 2009)
> 210	Human	HeLa cells	Immunoblot (N-terminal serum‡)	(Rega *et al.*, 2001)
≈420 **≈285** **≈200**	Human	Lymphoblastoid cells	Immunoblot (affinity-purified BF1§)	(Passani *et al.*, 2000)
≈420 **≈380**	Human	Frontal cortex	Immunoblot (affinity-purified BF1§)	(Passani *et al.*, 2000)
≈290 **≈200** **≈150**	Cow	Adrenal medulla (neuroendocrine secretory granules)	Immunoprecipitation-Immunoblot	(Wegrzyn *et al.*, 2010)
≈490	Mouse	Whole brain	Immunoblot (affinity-purified BF1§)	(Passani *et al.*, 2000)

* Estimated molecular mass. Greater/less than symbols used for proteins extrapolated beyond molecular weight markers by SDS-PAGE.

† Immunoblot confirmation of silver-stained band after purification of FLAG-tagged hnRNP-L complex.

‡ Raised against N-terminal region peptide MERRGKYSSLERS (comparable to N-terminal epitope in **Figure 1.2**).

§ BF-1 antibody, raised against purified recombinant C-terminal 178 residues (up to and including the WW domain).

1.7. Aims and overview of thesis

Various transcriptional and protein forms of SETD2 have been previously reported, in a variety of systems, but without clear validation, comparability, or further research into their prevalence, identity, function or significance. Do *bona fide* alternative SETD2 isoforms exist? How are different forms generated? What are the implications for SETD2 function? The answers to these questions are currently unclear, yet crucially influence our understanding of endogenous SETD2 function *in vivo*.

This thesis therefore aims to identify and characterise alternative transcriptional and protein isoforms of mammalian SETD2, focusing primarily on forms expressed in murine C3H 10T½ fibroblasts.

In overview, Chapter 3 employs specific α-SETD2 antibodies and RNAi to characterise endogenous SETD2 isoforms in cultured mammalian cells. Chapter 4 further characterises these proteins by subcellular fractionation, electrophoretic and proteomic techniques. Chapter 5 explores the transcriptional diversity of the murine *Setd2* gene whilst Chapter 6 investigates an ectopically expressed isoform encoded by a novel transcript variant. Implications for our understanding of SETD2, the wider significance of findings, and future work are discussed in Chapter 7.

Chapter 2 Materials and methods

2.1. Materials and general procedures

Analytical-grade reagents were obtained from Sigma or VWR unless stated otherwise. For RNA analysis, all glassware was baked at 180°C for 4 hours and reagents were treated, where appropriate, with dimethylpyrocarbonate (DMPC) to inactivate ribonucleases. Unless stated otherwise, all procedures were performed at room temperature (23°C) and all centrifugation was performed in a Heraeus microfuge (Biofuge Fresco).

2.1.1. PCR primers

PCR primers were designed using Primer 3.0 (Rozen & Skaletsky, 2000) and specificity checked against mouse genome and cDNA sequences by BLAST (Altschul *et al.*, 1997). All primers were synthesised and purified by SePOP (Selective Precipitation Optimized Process)–desalting by Eurogentec.

2.1.2. Peptides and antibodies

The location of N-terminal, C-terminal and CT21 epitopes are indicated in **Figures 1.2A** and **1.3**. In-house antisera to N- and C-terminal regions of murine SETD2 (RefSeq NP_001074809.2) were previously raised by PTU/BS (Pentlands Science Park, Penicuik, Midlothian, UK) in rabbit and sheep immunised with SETD2 N-terminal and C-terminal peptides, respectively, as detailed in **Table 2.1**. Peptides contained an additional C-terminal cysteine for conjugation to keyhole limpet haemocyanin (Edmunds *et al.*, 2008).

2.2. Peptide affinity–purification of antisera

2.2.1. Generation of peptide-coupled affinity columns

SETD2 peptide–affinity columns were generated using SulfoLink® Immobilization Kit for Peptides (Thermo) according to the manufacturer's instructions. Two milligrams of peptide (**Table 2.1**) were coupled per column at approximately 50% and 95% efficiency for SETD2

N-terminal and C-terminal columns, respectively. Columns were stored in phosphate-buffered saline (PBS) pH 7.4 with 0.05% (*v*/*v*) sodium azide at 4°C in the dark.

Table 2.1 – α-SETD2 antibodies

Antibody	Peptide sequence	Residues*	Coding Exons†	Identity with human SETD2‡	Source
N-term	MERRGKYTSKLERESKRTSE-C	504–523	3	95%	NSL R2039 final bleed
C-term	RGTKRPDDRYDTPTSKKKVR-C	2063–2082	14–15	100%	NSL S950B final bleed
CT21	C-QKFGAVYKPKEDTE	2522–2535	21	100%	Everest Biotech eb07418

* from murine SETD2 protein (UniProt KB accession E9Q5F9, 2537 amino acids).
† exon numbers correspond to murine *Setd2* transcript (RefSeq: NM_001081340.2).
‡ human SETD2 protein (UniProt KB accession Q9BYW2, 2564 amino acids).

2.2.2. Affinity purification of α-SETD2 antisera

Antisera were clarified by centrifugation at 16,000*g* for 30 minutes and supernatants were passed through a 0.22 μm filter (mixed cellulose esters; MILLEX®-GS, Millipore) to remove remaining particulate material. Typically, 700 μl of α-SETD2 N-terminal rabbit serum or 400 μl of α-SETD2 C-terminal sheep serum were then diluted drop-wise with PBS to a final volume of 3.3 ml. Peptide–affinity columns were washed sequentially by centrifugation at 1000*g* for 2 minutes after rotating for 2 minutes each with: 2 ml PBS pH 7.4, 2 ml 10 mM HCl pH 2.0, 8 ml PBS, 2 ml 6 M guanidinium chloride, and finally 8 ml PBS. Columns were equilibrated with 6 ml PBS pH 7.4 for 10 minutes, before centrifugation as above, and then end-over-end rotation with diluted serum for 30 minutes. Unbound material was collected by centrifugation and was pooled with additional flow-through from 1 ml of PBS. Columns were washed 10 times with 2 ml PBS; wash fractions were collected separately. Four elution fractions were obtained by addition of 2 ml 0.2 M glycine–HCl pH 2.5 into pre-chilled tubes containing 400 μl 1 M Tris–HCl pH 9.0. All fractions were thoroughly mixed on ice, adjusted to 0.05% (*v*/*v*) sodium azide and stored at 4°C.

2.3. Cloning of epitope-tagged *Setd2* expression constructs

2.3.1. Plasmids and PCR primers

Plasmids and clones for generation of epitope-tagged *Setd2*(13»21) expression constructs are indicated in **Table 2.2**. Annotated sequences of epitope-tagged *Setd2*(13»21) constructs are provided in **Figure 6.1**.

Primers were designed to amplify a novel open reading frame (ORF) identified in *Setd2* clone CA319728 that initiates at the boundary of exons 12b and 13, and terminates in exon 21 (**Table 2.3** and **Figure 5.9A**). For directional cloning into either pcDNA3–HA or p3×FLAG–CMV-10 epitope-tagging vectors, forward primers contained a 5′ flanking sequence with a HindIII site whilst reverse primers contained an EcoRI site. The HinKoz13F1 primer contained a Kozak consensus sequence downstream of the HindIII site for translation initiation of C-terminal HA-tagged protein.

2.3.2. PCR amplification of *Setd2*(13»21) and cloning into pJET1.2/blunt vector

Setd2(13»21) DNA with appropriate flanking sequences for cloning and expression was amplified from purified *Setd2* clone BC059049 using Expand High Fidelity PCR system (Roche) and appropriate primers (**Table 2.3**). A 50 µl PCR reaction contained: 50 ng template in 1× TE buffer (10 mM Tris–HCl pH 8.0, 1 mM EDTA), 1× Expand High Fidelity PCR buffer (without $MgCl_2$), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.3 µM forward primer, 0.3 µM reverse primer, 0.5 µl Expand enzyme mix. Thermocycling conditions were: 94°C for 5 minutes; 25 cycles of [94°C for 1 minute; 57°C for 1 minute; 68°C for 1 minute]; 72°C for 7 minutes. PCR products corresponding to *Setd2*(13»21)–HA and FLAG–*Setd2*(13»21) were purified by gel extraction (section 2.17.2) and cloned into pJET1.2/blunt vector (**Table 2.2**) using the blunting protocol according to the manufacturer's instructions. Chemically competent *Escherichia coli* (NEB-10β) were transformed with ligation mixture by the heat-shock method according to the manufacturer's instructions. Transformants were selected overnight at 37°C on LB–agar plates containing 100 µg/ml ampicillin, and screened for correct inserts by PCR and restriction digestion using EcoRI-HF and HindIII (Promega).

Table 2.2 – Plasmids and clones

Plasmid / Clone	Derivative vector	Description	Source
BC059049	pYX-Asc	Mouse *Setd2* (exons 12»21) cDNA IMAGE clone	Source Bioscience
CA319728	pYX-Asc	Mouse *Setd2* (exons 12−12b−13»21) cDNA IMAGE clone	Source Bioscience
pJET1.2/blunt	-	PCR cloning vector	Fermentas
pcDNA3−NHK−HA	pcDNA3.1	Cytomegalovirus (CMV)-mediated over-expression of C-terminal haemagglutinin (HA)-tagged constructs in mammalian cells. [Plasmid contains NHK(human *SERPINA* α-1) flanked by EcoRI/HindIII sites]	Nicholay Kukushkin (University of Oxford)
p3×FLAG−CMV-10	-	CMV-mediated over-expression of N-terminal, triple FLAG–tagged constructs in mammalian cells.	Sigma
pcDNA3−*Setd2*(13»21)−HA	pcDNA3−NHK−HA	For expression of C-terminal HA-tagged SETD2(13»21); *Setd2*(13»21) is cloned between EcoRI and HindIII sites of vector.	This study
p3×FLAG−*Setd2*(13»21)−CMV-10	p3×FLAG−CMV-10	For expression of N-terminal 3×FLAG-tagged SETD2(13»21); *Setd2*(13»21) is cloned between EcoRI and HindIII sites of vector.	This study

Table 2.3 – PCR primers for generation of epitope-tagged constructs

Construct	Primer	Direction	Sequence (5' → 3')	Tm (°C)*
pcDNA3−*Setd2*(13»21)−HA	HinKoz13F1	Forward	CCAAGCTTgccaccatggAGGTGTATCGGATTCCAAAGAA	68
	21EcoR1	Reverse	ATGAATTCCTCTAGTTCAGTGTCCTCTTTAGG	70
p3×FLAG−*Setd2*(13»21)−CMV-10	Hin13F1	Forward	CCAAGCTTGAGGTGTATCGGATTCCAAAGA	64
	21TGAEcoR1	Reverse	CTGAATTCAGGTCACTCTAGTTCAGTGTC	62

HindIII and EcoRI restriction sites are underlined; Kozak consensus sequence is in lower-case. * Tm calculated as 2 × AT + 4 × CG

2.3.3. Sub-cloning of *Setd2*(13»21) into epitope-tagging vectors

Setd2(13»21) fragments were excised from pJET1.2–*Setd2*(13»21) clones by EcoRI-HF/HindIII double-digestion and purified by gel extraction (section 2.17.2). Recipient vectors, pcDNA3−NHK−HA and p3×FLAG−CMV-10 (**Table 2.2**) were digested with EcoRI-HF and HindIII, dephosphorylated with Antarctic phosphatase (NEB) and digested vectors purified by gel extraction (section 2.17.2). Purified *Setd2*(13»21)−FLAG and *Setd2*(13»21)-HA fragments were ligated into processed p3×FLAG−CMV-10 and pcDNA3−HA vectors, respectively, at 16°C overnight using DNA Ligation Mighty Mix (Takara) according to the manufacturer's instructions. NEB-10β *E. coli* were transformed with 10 μl ligation reaction and selected as described in section 2.3.2. Colonies with correct inserts were screened by PCR and restriction digestion with EcoRI-HF and HindIII. Epitope-tagged constructs were verified by multiple overlapping rounds of DNA sequencing (section 2.17.3).

2.4. Cell culture and treatment

Murine fibroblasts (C3H 10T½), human lung fibroblasts (MRC5), and human embryonic kidney–derived cells (HEK 293T) were cultured in Dulbecco's modified Eagle's medium (DMEM; Life Technologies) containing 10% (*v*/*v*) foetal calf serum (FCS) and 2 mM glutamine (Life Technologies) at 37°C with 6% CO_2. HEK293T cells were cultured in the presence of 50 U of penicillin G and 50 μg streptomycin sulphate (Life Technologies). Cultures were quiesced at 80–95% confluence in DMEM containing 0.5% (*v*/*v*) FCS and 2 mM glutamine for 18–24 hours. Cell cultures were confirmed free of mycoplasma by EZ-PCR mycoplasma test kit (Geneflow).

Murine embryonic stem cells (*CGBP*$^{-/-}$ ESCs) were kindly provided by Dr Catherine Hazzalin. ESCs were harvested in PBS before lysis in TX−DOC buffer (section 2.7.2).

Drosophila S2 cells were cultured at 23°C at ambient CO_2 in Schneider's medium (Lonza) containing 10% (*v*/*v*) heat-treated FCS (kindly provided by Caroline Schneider).

Where indicated, compounds were added directly to cell culture media from 1000× stock solutions. Emetine (Sigma) was used at a final concentration of 10 μg/ml; recombinant human epidermal growth factor (EGF, Promega) was used at a final concentration of 50 ng/ml.

2.5. Transient transfection of 10T½ cells with small-interfering RNA (siRNA)

Cells were transiently transfected with 50 nM siRNA (**Table 2.4**) by Lipofectamine™ 2000 or RNAiMAX (Life Technologies) based on the procedure described in (Edmunds *et al.*, 2008). siRNA stock solutions were dissolved at 10 μM in ***siRNA buffer*** (6 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid)–KOH pH 7.5 and 0.2 mM $MgCl_2$). 10T½ cells at 40–50% confluence were washed once with, and transferred to, OptiMAX® medium (Life Technologies) pre-warmed to 37°C. Equal volumes of buffer A (500 nM siRNA [or equal volume of siRNA buffer for mock transfections] in OptiMAX®) and buffer B (1 μl Lipofectamine™ 2000 or RNAiMAX in 50 μl OptiMAX®) were incubated separately at room temperature for 10 minutes, before combining with gentle mixing and incubation for a further 20 minutes. AB mix (0.2 volumes) was added drop-wise to cells in OptiMAX® with continuous mixing. Cells were transfected for 5 hours at 37°C before adding an equal volume of DMEM containing 20% (*v*/*v*) FCS and 2 mM glutamine. Cells were quiesced 24 hours after transfection and harvested after a further 24 hours by TX–DOC lysis (section 2.7.2).

2.6. Transient transfection of 10T½ cells by electroporation

Exponentially growing, sub-confluent 10T½ cells cultured in 175 cm^2 flasks were harvested by trypsinization and sedimented by centrifugation at 300*g* for 5 minutes. Cells were resuspended at 6×10^7 cells/ml in ***NCS-medium*** (DMEM containing 10% (*v*/*v*) new-born calf serum (NCS)) and 250 μl were added to electroporation cuvettes containing 3 μg purified plasmid (section 2.17.1) and 500 μg sonicated herring sperm DNA. Cells were immediately electroporated (230 V, 1075 μF for pulse times of 40–45 ms), then supplemented with 750 μl

Table 2.4 – Small-interfering RNAs (siRNAs)

siRNA	*Setd2 e*xon*	Sense sequence (5' → 3')	Source	Reference
Setd2(52)	3	CAUGUUACUUUGAUCUUAU-dTdT	Sigma	This study
Setd2(53)	3	GUGUUAUCUGGGCAGUCUU-dTdT	Sigma	This study
Setd2(54)	3	GGAAAUAAGCAGCAAUUCA-dTdT	Sigma	This study
Setd2(0)	13	GUAUCGGAUUCCAAAGAAA-dTdT	Dharmacon	(Edmunds *et al.*, 2008)
Setd2(3)	21	GAGCUGAAGUACUGUAAGA-UU	Dharmacon	(Edmunds *et al.*, 2008)
Non-targeting	-	UAGCGACUAAACACAUCAA-UU	Dharmacon	(Edmunds *et al.*, 2008)

* exon numbers correspond to murine *Setd2* (NCBI RefSeq: NM_001081340.2).

NCS-medium and incubated at room temperature for 5 minutes. After centrifuging through 10 ml NCS-medium at 300*g* for 5 minutes, cells were resuspended in fresh DMEM containing 10% (*v*/*v*) FCS and 2 mM glutamine and plated into three 90 mm dishes. Cells were quiesced for 24 hours before harvesting (48 hours after transfection) by nuclease–salt extraction (section 2.7.3) or differential detergent extraction (section 2.7.4).

2.7. Preparation of cell lysates and subcellular fractions

2.7.1. Phosphatase and protease inhibitors

Phosphatase inhibitors consisted of: 20 mM sodium β-glycerophosphate, 1 μM microcystin-LR (Enzo Life Sciences) and 100 μM sodium orthovanadate.

Protease inhibitors were added immediately before use at the following final concentrations: 2 μg/ml aprotinin, 5 μg/ml leupeptin, 2 mM phenylmethanesulphonyl fluoride, 50 μg/ml soybean trypsin inhibitor, 100 nM tosyl lysyl chloromethyl ketone, and 100 nM tosyl phenylalanyl chloromethyl ketone.

2.7.2. TX–DOC lysis and extraction of TX-DOC-resistant material

Preparation of Triton X-100–deoxycholate (TX–DOC) extracts was adapted from (Mahadevan *et al.*, 1988). Culture medium was aspirated from quiesced, confluent cells cultured in 90 mm (or 145 mm) dishes and then cells were lysed directly by scraping in 200 μl (or 500 μl) ice-cold ***TX–DOC buffer*** (20 mM HEPES–KOH pH 8.0, 100 mM NaCl, 5 mM EDTA (ethylenediaminetetraacetic acid), 1% (*v*/*v*) Triton X-100, 1% (*w*/*v*) sodium deoxycholate, 10 mM sodium butyrate, 5 mM NaF, phosphatase and protease inhibitors). Lysates were incubated on ice for 10 minutes before centrifugation at 16,000*g* for 15 minutes at 4°C. The TX–DOC supernatant was transferred to a new tube and either processed directly for electrophoresis (section 2.8) or precipitated with trichloroacetic acid (TCA) (section 2.7.6) as indicated.

The TX-DOC-resistant pellet was resuspended in 200 μl (or 500 μl) ***urea-solubilisation buffer*** (9 M urea, 100 mM dithiothreitol (DTT), 0.1% (*v*/*v*) NP-40) containing protease

inhibitors, and rotated end-over-end for 30 minutes at room temperature. Centrifugation at 16,000*g* for 15 minutes at 4°C yielded a urea-soluble supernatant and urea-insoluble pellet, which were immediately processed for electrophoresis (section 2.8) or stored at −20°C.

2.7.3. Subcellular fractionation by nuclease digestion and high-salt extraction

This method was modified from a procedure to generate a subcellular fraction enriched for components of the nuclear substructure from C3H 10T½ cells (Rampersaud, 2003), which is based on nuclease digestion and high-salt fractionation by (Dwyer & Blobel, 1976). The method is outlined in **Figure 4.1** in Chapter 4.

After medium aspiration, cells cultured in 90 mm (or 145 mm) dishes were scraped directly in 200 µl (or 500 µl) of ice-cold ***hypotonic lysis buffer*** (0.1% (*v*/*v*) Triton X-100, 10 mM HEPES–KOH pH 8.0, 250 mM sucrose, 1.5 mM $MgCl_2$, 10 mM KCl, 10 mM sodium butyrate, phosphatase and protease inhibitors). After incubation on ice for 10 minutes, the sample was centrifuged at 800*g* for 5 minutes at 4°C to yield a hypotonic supernatant (Hypo-SN). The hypotonic pellet was resuspended in 292 µl (or 584 µl) of ice-cold ***extraction buffer*** (10 mM Tris–HCl pH 7.5, 300 mM sucrose, 0.1 mM $MgCl_2$, 1 mM DTT, 10 mM sodium butyrate, phosphatase and protease inhibitors) supplemented with 100 µl 0.1 mM $MgCl_2$, 6.5 µl deoxyribonuclease I (DNase I at 10 U/µl; Roche) and 2 µl ribonuclease A (RNase A at 10 mg/ml; Sigma). Samples were digested with end-over-end rotation for 30 minutes at room temperature. Samples were centrifuged at 16,000*g* for 10 minutes at 4°C to yield an initial nuclease-digest supernatant (NAse). The pellet was washed by resuspending in 100 µl (or 200 µl) of ice-cold extraction buffer, re-centrifuged as above and the supernatant pooled with the preceding NAse fraction. The nuclease-resistant pellet was resuspended in 200 µl (or 400 µl) of ice-cold extraction buffer and then an equal volume of ice-cold ***2 M NaCl buffer*** (2 M NaCl, 100 mM Tris–HCl pH 7.5, 300 mM sucrose, 1 mM DTT and protease inhibitors) was slowly added drop-wise with mixing. The sample was extracted by end-over-end rotation for 30 minutes at 4°C and subsequently centrifuged at 16,000*g* for 10 minutes at 4°C to yield an initial high-salt supernatant. The salt-resistant pellet was washed by resuspension in 100 µl (or 200 µl) of ice-cold extraction buffer, re-centrifuged as above and the supernatant pooled with the preceding high-salt supernatant. Soluble

supernatant fractions were precipitated with TCA (as described in section 2.7.6). Samples were immediately processed for electrophoresis (section 2.8) or stored at −20°C.

2.7.4. Differential detergent fractionation to produce a p66$^{C/CT21}$-enriched fraction

A subcellular fractionation procedure employing differential detergent extraction and centrifugation was developed to obtain a p66$^{C/CT21}$-enriched extract. The overall method is outlined in **Figure 4.5** in Chapter 4.

Quiesced, confluent C3H 10T½ cells in 90 mm (or 145 mm) dishes were lysed, after medium aspiration, by scraping directly in 200 µl (or 500 µl) of ice-cold ***hypotonic lysis buffer*** (10 mM HEPES–KOH pH 8.0, 0.1% (*v*/*v*) Triton X-100, 10 mM KCl, 1.5 mM $MgCl_2$, 250 mM sucrose, 10 mM sodium butyrate, phosphatase and protease inhibitors). Lysates were incubated on ice for 10 minutes before centrifugation at 800*g* for 5 minutes at 4°C to yield a hypotonic supernatant (Hypo-SN). The hypotonic pellet (Hypo-P) was thoroughly resuspended by flicking in 200 µl (or 400 µl) of ice-cold ***Triton-extraction buffer*** (20 mM HEPES–KOH pH 8.0, 1% (*v*/*v*) Triton X-100, 100 mM NaCl, 5 mM EDTA, 10 mM sodium butyrate, 5 mM NaF, phosphatase and protease inhibitors). Pellets were extracted on a rotating wheel for 30 minutes at 4°C before centrifugation at 16,000*g* for 10 minutes at 4°C. The initial Triton-supernatant (TX-SN) was transferred to a new tube and the pellet resuspended briefly in 100 µl (or 200 µl) Triton-extraction buffer before re-centrifugation at 16,000*g* for 10 minutes at 4°C. The supernatant was pooled with the TX-SN fraction, which was then centrifuged at 100,000*g* for 60 minutes at 4°C (Beckman TLA100.3 rotor in a Beckman Optima-TLX ultracentrifuge) to yield the TX-S^{100} supernatant and the TX-P^{100} pellet. The latter was thoroughly resuspended in 200 µl (or 400 µl) of ice-cold TX–DOC buffer (described in section 2.7.2) by vigorous flicking and rotation at 4°C until completely solubilised (30–90 minutes). The resuspension was subjected to identical ultra-centrifugation conditions as above to yield the TD-S^{100} supernatant and TD-P^{100} pellet. All soluble supernatants were precipitated with TCA (section 2.7.6). Samples were either stored at −20°C or solubilised for immediate electrophoresis (section 2.8).

During the course of DDC development (**Figure 4.4**), hypotonic pellets from 145 mm dishes were extracted with 500 µl Triton-extraction buffer or TX–DOC buffer by end-over-rotation for 30 minutes at 4°C. These extracts were centrifuged at 16,000*g* for 10 minutes at 4°C to yield TX or TD supernatants, respectively, and detergent-resistant pellets. The soluble fractions were ultra-centrifuged as detailed above, and where indicated, TX–P^{100} pellets were resolubilised in TX–DOC buffer, or Triton-extraction buffer supplemented with either 1% (*v*/*v*) Tween-20 or NaCl at 1 M final concentration, then ultra-centrifuged at either 43,000*g* or 100,000*g* as above. Supernatants were precipitated with TCA (section 2.7.6). Samples were either stored at −20°C or solubilised for immediate electrophoresis (section 2.8).

2.7.5. Preparation of acid extracts

Detergent-resistant pellets after TX–DOC lysis (section 2.7.2) or Triton-extraction (section 2.7.4) were extracted with 75 µl of ice-cold 0.4 M HCl for 60 minutes on ice. Samples were centrifuged at 16,000*g* for 5 minutes at 4°C and supernatants transferred to new tubes. Acid extracts were precipitated overnight (16–24 hours) at −20°C with 10 volumes of acetone (stored at −20°C). Protein was recovered by centrifugation at 16,000*g* for 15 minutes at 4°C and pellets were washed three times with 1 ml acetone (stored at −20°C) before briefly air-drying at room temperature.

2.7.6. Trichloroacetic acid (TCA) precipitation

Samples were precipitated by addition of an equal volume of 50% (*w*/*v*) TCA (stored at 4°C) and incubated on ice for 60 minutes with occasional agitation by vortex. Precipitated material was sedimented by centrifugation at 16,000*g* for 15 minutes at 4°C and subsequently washed three times with 1 ml acetone (stored at −20°C), ensuring rigorous pellet disruption at each step. Acetone was removed by evaporation at room temperature for several minutes. Samples were either stored at −20°C or solubilised for immediate electrophoresis as described below.

2.8. Protein electrophoresis

2.8.1. Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE)

Three solubilisation methods were employed for SDS-PAGE, depending on the nature of the sample.

(i) Soluble protein extracts were adjusted to 1× ***SDS-PAGE loading buffer*** by addition of 4× stock (40% (*v*/*v*) glycerol, 240 mM Tris–HCl pH 6.8, 8% (*w*/*v*) SDS, 0.04% (*w*/*v*) bromophenol blue, and 4% (*v*/*v*) β-mercaptoethanol). Samples were immediately heated at 95°C for 3 minutes. Where indicated, samples were adjusted to 0–4 M urea and 1× SDS-PAGE loading buffer before thermal denaturation under specified conditions.

(ii) TCA-precipitated samples and subcellular fractions in pellet form that derived from 90 mm (or 145 mm) dishes were solubilised in ***urea/SDS-PAGE loading buffer*** by addition of 100 µl (or 200 µl) of urea-solubilisation buffer (9 M urea, 100 mM DTT, 0.1% (*v*/*v*) NP-40) and rotation for 30 minutes at room temperature, before adding an equal volume of 2× SDS-PAGE loading buffer.

(iii) Acetone-precipitated acid extracts were solubilised in 60 µl 8 M urea for 15 minutes, before dilution to 4 M urea with Milli-Q water, and addition of an equal volume of 2× SDS-PAGE loading buffer.

To avoid carbamylation, samples containing urea were *not* heated. Immunodetection of $p66^{C}$ with α-SETD2 C-terminal antibody, however, *requires* heating at 95°C in the presence of urea; unless specified otherwise, this was achieved by heating aliquots of samples at 95°C for 10 or 15 minutes.

SDS-PAGE was performed according to standard methods (Laemmli, 1970) with the following modifications. Mini gels [200 mm × 80 mm] or large gels [250 mm × 175 mm] were prepared using Protogel™ solutions (National Diagnostics). Resolving gels contained 375 mM Tris–HCl pH 8.8 and 0.1% (*w*/*v*) SDS with: (i) 8% (*w*/*v*) acrylamide and 0.21% (*w*/*v*)

bis-acrylamide for 8% gels; (ii) 10% (*w*/*v*) acrylamide and 0.27% (*w*/*v*) *bis*-acrylamide for 10% gels; or (iii) 15% (*w*/*v*) acrylamide and 0.4% (*w*/*v*) *bis*-acrylamide for 15% gels. Stacking gels contained 125 mM Tris–HCl pH 6.8, 0.1% (*w*/*v*) SDS, 5% (*w*/*v*) acrylamide and 0.13% (*w*/*v*) *bis*-acrylamide. Resolving gels were polymerised by addition of 0.1% (*v*/*v*) *N*, *N*, *N*', *N*'-tetramethyl-ethane-1,2-diamine (TEMED) and 0.1% (*w*/*v*) ammonium persulphate (APS), whilst 0.3% (*v*/*v*) TEMED and 0.05% (*w*/*v*) APS were added for stacking gels. Each well was loaded with 1/10th of an extract derived from a 90 mm dish (1/20th from a 145 mm dish), which was typically a volume of 20–30 µl. Gels were run in ***SDS-PAGE running buffer*** (25 mM Tris pH 8.3, 192 mM glycine, 0.11% (*w*/*v*) SDS) at a constant voltage of 150 V until bromophenol blue dye had migrated off the gel (typically, 100–110 minutes for SETD2 analysis on 8% gels). Gels were then either Coomassie-stained (section 2.9) or transferred to polyvinylidene fluoride (PVDF) membrane for immunoblotting (sections 2.10 and 2.11).

2.8.2. Acetic acid–urea polyacrylamide gel electrophoresis (AU-PAGE)

TCA-precipitated samples and subcellular fractions in pellet form were solubilised with 4 M urea, 5% (*v*/*v*) β-mercaptoethanol and 5% (*v*/*v*) acetic acid. For one-dimensional AU-PAGE, samples were solublised in 130 µl 4 M urea and 7.5 µl β-mercaptoethanol at room temperature for 15 minutes with regular mixing by vortex. Samples were supplemented with 7.5 µl glacial acetic acid and 5 µl methyl green dye, and solubilised at room temperature for a further 15 minutes with regular mixing by vortex. To facilitate loading of the entire TD-S^{100} extract from a 145 mm dish for 2D–AU/SDS–PAGE, samples were resolubilised with 87 µl 4 M urea, 5 µl β-mercaptoethanol, 5 µl glacial acetic acid and 3.3 µl methyl green.

AU-polyacrylamide gels were made 250 mm × 175 mm and 0.75 mm thick. Resolving gels contained 4 M urea, 5% (*v*/*v*) acetic acid, 10% (*w*/*v*) acrylamide and 0.13% (*w*/*v*) *bis*-acrylamide. For histones, 15% (*w*/*v*) acrylamide and 0.2% (*w*/*v*) *bis*-acrylamide were required. Resolving gels were polymerised by addition of 0.2% (*w*/*v*) APS and 0.4% (*v*/*v*) TEMED. Stacking gels (20 mm deep) contained 4% (*w*/*v*) acrylamide, 0.052% (*w*/*v*)

bis-acrylamide, 8 M urea and 5% (*v*/*v*) acetic acid; they were polymerised by addition of 0.4% (*w*/*v*) APS and 0.8% (*v*/*v*) TEMED.

Wells were loaded with 20 µl 4 M urea/10% (*w*/*v*) polyethylene glycol (PEG) and pre-electrophoresis was performed in 5% (*v*/*v*) acetic acid buffer, overnight at a constant voltage of 150 V until current stabilised. Running buffer was then replaced anew and urea/PEG removed from wells.

For one-dimensional AU-PAGE, 1/10th extract from a 145 mm dish was loaded (typically 15 µl); for two-dimensional AU/SDS-PAGE, the entire TD-P^{100} fraction (100 µl) from a single 145 mm dish was analysed. Electrophoresis was performed at a constant current of 15 mA (per gel): for 7 hours and 45 minutes for α-SETD2 analysis on 10% gels; for 5 hours and 20 minutes for histones on 15% gels. Gels were subsequently stained with Coomassie (section 2.9), transferred to PVDF (section 2.10.2) or processed for two-dimensional electrophoresis (section 2.8.3).

2.8.3. Two-dimensional AU/SDS–PAGE

For two-dimensional analysis, first-dimension AU-PAGE was performed as described in section 2.8.2, ensuring critically: (i) gel thickness of 0.75 mm; (ii) wide-spaced wells; and (iii) blank lanes between each sample loaded. The second-dimension 10% SDS-gel (250 mm × 175 mm) was prepared as in section 2.8.1 but needed to be 1 mm thick to accommodate swollen first-dimension gel strips. Gel strips were loaded into 70 mm–wide lanes generated with custom-made combs that enable two such lanes to be loaded in parallel, along with a single well necessary for loading molecular weight markers.

First-dimension AU-PAGE was performed exactly as in section 2.8.2 with the entire p66$^{C/CT21}$-enriched TD–P^{100} extract from a single 145 mm dish loaded in 100 µl. After electrophoresis, gel strip locations were indicated by gridlines drawn on both gel plates, before disassembly. Avoiding movement of the gel, one plate was removed and appropriate gel strips were excised and their orientation marked using a clean block blade.

Reproducible localisation of $p66^{C/CT21}$ was experimentally confirmed, allowing the top 20 mm of gel to be discarded. The topmost 60 mm segment was processed further: strips were incubated three times in 100 ml 125 mM Tris–HCl pH 6.8 for 5 minutes for each incubation, on a rotating platform to neutralise AU-PAGE buffer. Strips were then incubated three times in 100 ml 2D-equilibration buffer (100 mM Tris–HCl pH 6.8, 10% (*v*/*v*) glycerol, 0.1% (*w*/*v*) SDS, 0.01% (*w*/*v*) bromophenol blue) for 5 minutes on a rotating platform. After removing unpolymerised acrylamide from the wells of the SDS gel by flushing with 1× SDS-PAGE running buffer, the majority of liquid was removed with blotting paper (leaving a residual 1–2 mm layer of liquid to help avoidance of air bubbles when loading gel strips). Equilibrated first-dimension gel segments were carefully loaded, avoiding distortion of dimensions, and air bubbles at the interface. Electrophoresis was performed at a constant voltage of 70 V until the dye front had migrated into the resolving gel. Electrophoresis was continued at 70–90 V overnight (14–18 hours) until the 28 kDa marker reached the bottom. Gels were stained with Coomassie (section 2.9) or transferred in Towbin buffer for immunoblotting (section 2.10.1).

2.9. Coomassie Brilliant Blue staining

Polyacrylamide gels were stained with Coomassie Brilliant Blue stain (0.1% (*w*/*v*) Coomassie Brilliant Blue R-250; 25% (*v*/*v*) isopropanol; 10% (*v*/*v*) acetic acid) for 4 hours at room temperature. Gels were destained with several changes of ***Coomassie destaining solution*** (25% (*v*/*v*) methanol and 7.5% (*v*/*v*) acetic acid).

2.10. Western transfer

Common to all procedures described below, gels were transferred to polyvinylidene fluoride (PVDF) membrane (Immobilon P™, Millipore), which was pre-wetted with methanol and equilibrated in the appropriate transfer buffer for at least 10 minutes before assembly of transfer apparatus.

2.10.1. Western transfer after SDS-PAGE

SDS-polyacrylamide gels were equilibrated with gentle rocking for 30 minutes at room temperature in either ***SDS transfer buffer*** (12.5 mM Tris pH 8.3, 96 mM glycine, 0.05% (*w*/*v*) SDS, 10% (*v*/*v*) methanol) or ***Towbin transfer buffer*** (25 mM Tris pH 8.3, 192 mM glycine, 20% (*v*/*v*) methanol) (Towbin *et al.*, 1979), as required for specific antibodies (**Table 2.5**). Gels were transferred to PVDF in SDS or Towbin transfer buffers at 200 mA constant current, overnight (16–24 hours) at 4°C with buffer agitation.

2.10.2. Western transfer after AU-PAGE

AU-polyacrylamide gels were equilibrated in 5% (*v*/*v*) acetic acid/10% (*v*/*v*) methanol for 30 minutes followed by 10 minutes in ***AU transfer buffer*** (0.1% (*v*/*v*) acetic acid and 10% (*v*/*v*) methanol). Gels were transferred to PVDF in AU transfer buffer at 200 mA constant current, overnight (16–24 hours) at 4°C with buffer agitation.

2.10.3. Western transfer after Coomassie staining

After staining with Coomassie Brilliant Blue (section 2.9), gels were thoroughly destained with several changes of Coomassie destaining solution at 50°C for 4 hours, before leaving overnight in destaining solution at room temperature with gentle agitation. Gels were equilibrated in SDS-PAGE running buffer for 60 minutes at room temperature, followed by two washes in Towbin transfer buffer for 15 minutes. Gels were then transferred to PVDF in Towbin buffer as described in section 2.10.1.

2.10.4. Ponceau S staining

Even transfer was assessed by staining PVDF membranes with Ponceau S stain (0.4% (*w*/*v*) Ponceau S; 5% (*v*/*v*) acetic acid) for several minutes and then destaining with several changes of methanol. Membranes were air-dried at room temperature for 30 minutes before imaging and immunoblotting.

2.11. Immunoblotting

2.11.1. Peptide–dot blotting

Nitrocellulose membrane (Hybond™-C Extra, GE Healthcare) was spotted with 1 μl of SETD2 peptide (dissolved at 1 pmol/μl in PBS) and left to dry for 20 minutes. Membranes were incubated in ***blocking solution*** (5% (*w*/*v*) skimmed milk powder, PBS, 0.1% (*v*/*v*) Tween-20) for one hour at room temperature, and washed twice in PBS/0.1% (*v*/*v*) Tween-20 for 5 minutes, before immunoblotting as described in section 2.11.2.

2.11.2. Immunoblotting

Air-dried membranes were briefly rinsed in 1× PBS before incubation with primary antibody diluted in blocking solution (as specified in **Table 2.5**) for one hour at room temperature. Membranes were washed three times with PBS/0.1% (*v*/*v*) Tween-20 for 10 minutes, except for α-Lamin B immunoblots, which were washed thrice with PBS/0.1% (*v*/*v*) Triton X-100 for 15 minutes. Membranes were subsequently incubated for one hour at room temperature with appropriate horseradish peroxidase (HRP)–conjugated secondary antibodies diluted in blocking solution (as specified in **Table 2.5**). Membranes were washed three times with PBS/0.1% (*v*/*v*) Tween-20 for 10 minutes except for α-Lamin B immunoblots, which were washed as previously stated. Secondary antibody was detected by enhanced chemiluminescence (section 2.11.3) and exposure to X-ray film (Kodak).

For antibody characterisation, membranes were divided into 4 mm–wide strips and probed in eight-lane trays as above, except five washes in PBS/0.1% (*v*/*v*) Tween-20 were performed for five minutes each.

For reprobing with antibodies from different species, membranes were stripped with 25 mM glycine–HCl pH 2.0/1% (*w*/*v*) SDS for 20 minutes at room temperature. Membranes were then washed twice in PBS/0.1% (*v*/*v*) Tween-20 for 10 minutes and incubated overnight with blocking solution at 4°C. Membranes were briefly washed with PBS/0.1% (*v*/*v*) Tween-20 before primary antibody incubation.

Table 2.5 – Antibodies used in immunoblotting

Antibody	Species	Type	Source	Dilution (Concentration)	Transfer condition	ECL system
Primary antibodies						
α-FLAG (M2)	Mouse	Purified monoclonal IgG_1	Sigma (F3165)	1:5000	Towbin	Coumaric acid
α-HA	Rabbit	Purified polyclonal IgG	Santa Cruz (sc-805)	1:500 (0.2 µg/ml)	Towbin	Millipore
α-H3K27me3	Rabbit	Protein A–purified polyclonal IgG	Millipore (07-449)	1:10,000	SDS	Millipore
α-H3K36me3	Rabbit	Peptide affinity–purified polyclonal IgG	Abcam (ab9050)	1:5000 (0.2 µg/ml)	SDS	Millipore
α-JUN	Rabbit	Purified polyclonal IgG	Santa Cruz (sc-1694)	1:1000 (0.2 µg/ml)	Towbin	Coumaric acid
α-Lamin B	Goat	Purified polyclonal IgG	Santa Cruz (sc-6217)	1:1000 (0.2 µg/ml)	SDS	Millipore
α-Nucleophosmin	Goat	Purified polyclonal IgG	Santa Cruz (sc-6013)	1:2000 (0.4 µg/ml)	Towbin	Millipore
α-RNAPII (RPB1)	Rabbit	Purified polyclonal IgG	Santa Cruz (sc-899)	1:1000 (0.2 µg/ml)	SDS	Millipore
α-SETD2 N-terminal	Rabbit	Peptide affinity–purified polyclonals	This study (NSL R2039)	1:120 (2 µg/ml)	SDS/Towbin	4IPBA
α-SETD2 C-terminal	Sheep	Peptide affinity–purified polyclonals	This study (NSL S950B)	1:120 (2 µg/ml)	SDS/Towbin	Coumaric acid
α-SETD2 CT21	Goat	Peptide affinity–purified polyclonals	Everest Biotech (eb07418)	1:1000 (0.5 µg/ml)	SDS/Towbin	Millipore
Secondary antibodies						
α-rabbit IgG	Donkey	Affinity-purified polyclonal HRP-conjugated IgG	GE Healthcare (NA934)	1:5000*		
α-sheep/goat IgG	Donkey	Purified polyclonal HRP-conjugated IgG	AbD Serotec (STAR88P)	1:5000		
α-mouse IgG	Sheep	Affinity-purified polyclonal HRP-conjugated IgG	GE Healthcare (NA931)	1:5000		

*1:2500 for α-JUN immunoblot.

2.11.3. Enhanced chemiluminescence

Three different enhanced chemiluminescence (ECL) systems were used for detection of different primary antibodies (specified in **Table 2.5**):

(i) Coumaric acid–based ECL (100 mM Tris–HCl pH 8.0, 1.25 mM 3-aminophthalhydrazide (luminol), 200 μM *p*-coumaric acid, 0.01% (*v*/*v*) H_2O_2). Membranes were incubated for one minute.

(ii) 4-iodophenylboronic acid (4IPBA)–based ECL (100 mM Tris–HCl pH 8.8, 1.25 mM luminol, 2 mM 4IPBA, 0.02% (*v*/*v*) H_2O_2) derived from (Haan & Behrmann, 2007). Membranes were incubated for one minute.

(iii) High-sensitivity, commercial ECL: Immobilon™ ECL (Millipore). Membranes were incubated for five minutes.

2.11.4. Peptide competition

Primary antibody solutions were supplemented with SETD2 peptides (100 nM or 1 μM final concentration) and rotated for 30 minutes at room temperature, before immunoblotting membranes as in section 2.11.2.

2.11.5. Quantification and mass estimation of immunoreactive proteins

ImageJ 1.44o was used to quantify immunoblot signals (Abramoff *et al.*, 2004). Signals were averaged from at least two exposures per experiment and normalised to Ponceau S stain. RNAi data were expressed as mean % non-target control signal ($n \geq 3$). Protein stability data were expressed as mean % untreated signal ($n \geq 3$). Protein decay after inhibition of protein synthesis was assumed to decay exponentially with first-order kinetics: ($P_t = P_0 e^{-kt}$) where P_t is protein signal at time t, and P_0 is untreated signal (Schwanhausser *et al.*, 2011). The rate-constant (k) was estimated by linear regression ($\ln P_t = \ln P_0 - kt$) and half-life ($t_{½}$) calculated as ($t_{½} = \ln 2 / k$).

Molecular mass of immunoreactive proteins was estimated by interpolating *log*(molecular mass) versus relative mobility of protein standards by SDS-PAGE (Weber & Osborn, 1969). Mean values were calculated from at least five independent biological experiments.

2.12. In-gel digestion and mass spectrometry

All procedures were performed at room temperature unless stated otherwise. Polyacrylamide gels were prepared and stained with filtered solutions (0.22 μm filter, mixed cellulose esters; MILLEX®-GS, Millipore) to minimise keratin contamination.

2.12.1. In-gel trypsin digestion and C18 purification

Coomassie-stained spots from 2D−AU/SDS gels were excised with a clean scalpel and cut into ≈1 mm pieces. Samples were washed with 100 μl ***solution B*** (25 mM NH_4HCO_3, 50% (*v*/*v*) acetonitrile) for 30 minutes to remove salts; supernatants were discarded and the wash procedure repeated until samples were completely destained. Samples were dehydrated in 100 μl acetonitrile for 10 minutes; supernatants were removed and samples dried in a centrifugal evaporator for 10 minutes. To reduce disulphides, 100 μl DTT solution (10 mM DTT in 25 mM NH_4HCO_3) was added and samples incubated at 56°C for 30 minutes. After discarding supernatants, samples were washed twice with 100 μl of 25 mM NH_4HCO_3 and dehydrated with acetonitrile as above. Samples were incubated with 100 μl alkylation solution (55 mM iodoacetamide, 25 mM NH_4HCO_3) for 60 minutes at room temperature in the dark. After discarding supernatants, samples were washed twice with 100 μl solution B for 10 minutes, dehydrated with acetonitrile, and dried in a centrifugal evaporator for 10 minutes. Samples were digested for 30 minutes at 37°C by addition of 20 μl trypsin (Promega, cat. No V5111) which had been freshly reconstituted at 10 μg/ml in 25 mM NH_4HCO_3. Sufficient 25 mM NH_4HCO_3 was added to ensure gel pieces were covered with 2–3 mm of liquid, before incubating overnight at 37°C. After addition of 1 μl formic acid, supernatants containing peptides were transferred to new tubes. Gel fragments were incubated with 50 μl of 0.1% (*v*/*v*) formic acid at room temperature for 30 minutes, after which supernatants were pooled with peptides. Gel pieces were finally incubated with 50 μl of 0.1% (*v*/*v*) formic acid in 50% (*v*/*v*) acetonitrile for 30 minutes at

room temperature. Supernatants were combined with pooled peptides, and peptide samples were dried almost completely in a centrifugal evaporator.

Tryptic peptides were desalted using C18 columns coupled to a vacuum manifold. C18 columns were prepared by packing a P10 pipette tip with a small disk punched from Empore™ Octadecyl C18 material (Sigma®, 47 mm diameter). C18 columns were pre-wetted with 50 µl of 0.1% (*v*/*v*) formic acid/acetonitrile before equilibration with 50 µl of 0.1% (*v*/*v*) formic acid/Milli-Q. Peptides dissolved in 10 µl of 0.1% (*v*/*v*) formic acid were slowly loaded onto columns and the flow-through fractions retained. After washing with 50 µl of 0.1% (*v*/*v*) formic acid, peptides were eluted into single tubes by, firstly, adding 20 µl of 0.1% (*v*/*v*) formic acid/50% (*v*/*v*) acetonitrile, and secondly, adding 20 µl of 0.1% (*v*/*v*) formic acid/70% (*v*/*v*) acetonitrile. Peptides were dried in a centrifugal evaporator and stored at −20°C pending mass spectrometry.

2.12.2. In-gel digestion with chymotrypsin

Excised gel segments were incubated overnight with 200 µl wash solution (50% (*v*/*v*) methanol, 5% (*v*/*v*) acetic acid) at room temperature. After replacement with a further 200 µl wash solution and incubation for 2½ hours, gel fragments were dehydrated twice with 200 µl acetonitrile for 5 minutes. Samples were dried in a centrifugal evaporator for 3 minutes before incubation with 30 µl DTT solution (10 mM DTT, 100 mM NH_4HCO_3) for 30 minutes. Supernatants were replaced with 30 µl alkylating solution (100 mM iodoacetamide, 100 mM NH_4HCO_3). After 30 minutes, alkylating solution was removed and samples were dehydrated with 200 µl acetronitrile for 5 minutes and dried for 3 minutes in a centrifugal evaporator. After rehydration in 200 µl of 100 mM NH_4HCO_3 for 10 minutes, samples were dehydrated as above. Samples were digested with 30 µl of 20 ng/µl sequencing-grade chymotrypsin (Roche, cat no. 1418467) dissolved in 50 mM NH_4HCO_3. After incubation on ice for 10 minutes with occasional mixing by vortex, excess chymotrypsin solution was removed; 50 µl of 50 mM NH_4HCO_3 was subsequently added and digestion performed overnight at 37°C. Samples were incubated for 10 minutes with an additional 30 µl of 50 mM NH_4HCO_3 and the supernatant containing chymotryptic peptides transferred to a new tube. Gel fragments were further extracted twice by

incubation in 50 µl extraction buffer (5% (*v*/*v*) formic acid, 50% (*v*/*v*) acetonitrile) for 10 minutes and peptides pooled. After drying in a centrifugal evaporator, peptides were resolubilised in 20 µl of 1% (*v*/*v*) acetic acid for 60 minutes and stored at 4°C pending mass spectrometry.

2.12.3. Mass spectrometry

Samples digested with trypsin were analysed by MALFI-TOF/TOF (matrix-assisted laser desorption ionisation—time-of-flight in tandem) mass spectrometry in collaboration with Dr Benjamin Thomas and Gabriela Ridlova (Proteomics Facility, Sir William Dunn School of Pathology, University of Oxford). Analysis of chymotrypsin-digested samples by nano-UPLC-MS/MS (nano–ultra performance liquid chromatography–tandem mass spectrometry) was performed by Dr Mukram Mackeen (Centre for Cellular and Molecular Physiology, University of Oxford).

2.13. RNA extraction

2.13.1. Total RNA extraction

Total RNA was extracted by a modification of the acid guanidinium isothiocyanate–phenol–chloroform method (Chomczynski & Sacchi, 1987). Cells cultured in 90 mm dishes were lysed directly after medium aspiration by scraping in 0.6 ml ice-cold ***solution D*** (4 M guanidinium isothiocyanate, 25 mM sodium citrate pH 7.0, 0.5 % (*w*/*v*) sarkosyl, 100 mM β-mercaptoethanol). Lysate was thoroughly mixed after sequential addition of 67 µl of 2 M sodium acetate pH 4.0, 600 µl water-saturated phenol, and 240 µl chloroform:isoamyl alcohol (49:1). After incubation on ice for 10 minutes, the lysate was centrifuged at 16,000*g* for 10 minutes at 4°C. The upper aqueous phase was transferred to a new tube and RNA precipitated by addition of 833 µl isopropanol and incubation at −20°C for at least 60 minutes. After centrifugation at 16,000*g* for 15 minutes at 4°C, the pellet was resuspended in 300 µl solution D and the RNA precipitated with 200 µl isopropanol at −20°C for 2 hours. The RNA was recovered by centrifugation as above and the pellet washed sequentially in 70% (*v*/*v*) ethanol and absolute ethanol (both stored at −20°C).

Pellets were briefly air-dried before dissolving in 50 µl sterile DMPC-treated Milli-Q. RNA was subsequently quantified by a NanoDrop™ ND-1000 spectrophotometer.

2.13.2. Poly(A) RNA selection

Poly(A)+ RNA was purified from total RNA (extracted as in section 2.13.1) using Promega PolyATract® mRNA Isolation System III according to the manufacturer's instructions. Total RNA (1 mg) was solubilised in 0.5 ml nuclease-free water and heated at 65°C for 10 minutes before adding 13 µl 20× SSC and 3 µl biotinylated-oligo(dT) probe. Samples were mixed gently and incubated at room temperature for 10 minutes. Streptavidin paramagnetic particles (SA-PMPs) were washed three times with 300 µl 0.5× SSC and resuspended in 100 µl 0.5× SSC before addition of RNA samples. SA-PMPs were incubated at room temperature for 10 minutes and mixed gently by inversion every 1–2 minutes. SA-PMPs were captured magnetically and poly(A)-minus supernatants transferred to new tubes. SA-PMPs were washed four times with 300 µl 0.1× SSC and individual wash fractions retained. Poly(A)+ RNA was eluted sequentially with 100 µl, and then 150 µl, nuclease-free water. Contaminating SA-PMPs were removed from fractions by centrifugation at 16,000*g* for 1 minute at 4°C.

RNA from poly(A)+ and wash fractions was precipitated by addition of 0.1 volumes of 3 M sodium acetate pH 5.2 and one volume of isopropanol, and incubation overnight at −20°C. RNA was sedimented by centrifugation at 16,000*g* for 20 minutes at 4°C; then washed sequentially with 70% (*v*/*v*) ethanol and absolute ethanol, before briefly air-drying.

2.14. RNA gels and electrophoresis

RNA samples were separated by denaturing electrophoresis through 1% (*w*/*v*) agarose/MOPS gels containing formaldehyde according to (Hazzalin *et al.*, 1997). RNA samples (3–5 µg) in 7.3 µl DMPC-treated Milli-Q were mixed with 10 µl formamide, 0.7 µl of 37% (*v*/*v*) formaldehyde, and 2 µl ***5× formaldehyde gel–electrophoresis buffer*** (0.1 M 3-(N-morpholino)propanesulphonic acid (MOPS) pH 7.0, 40 mM sodium acetate, 5 mM EDTA pH 8.0). Samples were heated at 65°C for 15 minutes, immediately cooled on ice, and

then mixed with 2 µl formaldehyde gel–loading buffer (50% (*v*/*v*) glycerol, 1 mM EDTA pH 8.0, 0.25% (*w*/*v*) bromophenol blue, 0.25% (*w*/*v*) xylene cyanol FF) and 1 µl of 1 mg/ml ethidium bromide. RNA gels contained 1% (*w*/*v*) agarose, 1.23% (*v*/*v*) formaldehyde and 1× formaldehyde gel–electrophoresis buffer, and were subject to pre-electrophoresis at 65 V for 5 minutes in 1× formaldehyde gel–electrophoresis buffer, before sample loading. Prepared identically to RNA samples, 3 µg RNA markers (Promega G319A) were also loaded for size estimation. Electrophoresis was performed at 65 V for 2 hours, after which gels were documented under a UV transilluminator.

2.15. Northern blotting

2.15.1. Northern transfer to nylon membranes

RNA was transferred to positively charged nylon membrane (Hybond™-N, GE Healthcare) by alkaline capillary transfer in 0.05 M NaOH for 4 hours. Membranes were neutralised in ***2× SSC*** (0.3 M NaCl, 30 mM sodium citrate pH 7.0) before drying at room temperature.

2.15.2. Preparation of radiolabelled probes

DNA for probe generation was generated by restriction digestion (**Table 2.6**); appropriate fragments were purified by gel extraction (section 2.17.2) and quantified by spectrophotometry.

Table 2.6 – DNA probes for northern blotting

Probe	Description	Sequence*	Size (bp)	Derivation
***Setd2*(15»19)**	Mouse *Setd2* 3' region	1034–1640	893	KpnI/BstXI digestion of CA319728 clone
***Setd2*(15»21)**	Mouse *Setd2* 3'region	1034–2377	1343	KpnI/Sac digestion of CA319728 clone

* exon numbers correspond to murine *Setd2* (NCBI RefSeq: NM_001081340.2).

Random-primed radiolabelled probes were prepared from 25 ng DNA and α-[^{32}P]-dCTP using Rediprime™ II DNA Labelling System (GE Healthcare) according to the manufacturer's instructions. Unincorporated α-[^{32}P]-dCTP was removed with SigmaSpin™ Post-Reaction Purification Columns (Sigma). The *Gapdh* probe was a 1 kb fragment of murine *Gapdh* cDNA, kindly prepared by Dr Nicholas Crump.

2.15.3. Hybridisation of radiolabelled probes

Membranes were blocked in ***Church buffer*** (0.5 M $NaHPO_4$ pH 7.2, 1% (*w*/*v*) bovine serum albumin fraction V, 1 mM EDTA pH 8.0, 7% (*w*/*v*) SDS) for 1 hour at 65°C (Church & Gilbert, 1984). Radiolabelled probes were denatured by heating at 100°C for 5 minutes followed by immediate incubation on ice for 5 minutes. Probes were added directly to Church buffer and hybridised at 65°C for 16–20 hours. Membranes were washed twice with 2× SSC/0.1% (*w*/*v*) SDS for 10 minutes at 65°C, then twice with 0.2× SSC/0.1% (*w*/*v*) SDS for 10 minutes at 65°C, before exposing to X-ray film (Kodak) between intensifying screens at −80°C.

For sequential hybridisation with *Gapdh* probe, membranes were first stripped of *Setd2* probe with three 30 minute incubations of a boiling solution of 0.1% (*w*/*v*) SDS, then washed twice at room temperature with 2× SSC, before blocking and hybridisation as above.

2.16. Reverse transcription–PCR (RT-PCR)

2.16.1. DNase treatment

Contaminating genomic DNA was removed from total RNA by TURBO DNA-free™ Kit (Life Technologies). Typically, 5 μg total RNA was digested in a 50 μl volume with 1 μl Turbo DNase I at 37°C for 30 minutes. After addition of 5 μl DNase Inactivation Reagent and incubation at room temperature for 2 minutes, samples were centrifuged at 16,000*g* for 2 minutes at room temperature and supernatants transferred to new tubes. DNase-treated RNA was immediately reverse transcribed or stored at −20°C.

2.16.2. Reverse transcription

First-strand cDNA was synthesized from DNase-treated total RNA using murine moloney leukaemia virus (MMLV) reverse transcriptase (RT; Life Technologies). Reactions were prepared with 2 µl 10× first-strand synthesis buffer, 4 µl of 2.5 mM dNTPs, 10 µl nuclease-free water and 1 µl of 50 µM random hexamers (2.5 µM final; Life Technologies). One microlitre of total RNA (≤ 1 µg) was added, and denatured at 70°C for 5 minutes before immediate chilling on ice for 5 minutes. Reactions were thoroughly mixed with 1 µl RNase inhibitor (40 U/µl; Life Technologies) and either 1 µl MMLV-RT (100 U/µl) for RT+ reactions or 1 µl nuclease-free water for RT-negative reactions. Samples were incubated at room temperature for 10 minutes (annealing); 42°C for 60 minutes and 44°C for 10 minutes (reverse transcription); 99°C for 5 minutes (RT inactivation). cDNA was stored at 4°C for immediate use or −20°C for long-term storage.

2.16.3. PCR

PCR was performed using the AmpliTaq Gold® (Life Technologies) system:

7.4 µl	PCR-grade water (Sigma)	
2.5 µl	10× PCR Gold buffer (without $MgCl_2$)	1× final
2.5 µl	10–25 mM $MgCl_2$	1–2.5 mM final
2.5 µl	2 mM dNTPs	0.2 mM final
0.25 µl	2.5 µM Forward primer	0.25 µM final
0.25 µl	2.5 µM Reverse primer	0.25 µM final
5 µl	First-strand cDNA (1:100 dilution)	
0.1 µl	AmpliTaq Gold® polymerase (5 U/µl)	0.5 U final

Primers are specified in

Table 2.8. RT-negative and non-template controls were prepared in parallel, where cDNA template was replaced with RT-negative cDNA or PCR-grade water, respectively. Thermocycling conditions are detailed in **Table 2.7**. All PCR reactions employed a common initial denaturation at 95°C for 10 minutes and a final elongation at 72°C for 7 minutes.

Table 2.7 – RT-PCR conditions

PCR reaction	Stage	Denaturation	Annealing	Elongation	Cycles
12F | 12cR	1	95°C for 30 s	54°C for 30 s	72°C for 120 s	45
8F | 9.5R	1	95°C for 30 s	55°C for 30 s	72°C for 60 s	40
12F | 12bR					
12bF | 13R					
12bF | 14R					
12bF | 15R					
17F | 18.5R					
12F | 13R	1	95°C for 30 s	53°C for 30 s	72°C for 120 s	40
12F | 14R					
35F | 13R					
6F | 10.5R	1	94°C for 60 s	55°C for 60 s	68°C for 120 s	15
	2	94°C for 60 s	55°C for 60 s	68°C for 120 s + 5 s/cycle	35
11F | 12bR	1	94°C for 60 s	55°C for 60 s	68°C for 120 s	15
	2	94°C for 60 s	55°C for 60 s	68°C for 120 s + 5 s/cycle	35

Table 2.8 – PCR primers for RT-PCR and DNA sequencing

Setd2 primer	Sequence (5' → 3')	Length	Tm (°C)[§]
6F	TCATGGCCCTGAAAAATGAC	20	58
8F	GACAGCTGAGGGTTGGATTT	20	60
9.5R	TCCAAACCTGGATTTTCCTG	20	58
10F	GCACTGATGGAAAATGGTGA	20	58
10.5R	TTTAACCGAAGGGGTAAGCA	20	60
11F[*]	TCATTTCTGGAACGTCATGG	20	58
12F	GAGGAAACACCATCCCAAGA	20	60
12bF	CATTTCCAACTTCGTGCTGA	20	58
12bF2[†]	GCGTGCATTCTGTGAAGAAA	20	58
12bR	CACAGAATGCACGCTAGGGTCT	22	68
12cR[‡]	TGCTGAGGCTAAAGTTGTGC	20	60
13R	TGTGCTCTCCTTTTCAGTTTGA	22	62
14R	TCATCTGGCCTTTTTGTTCC	20	58
15R	ATGTCATCCCCAAGTTCTGC	20	60
17F	ATCTACTGGACCTGCCACCA	20	62
18.5R	GGTGCATCTGACCCTTCAAT	20	58

[*] documented as Setd2–(35)F.
[†] documented as Setd2–(45)F.
[‡] documented as Setd2–12b.5R
[§] calculated as 2 × AT + 4 × CG

All PCR products were analysed by gel electrophoresis (section 2.17.2). Abundant PCR products were purified by gel extraction and identified by both forward and reverse sequencing using appropriate primers (detailed in section 2.17.3).

2.17. DNA extraction, electrophoresis and sequencing

2.17.1. Plasmid purification

For cloning, plasmids were isolated by standard alkaline-lysis (Sambrook *et al.*, 1989). For DNA sequencing and transfection of 10T½ cells, plasmids were further purified by phenol–chloroform extraction, ethanol precipitation, and PEG precipitation by standard procedures (Sambrook *et al.*, 1989).

2.17.2. DNA electrophoresis and gel purification

DNA samples and markers (1 kb or 100 bp, Life Technologies) were mixed with 5 volumes of 6× DNA loading buffer (60% (*v*/*v*) glycerol, 120 mM EDTA pH 8.0, 0.3% (*w*/*v*) Orange G) and separated on 0.8–2.0% (*w*/*v*) agarose/TAE (40 mM Tris–acetate pH 8.0, 1 mM EDTA) gels containing 0.5 μg/ml ethidium bromide (EtBr). Electrophoresis was performed at 100 V in 1× TAE containing 0.5 μg/ml EtBr until adequate resolution was achieved, judged by UV transillumination. DNA was extracted from gels and purified using GeneClean Turbo Kit (MP Biomedicals).

2.17.3. DNA sequencing

DNA was pre-sequenced using BigDye v3.1 system (Life Technologies) according to the manufacturer's instructions. Thermocycling conditions were: 96°C for 1 minute; 25 cycles of [96°C for 10 s; 53°C for 5 s; 60°C for 4 minutes]. Sequencing was performed by Bioscience Lifescience (University of Oxford). Chromatograms were analysed using Chromas (Technelysium); sequences were identified by BLAST (Altschul *et al.*, 1997); and aligned to *Setd2* using ClustalW (Larkin *et al.*, 2007).

Chapter 3 Immunodetection of p290^{SETD2} and putative SETD2 isoforms

3.1. Introduction

The existence of putative SETD2 protein isoforms in mammalian cells has been previously suggested by various immunoblotting studies and identification by mass spectrometry (Passani *et al.*, 2000; Rega *et al.*, 2001; Garcia *et al.*, 2004; Fujiki *et al.*, 2009; Wegrzyn *et al.*, 2010). However, the nature of these putative forms has received little further, rigorous analysis; their exact identity and functional significance remains unclear. This chapter aims to identify endogenous forms of SETD2 present in mammalian cells through use of antibodies and RNA interference. Experiments presented throughout this thesis predominantly involve murine C3H 10T½ fibroblasts as a model system, because aspects of SETD2 biology relevant to the dynamics and quantification of histone modification, and gene induction have been characterised previously (Edmunds *et al.*, 2008).

3.1.1. Antibodies as tools to investigate the identity and function of endogenous proteins

The combination of immunoblotting and size-analysis by electrophoresis is a powerful technique to assess the complement of full-length endogenous isoforms of a protein of interest.

Antibodies have several advantages as analytical reagents: they are capable of exquisite specificity of interaction and are highly sensitive with the capacity to detect sub-picomoles of antigen. However, there are several disadvantages. Firstly, cross-reactivity of serum can confound analysis, as is exemplified by problems in cross-reactivity of antibodies to epitopes in differently modified histones (Egelhofer *et al.*, 2011). Although monoclonal antibodies can partially alleviate problems of cross-reactivity, crude polyclonal serum is highly susceptible to non-specific binding and cross-reactivity. Specificity can be improved by peptide affinity–purification but may not improve problems with cross-reactivity to

related epitopes or mimotopes. Polyclonal sera, unlike monoclonal antibodies, are also more susceptible to antigenic drift, making comparable analyses between different bleeds and commercial batches of antibody sometimes difficult (due to batch-to-batch variability in terms of specificity and avidity). The interpretation of antibody signals is also complicated due to potential epitope occlusion by post-translational modifications that may or may not be known *a priori* (Clayton *et al.*, 2000). Even antibodies raised against identical peptide immunogens can exhibit different affinities and specificities towards epitopes and capacities to detect endogenous proteins from cells (Wilkinson *et al.*, 1999; Clayton *et al.*, 2006). Finally, antibodies may recognise specific conformations of peptides that may only exist under specific native or denatured conditions, limiting their utility in certain techniques (Steger *et al.*, 2008).

3.1.2. RNA interference to investigate protein identity and function

RNA interference (RNAi) involves post-transcriptional silencing of gene expression by small double-stranded RNA molecules (≈20–21 bp). Double-stranded RNA (dsRNA) is recognised and cleaved by Dicer into single-stranded RNA molecules that associate with argonaute proteins to direct translation inhibition and cleavage of target mRNAs that are complementary to the argonaute-bound siRNA (Matzke & Birchler, 2005).

Whilst production of large dsRNA in mammalian cells triggers interferon responses, the transfection of small dsRNA bypasses this response (Matzke & Birchler, 2005). RNAi may be induced transiently by lipofectamine-mediated transfection of siRNA molecules (Whitehead *et al.*, 2009), or may be more stably directed by expression of short hairpin RNAs from lentiviral vectors (Stewart *et al.*, 2003).

Transient siRNA transfection thus allows rapid, efficient alterations in mammalian systems. Genomic manipulations in mammals are complex, resource-intensive and may be embryonic lethal. Unlike gene-knockout by homologous recombination, which may interfere with independent processes across a gene (for example, use of alternative promoters), RNAi has the potential to eliminate specific transcript isoforms at the post-transcriptional level. Furthermore, transient siRNA transfection avoids secondary, adaptive

events that may compensate for gene knockout in traditional genomic manipulations. Combined with immunological screening of the endogenous proteins, RNA interference is an important technique for the identification of proteins and their effects on cellular function (Hannon, 2003).

Nevertheless, several caveats with RNAi exist. Firstly, the success of RNAi depends on the susceptibility of the mRNA to hybridisation with the siRNA-argonaute complex, which is influenced not only by the degree of complementarity between the siRNA sequence and the mRNA target, but also on the accessibility of the mRNA target sequence, which may be influenced by the formation of secondary structures and occlusion by RNA-binding proteins. Secondly, for unclear reasons, RNAi effects on proteins and transcripts may not be complete—*i.e.* only partial reductions are observed. Thirdly, because RNAi targets RNA rather than protein directly, proteins that turnover relatively slowly, may not be amenable to knockdown by transient or even longer-term siRNA exposure. Fourthly, siRNAs are prone to off-target effects; thus ideally, multiple reagents should be screened to ensure specificity. Finally, since RNAi targets entire transcripts to deplete proteins, RNAi cannot be used to address the functions of individual domains of a protein (Hannon, 2003).

3.2. Objectives

Previous analyses of mammalian SETD2 proteins conflict with respect to reported molecular weights as determined by SDS-PAGE (**Table 1.4**). Evidence for alternative forms has been reported from independent studies by immunoblotting and mass spectrometry, but no rigorous evaluation or further analysis has so far been developed. The objectives of this chapter were thus as follows:

- To generate specific, affinity-purified antibodies to SETD2.
- To characterise the endogenous forms of SETD2 by immunoblotting protein extracts from murine C3H 10T½ fibroblasts and other mammalian cell lines.
- To evaluate putative SETD2 isoforms by transient siRNA transfection.

3.3. Results

Polyclonal antisera from rabbit and sheep were previously generated against peptides derived from an N-terminal and C-terminal region of SETD2, respectively, as detailed in Table 3.1 below (Edmunds *et al.*, 2008). The N-terminal peptide was raised against an unstructured region upstream of the AWS domain whilst the C-terminal peptide partially overlaps a predicted coiled-coil motif, upstream of the QP-rich region (**Figure 1.2**).

Table 3.1 – α-SETD2 antibodies

Antibody	Peptide sequence	Residues*	Coding Exons†	Identity with human SETD2‡	Source
N-term	MERRGKYTSKLERESKRTSE-C	504–523	3	95%	NSL R2039 final bleed
C-term	RGTKRPDDRYDTPTSKKKVR-C	2063–2082	14–15	100%	NSL S950B final bleed
CT21	C-QKFGAVYKPKEDTE	2522–2535	21	100%	Everest Biotech eb07418

* from murine SETD2 protein (UniProt KB accession E9Q5F9, 2537 amino acids).
† exon numbers correspond to murine *Setd2* transcript (RefSeq: NM_001081340.2).
‡ human SETD2 protein (UniProt KB accession Q9BYW2, 2564 amino acids).

Preliminary peptide-competition experiments demonstrated that, despite specific detection of SETD2 peptides, the antisera contained polyclonal antibodies that substantially cross-react with other proteins—unrelated to SETD2 sequence—in extracts of murine C3H 10T½ fibroblasts (*cf.* **Figure 3.1D**, lane 1 versus 2; **Figure 3.3D**, lane 1 versus 2). SETD2-specific polyclonal antibodies were therefore purified from antisera by peptide affinity chromatography. The efficiency and specificity of affinity-purification was evaluated by peptide competition immunoblotting against soluble Triton X-100–deoxycholate (TX–DOC) extracts from 10T½ fibroblasts. TX–DOC lysis yields a complex soluble fraction of cytoplasmic, nucleoplasmic and detergent-soluble membrane proteins, separated from a TX-DOC-insoluble fraction enriched in chromatin and cytoskeletal proteins, such as vimentin and actin, which sediment after centrifugation at 16,000*g* for 15 minutes (*cf.* **Figure 3.4-iv**) (Mahadevan *et al.*, 1988).

3.3.1. Affinity-purified rabbit antibodies to an N-terminal region of SETD2 detect p290^{N} and p120^{N} proteins in murine fibroblasts

3.3.1.1. *Affinity-purification of N-terminal-specific SETD2 antibodies*

The population of polyclonal antibodies that specifically bind epitopes of the peptide of interest constitute a minority of the total immunoglobulin present in antisera. The remaining antibodies recognise a wide-variety of unrelated epitopes, contributing to cross-reactivity: undesirable detection of unrelated proteins by immunoblotting and other immunological techniques. To thus obtain a reagent capable of specific detection of proteins corresponding to the N-terminal region of SETD2, polyclonal antibodies were affinity-purified from rabbit serum using columns chemically coupled to N-terminal SETD2 peptide. In addition to removing common serum contaminants (such as albumin and serotransferrin; **Figure 3.1A**, lanes 3–6), the purification procedure importantly removed the majority of background immunoglobulin that reacted non-specifically with the complex mixture of proteins in cell extracts (**Figure 3.1D**, lane 1 versus 2).

To evaluate the efficiency and specificity of purification, pre-immune serum, antiserum and individual flow-through, wash, and elution fractions were used at equivalent dilutions to probe N-terminal and C-terminal SETD2 peptides spotted on nitrocellulose membrane. The load, flow-through, first wash, and elutions fractions all detected N-terminal SETD2 peptide (**Figure 3.1B-i*)*** but not C-terminal SETD2 peptide (**Figure 3.1B-ii**), and the recognition of N-terminal peptide was specifically competed by pre-adsorption of the fractions with cognate peptide only (**Figure 3.1B-iii**).

When TX−DOC extracts from mouse fibroblasts were immunoblotted using reagents at equivalent dilutions, the crude serum and flow-through detected multiple proteins (**Figure 3.1C**, lanes 1 and 2) whereas the elution fractions detected two predominant proteins: p120^{N} and p290^{N}, which migrated with apparent molecular masses of ≈120 kDa and above 250 kDa, respectively (**Figure 3.1C**, lanes 6−9; **Figure 3.1D**, lanes 5 and 7). Where protein bands migrated between molecular weight markers, the protein mass was estimated by

Figure 3.1 Generation of affinity-purified antibodies specific to an N-terminal region of SETD2

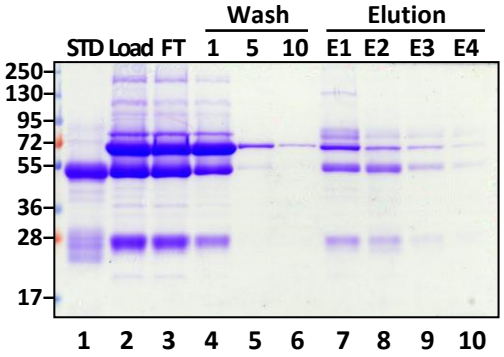
(A) Purification of α-SETD2 N-terminal polyclonal antibody from rabbit antiserum: Clarified α-SETD2 N-terminal rabbit antiserum was diluted in PBS and loaded onto an N-terminal SETD2 peptide–affinity column (Load). After collection of unbound flow-through (FT), the column was washed ten times with PBS, before bound polyclonals were recovered by low pH elution into four individual fractions (E1−E4). Equivalent proportions of purification fractions were resolved by 10% SDS-PAGE and Coomassie-stained. STD represents 5 μg of affinity-purified α-H3S10ph antibody. The approximate molecular mass of protein standards is indicated in kDa.

(B) Affinity-purified α-SETD2 N-terminal antibodies specifically react against N-terminal SETD2 peptide: Dot-blot analysis combined with peptide competition. Nitrocellulose membrane strips were spotted with 1 pmol N-terminal **(i, iii)** or C-terminal **(ii, iv)** SETD2 peptide and probed with pre-immune (PI) serum (lane 1), α-SETD2 N-terminal serum (lane 2) and purification fractions: flow-through (FT, lane 3); first, fifth and final (tenth) wash fractions (lanes 4, 5 and 6, respectively); and first and second elution fractions (lanes 7 and 8). For each dot-blot, immunoblotting (IB) reagent was pre-incubated alone **(i, ii)** or with 1 μM N-terminal SETD2 peptide **(iii, iv)**.

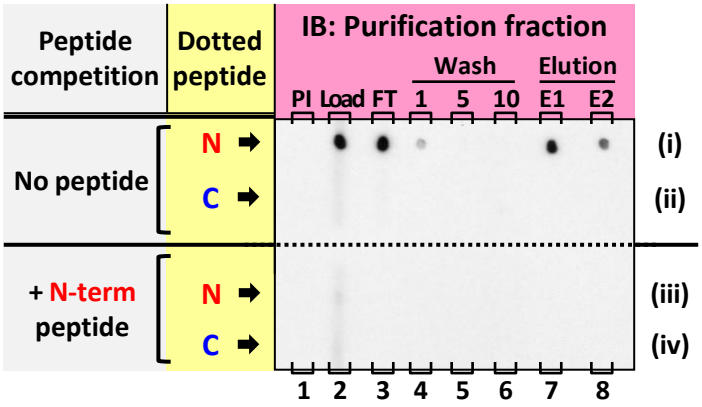
(C) Reactivity of α-SETD2 N-terminal serum and affinity-purified reagents towards mouse fibroblast proteins: Immunoreactivity of purification fractions against TX−DOC extracts from murine 10T½ fibroblasts, resolved by 8% SDS-PAGE. Membrane strips were probed with equivalent dilutions of α-SETD2 N-terminal serum (lane 1) and purification fractions: flow-through (FT, lane 2); first, fifth and final (tenth) wash fractions (lanes 3, 4 and 5, respectively); and four sequential elution fractions (lanes 6, 7, 8 and 9).

(D) Affinity-purified α-SETD2 N-terminal antibodies specifically detect p290^{N} and p120^{N} proteins: TX−DOC extract from quiesced 10T½ cells was resolved in a single lane by 8% SDS-PAGE and transferred to PVDF in SDS transfer buffer. Identical membrane strips were probed with equivalent dilutions of α-SETD2 N-terminal serum (lanes 1 and 2), flow-through (FT, lanes 3 and 4), elution fraction 1 (lanes 5 and 6), and elution fraction 2 (lanes 7 and 8). Each immunoblot reagent was pre-incubated in the absence (odd lanes) or presence (even lanes) of 1 μM N-terminal SETD2 peptide. Asterisks indicate non-specific proteins detected by the reagents, whilst arrowheads represent proteins that are specifically detected (eliminated by peptide pre-adsorption).

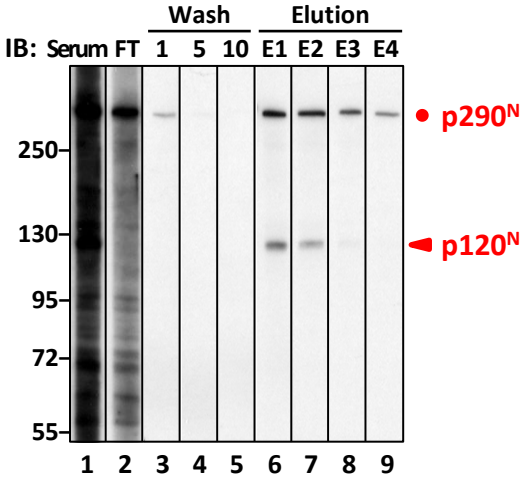
(A)


Wash
Elution
STD Load FT 1 5 10 E1 E2 E3 E4
250
130
95
72
55
36
28
17
1 2 3 4 5 6 7 8 9 10

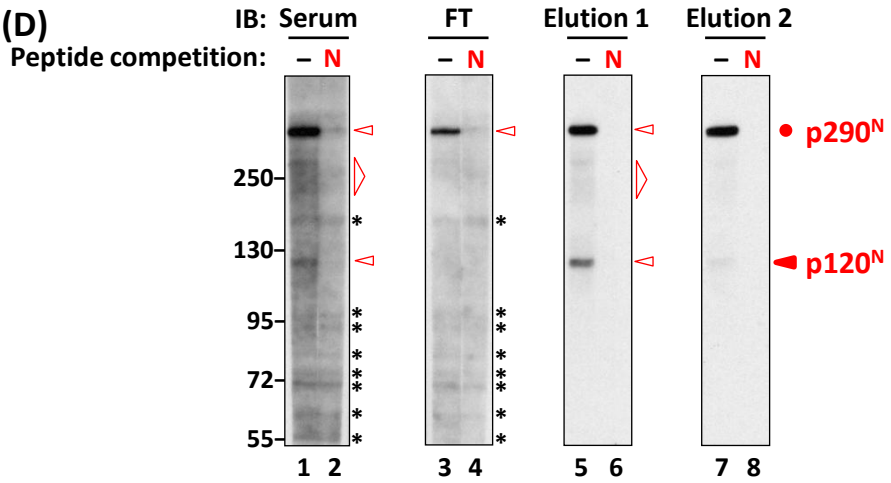
(B)


Peptide competition
Dotted peptide
IB: Purification fraction
Wash
Elution
PI Load FT 1 5 10 E1 E2
No peptide
N
C
+ N-term peptide
N
C
(i)
(ii)
(iii)
(iv)
1 2 3 4 5 6 7 8

(C)


Wash
Elution
IB: Serum FT 1 5 10 E1 E2 E3 E4
250
130
95
72
55
p290N
p120N
1 2 3 4 5 6 7 8 9

(D)


IB: Serum
FT
Elution 1
Elution 2
Peptide competition: – N – N – N – N
250
130
95
72
55
p290N
p120N
1 2 3 4 5 6 7 8

Laemmli plots (discussed in **Appendix A-1**). p290^{N} is named based on the 286 kDa mass of the protein predicted from the canonical *Setd2* gene model (RefSeq NP_001074809.2).

Recognition of these two proteins by the eluted antibodies was specific as judged by elimination of their immunodetection following pre-adsorption of antibody with N-terminal peptide (**Figure 3.1D**, lanes 5–8). Indeed only p290^{N} and p120^{N} proteins were eliminated by peptide competition when extracts were immunoblotted with crude serum, demonstrating that the additional proteins recognised by the serum were unrelated to SETD2 (**Figure 3.1D**, lane 1 versus 2). As expected, the cross-reactive proteins were detected only by non-specific polyclonal antibodies that were present in the flow-through fraction (**Figure 3.1D**, lane 3 versus 4).

While affinity-purification yielded virtually complete binding of immunoglobulins able to detect p120^{N} (*cf.* lack of p120^{N} signal in flow-through fraction compared to serum – **Figure 3.1D**, lane 1 versus 3), the purification of polyclonals detecting p290^{N} was less efficient, since the flow-through fraction still contained immunoglobulin that specifically recognised p290^{N} (**Figure 3.1D**, lane 1 versus 3). This was also reflected in the ability of the flow-through fraction to bind specifically to N-terminal peptide with intensity comparable to serum (**Figure 3.1B**, lane 2 versus 3). Alteration of peptide-coupling densities during affinity column generation and concentrations of serum load during purification did not improve yield (data not shown). Poor yield may derive from a proportion of specific polyclonal antibodies that, despite strong affinity for aqueous or membrane-adsorbed peptide, cannot bind peptides chemically cross-linked to the column matrix, potentially due to steric hindrance or epitope modification.

Nevertheless, the affinity-purified antibodies exclusively and specifically recognised epitopes within a defined N-terminal SETD2 sequence: they thus represent a substantially improved reagent over crude serum for specific characterisation of endogenous SETD2 proteins bearing these N-terminal epitopes.

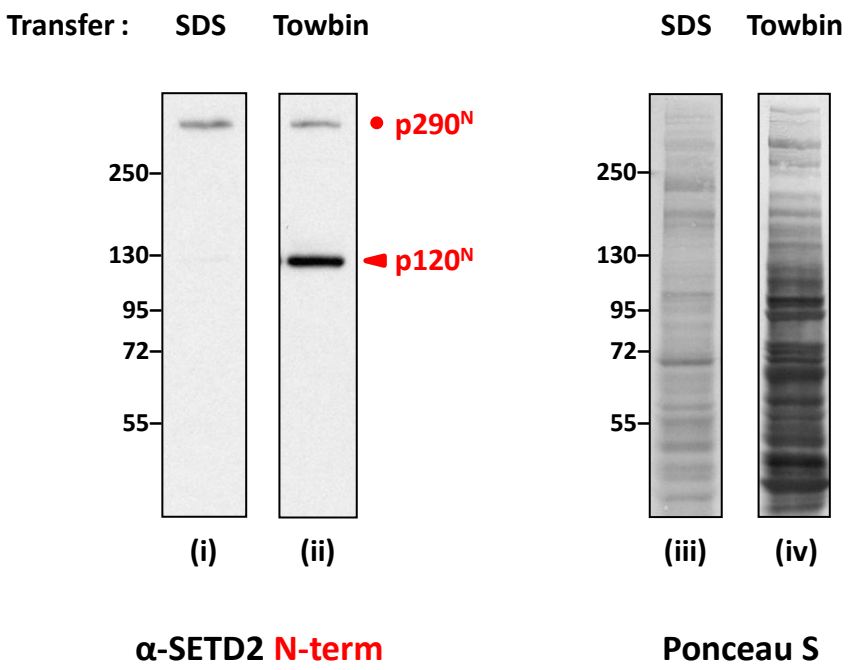
3.3.1.2. *Transfer conditions influence the relative intensities of p290^{N} and p120^{N}*

Based on the previous immunoblotting studies with crude antiserum, initial immunoblot experiments employed SDS transfer buffer (Edmunds *et al.*, 2008). Although affinity-purified α-SETD2 N-terminal antibody reproducibly detected p290^{N} and p120^{N}, variability in the relative levels of the two proteins was often noticed until it was discovered that the transfer condition was critical for reproducible detection of p120^{N} (**Figure 3.2**). This result is perhaps not surprising given that different protein profiles and intensities can be observed by Ponceau S stain following transfer in SDS or Towbin buffers (**Figure 3.2-iii/iv**).

Although it is unknown to what extent different transfer conditions lead to analysis of different sub-populations of protein molecules, no differences in behaviour between Towbin-transferred or SDS-transferred p290^{N} were observed during the course of this study. For this reason, and for comparability and the most reproducible detection of both p120^{N} and p290^{N} on the same membrane, transfer using Towbin buffer was the preferred method for analysis of N-terminal species. All further experiments presented using α-SETD2 N-terminal antibody were thus prepared by Towbin transfer to permit detection of both p290^{N} and p120^{N}.

Figure 3.2 Transfer conditions influence detection of p290^{N} and p120^{N} proteins by affinity-purified α-SETD2 N-terminal antibody

TX/DOC extracts from quiesced C3H 10T½ fibroblasts were separated by 8% SDS-PAGE and transferred to PVDF membrane in SDS buffer **(i, iii)** or Towbin buffer **(ii, iv)**. Membranes were stained with Ponceau S **(iii, iv)** to visualise total transferred protein and subsequently probed with affinity-purified N-terminal α-SETD2 antibody **(i, ii)**.


Transfer :
SDS
Towbin
SDS
Towbin
p290N
p120N
250
130
95
72
55
250
130
95
72
55
(i)
(ii)
(iii)
(iv)
α-SETD2 N-term
Ponceau S

3.3.2. Affinity-purified sheep antibodies to a C-terminal region of SETD2 specifically detect multiple proteins in murine fibroblasts

Crude α-SETD2 C-terminal serum was also purified by an affinity-column coupled to the C-terminal SETD2 peptide to which the serum was originally raised (**Figure 3.3A**). The purification yielded elution fractions that specifically detected C-terminal SETD2 peptide by peptide competition dot blot (**Figure 3.3B**).

The C-terminal antiserum and flow-through fraction detected multiple proteins in TX−DOC extracts from murine fibroblasts (**Figure 3.3C**, lanes 1−4) but the majority of these proteins were detected non-specifically as judged by peptide competition (**Figure 3.3D**, lane 1 versus 2, and lane 3 versus 4). p290^{C} and p45^{C} were the only obvious proteins detected by the antiserum that were depleted by pre-adsorption with C-terminal SETD2 peptide (**Figure 3.3D**, lane 1 versus 2). However, with removal of background cross-reactivity from polyclonal antibodies present in the flow-through (**Figure 3.3D**, lanes 3 and 4), the eluted antibodies were able to clearly detect multiple proteins in these extracts: p290^{C}, p250^{C}, p140^{C}, p80^{C} and p45^{C}. (**Figure 3.3C**, lanes 8−11). A protein of 66 kDa was also faintly detected by the affinity-purified antibody, which may or may not correspond to a specific and conspicuous p66^{C} protein, which will be described in section 3.3.3. Importantly, all of these immunoreactive proteins were detected specifically as judged by competition with C-terminal SETD2 peptide (**Figure 3.3D**, lane 5 versus 6).

Even when these immunoreactive signals were overexposed so that additional minor bands that migrate between ≈75 and ≈150 kDa were revealed (**Figure 3.3D**, lane 7), all signals were eliminated with pre-adsorption of the affinity-purified antibody with C-terminal SETD2 peptide (**Figure 3.3D**, lane 8), demonstrating the high degree of specificity of the purified polyclonal antibodies to this epitope. The 75−150 kDa species observed on immunoblot overexposure probably represent proteolytic degradation products of higher molecular weight immunoreactive proteins during sample preparation,

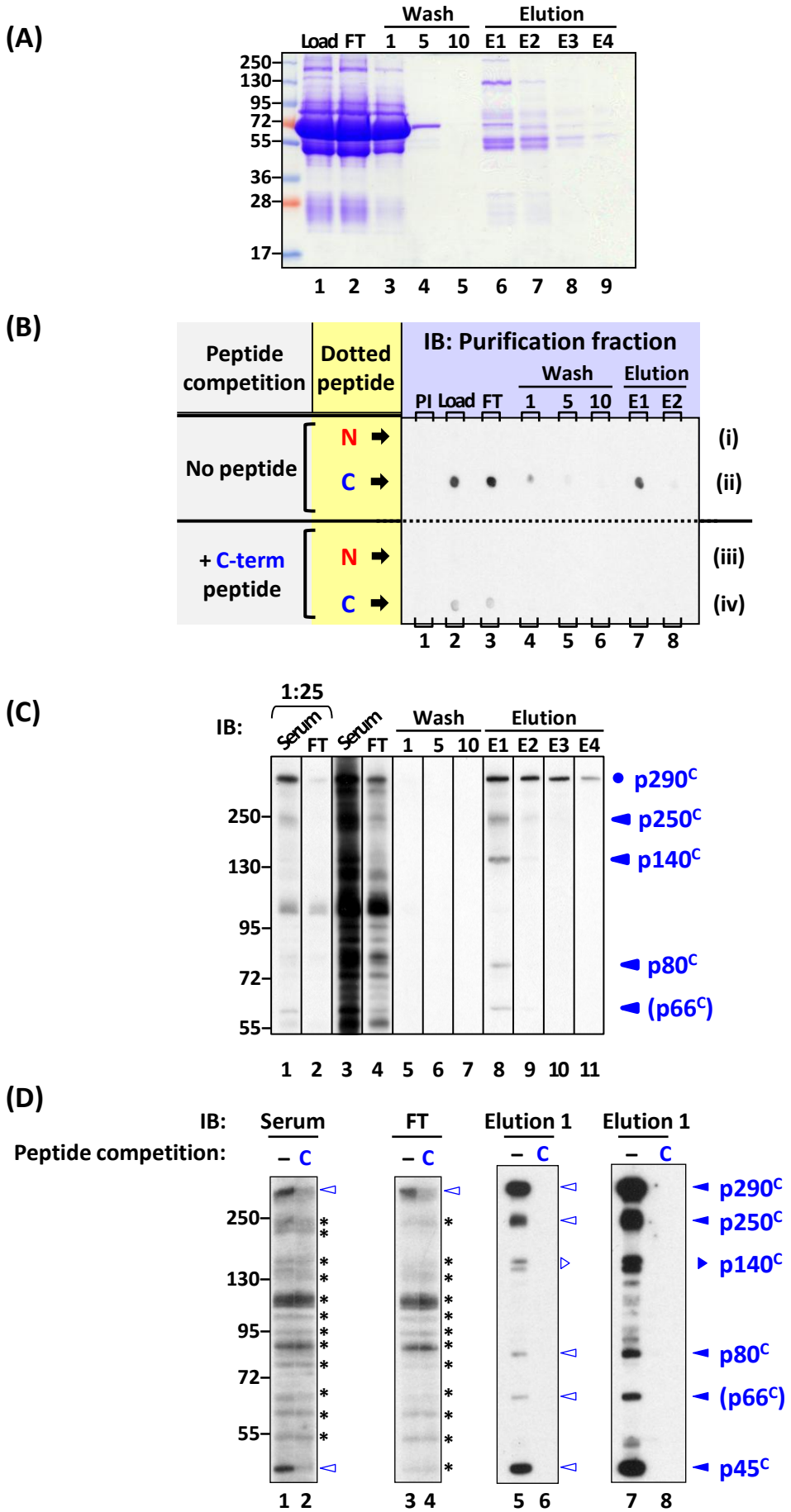
Figure 3.3 Generation of affinity-purified antibodies specific to a C-terminal region of SETD2

(A) Purification of α-SETD2 C-terminal polyclonal antibody from sheep antiserum: Clarified α-SETD2 C-terminal rabbit antiserum was diluted in PBS and loaded onto a C-terminal SETD2 peptide–affinity column (Load). After collection of unbound flow-through (FT), the column was washed ten times with PBS, before bound polyclonals were recovered by low pH elution into four individual fractions (E1−E4). Equivalent proportions of purification fractions were resolved by 10% SDS-PAGE and Coomassie-stained.

(B) Affinity-purified α-SETD2 C-terminal antibodies specifically react against C-terminal SETD2 peptide: Dot-blot analysis combined with peptide competition. Nitrocellulose membrane strips were spotted with 1 pmol N-terminal **(i, iii)** or C-terminal **(ii, iv)** SETD2 peptide and probed with pre-immune (PI) serum (lane 1), α-SETD2 C-terminal serum (lane 2) and purification fractions: flow-through (FT, lane 3); first, fifth and final (tenth) wash fractions (lanes 4, 5 and 6, respectively); and first and second elution fractions (lanes 7 and 8). For each dot-blot, immunoblotting (IB) reagent was pre-incubated alone **(i, ii)** or with 1 μM C-terminal SETD2 peptide **(iii, iv)**.

(C) Reactivity of α-SETD2 C-terminal serum and affinity-purified reagents towards mouse fibroblast proteins: Immunoreactivity of purification fractions against TX−DOC extracts from murine 10T½ fibroblasts, resolved by 8% SDS-PAGE. Membrane strips were probed with equivalent dilutions of α-SETD2 C-terminal serum (lane 3) and purification fractions: flow-through (FT, lane 4); first, fifth and final (tenth) wash fractions (lanes 5, 6 and 7, respectively); and four sequential elution fractions (lanes 8, 9, 10 and 11). Membranes strips 1 and 3 were probed with a 25-fold greater dilution of serum and FT fraction, respectively.

(D) Affinity-purified α-SETD2 C-terminal antibodies specifically detect multiple proteins: TX−DOC extract from quiesced 10T½ cells was resolved in a single lane by 8% SDS-PAGE and transferred to PVDF in SDS transfer buffer. Identical membrane strips were probed with equivalent dilutions of α-SETD2 C-terminal serum (lanes 1 and 2), flowthrough (FT, lanes 3 and 4), elution fraction 1 (lanes 5, 6, 7 and 8). Lanes 7 and 8 represent an over-exposure of lanes 5 and 6. Each immunoblot reagent was pre-incubated in the absence (odd lanes) or presence (even lanes) of 1 μM C-terminal SETD2 peptide. Asterisks indicate non-specific proteins detected by the reagents, whilst arrowheads represent proteins that are specifically detected (eliminated by peptide pre-adsorption).


(A)
Wash
Elution
Load FT 1 5 10 E1 E2 E3 E4
250
130
95
72
55
36
28
17
1 2 3 4 5 6 7 8 9
(B)
Peptide competition
Dotted peptide
IB: Purification fraction
Wash
Elution
PI Load FT 1 5 10 E1 E2
No peptide
N
C
(i)
(ii)
+ C-term peptide
N
C
(iii)
(iv)
1 2 3 4 5 6 7 8
(C)
IB:
1:25
Serum
FT
Serum
FT
Wash
Elution
1 5 10 E1 E2 E3 E4
250
130
95
72
55
p290C
p250C
p140C
p80C
(p66C)
1 2 3 4 5 6 7 8 9 10 11
(D)
IB:
Serum
FT
Elution 1
Elution 1
Peptide competition:
– C
– C
– C
– C
250
130
95
72
55
p290C
p250C
p140C
p80C
(p66C)
p45C
1 2
3 4
5 6
7 8

but there is always the possibility that they may represent forms of lower abundance that arise naturally *in vivo*.

Whilst the affinity-purified N-terminal antibodies consistently detected only p290^{N} and p120^{N} proteins between independent experiments, the intensity of different immunoreactive proteins detected by affinity-purified C-terminal antibodies often varied between experiments. Subtle variations may be influenced by transfer conditions (*cf.* **Figure 3.7C**) but no consistent correlations between transfer in SDS or Towbin buffer and band intensity were discernible across independent experiments. The origin of the variation in C-terminal signals is therefore an unresolved issue here, which could plausibly be influenced by subtle differences in cell culture confluence, length of quiescence or passage number, or differential changes to antibody integrity within a polyclonal population during storage over time.

Overall, the most reproducible C-terminal bands between independent experiments were p290^{C}, p140^{C} and the p66^{C} protein (described below). Nevertheless, regardless of the variability in patterns observed, the affinity-purified α-SETD2 C-terminal antibody importantly reacted only with proteins containing epitopes related to the C-terminal SETD2 peptide, suggesting that multiple α-SETD2 immunoforms may exist in murine fibroblasts.

3.3.3. p66^{C} is detected by C-terminal α-SETD2 antibody—when extracts are heated in urea

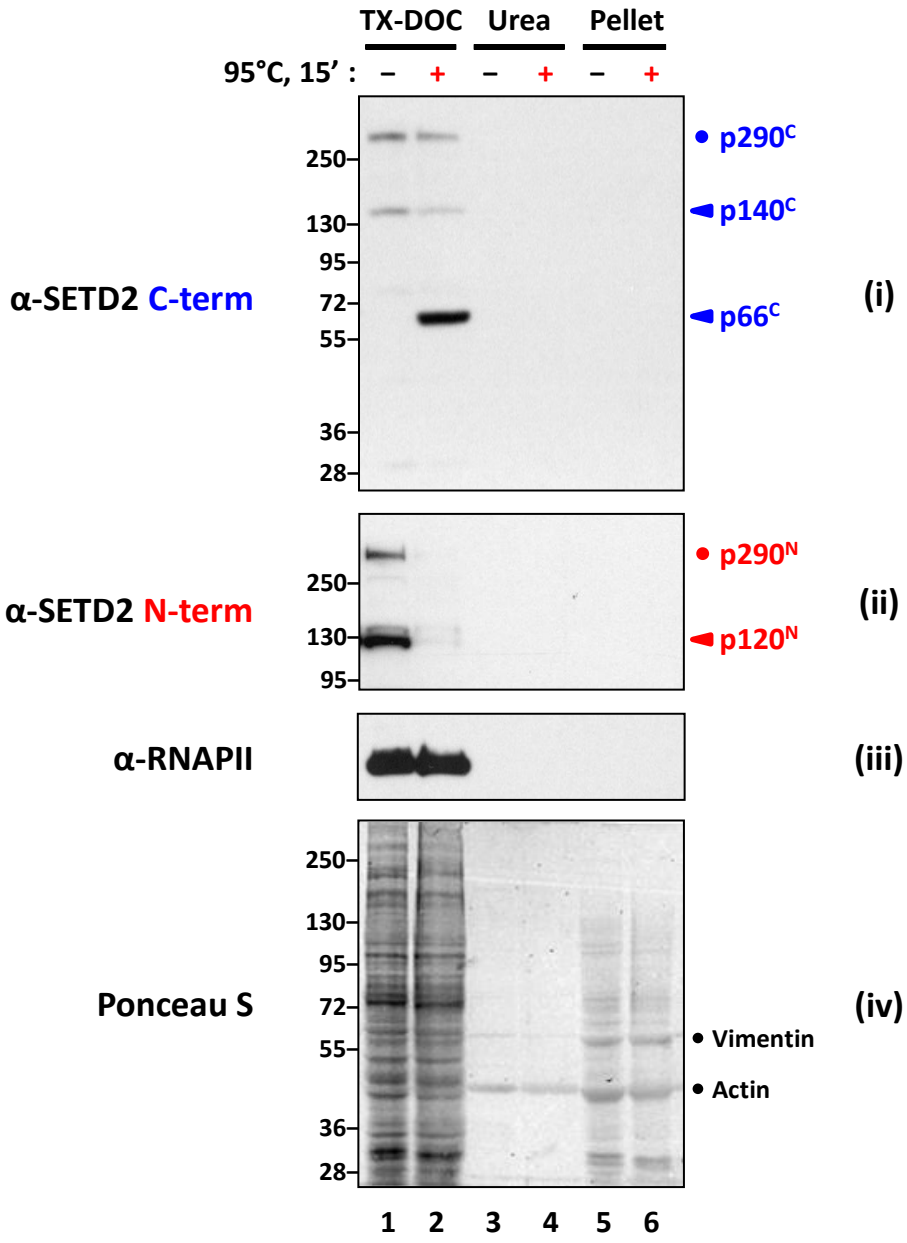
Due to differences in structure, interactions and subcellular localisation, proteins may be differentially solubilised in specific buffers, which is the basis for subcellular fractionation based on differential extraction (Berezney & Coffey, 1974; Ramsby & Makowski, 1999). SDS is frequently employed for complete solubilisation due to its strong denaturation of protein and membrane structure, and compatibility with SDS-PAGE for molecular weight determination (Weber & Osborn, 1969; Laemmli, 1970). Chaotropic urea and non-ionic detergents are preferred reagents for solubilisation for isoelectric focusing and two-dimensional (2D)–PAGE (O'Farrell, 1975). Lysis in SDS-containing Laemmli buffer yielded

extracts that were extremely susceptible to protein degradation resulting not only from proteolysis but also due to a sonication step that was necessary to reduce lysate viscosity (data not shown). For this reason, TX−DOC lysis was chosen as the preferred method of preparing protein extracts from mouse fibroblasts. However, during preliminary screening of extraction conditions, the use of Laemmli buffer in combination with urea yielded an initially perplexing result: when lysis buffer was supplemented with 4 M urea, an additional protein at ≈66 kDa was strongly detected by the α-SETD2 C-terminal antibody, yet attempts to detect this protein by extracting SDS-insoluble material with urea repeatedly failed (data not shown). Immunodetection of this p66^{C} protein was eventually realised to depend on an unusual *denaturation*—rather than extraction—condition, which will now be described.

The aforementioned immunoblotting analysis considered proteins present solely in the TX-DOC-soluble fraction and thus it was also important to assess the extent of immunoform retention in TX-DOC-resistant fractions. The TX−DOC-resistant pellet was therefore further extracted under harsher conditions (9 M urea, 0.1 M DTT, 0.1% (*v*/*v*) NP-40) to yield a urea-soluble extract (**Figure 3.4**, lanes 3 and 4); the urea-resistant pellet was solubilised in the presence of SDS-PAGE loading buffer containing 4 M urea, and sonicated (**Figure 3.4**, lanes 5 and 6). The inclusion of urea was based on the composition of O'Farrell's solubilisation buffer and on previous observations that histones are only reproducibly solubilised for SDS-PAGE in the presence of 4 M urea. The initial TX−DOC supernatant was precipitated with TCA, so that an equal proportion of each fraction, solubilised comparably in urea/SDS-PAGE loading buffer, could be analysed by SDS-PAGE. Sequential extraction demonstrated solubilisation of the abundant, detergent-resistant cytoskeletal elements, vimentin and actin (**Figure 3.4-iv**, lanes 3–6). The α-SETD2 species, p290$^{N/C}$, p140^{C} and p120^{N}, in addition to RNAPII, were almost completely extracted in the initial TX−DOC supernatant (**Figure 3.4:i−iii**, lanes 1 and 2). This buffer is therefore justified for preparing single SETD2-enriched extracts that maintain p290 integrity, appropriate for analysing the results of RNAi and changes in SETD2 levels after cell culture treatments (see sections 3.3.6 and 3.3.7).

Figure 3.4 Affinity-purified α-SETD2 C-terminal antibody detects p66^{C} in TX–DOC extracts after thermal denaturation

Quiesced 10T½ fibroblasts were lysed in TX–DOC buffer and centrifuged at 16,000*g* for 15 minutes to yield TX–DOC supernatant (lanes 1 and 2) and a detergent-resistant pellet, which was further extracted with urea-solubilisation buffer to yield the Urea supernatant (lanes 3 and 4) and final pellet (lanes 5 and 6). The TX–DOC extract was precipitated with TCA and all samples adjusted to urea/SDS-PAGE loading buffer before incubation at room temperature (odd lanes marked with "–") or at 95°C for 15 minutes (even lanes marked with "+"). Equivalent proportions of each fraction were separated by 8% SDS-PAGE and transferred to PVDF in SDS transfer buffer. Membranes were stained with Ponceau S **(iv)** before immunoblotting with α-SETD2 C-terminal **(i)**, N-terminal **(ii)**, or α-RNAPII **(iii)** antibodies.

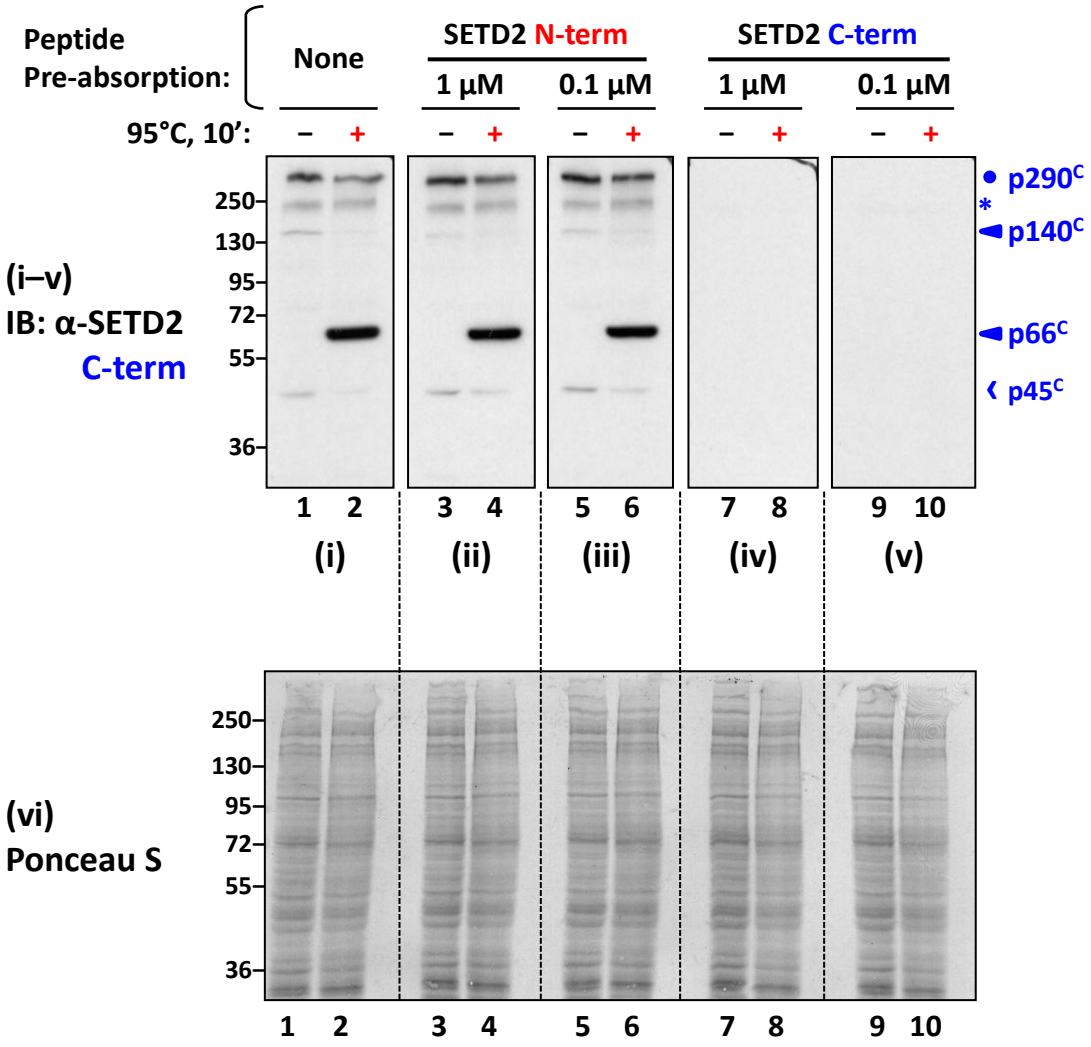

TX-DOC
Urea
Pellet
95°C, 15' : – + – + – +
α-SETD2 C-term
250
130
95
72
55
36
28
p290C
p140C
p66C
(i)
α-SETD2 N-term
250
130
95
p290N
p120N
(ii)
α-RNAPII
(iii)
Ponceau S
250
130
95
72
55
36
28
Vimentin
Actin
(iv)
1 2 3 4 5 6

When the aforementioned extracts were denatured at 95°C for 15 minutes, a striking additional protein of ≈66 kDa was strongly detected by the α-SETD2 C-terminal antibody, exclusively in the TX–DOC supernatant. After thermal denaturation in the presence of urea, the levels of p290^{C} and p140^{C} were reduced, but by a level incomparable to the increase in intensity of p66^{C} (**Figure 3.4-i**, lane 1 versus 2). N-terminal signals, however, appear to be severely decreased by thermal denaturation (**Figure 3.4-ii**, lane 1 versus 2). The intensity of RNAPII signal appears relatively unaffected by heating (**Figure 3.4-iii**, lane 1 versus 2). Indeed, Ponceau–stain revealed little change in total protein amount after thermal denaturation; importantly, no additional protein at ≈66 kDa appeared after heating (**Figure 3.4-iv**, lane 1 versus 2).

Specific detection of p66^{C} by α-SETD2 C-terminal antibody was confirmed by peptide competition: pre-adsorption of antibody with either 0.1 μM or 1 μM C-terminal SETD2 peptide eliminated detection of p66^{C}, p290^{C} and p140^{C} (**Figure 3.5-iv/v**). Pre-adsorption with N-terminal SETD2 peptide acted as a control for the effect of non-cognate peptide pre-adsorption on immunoblot signals (**Figure 3.5-ii/iii**). As mentioned above, additional bands, such as p45^{C}, were variably observed between experiments. Detection of this band, along with a poorly-resolved smear just below 250 kDa, were confirmed as specifically related to C-terminal SETD2 peptide sequence by peptide competition (**Figure 3.5**). The latter species may be the result of protein degradation.

Figure 3.5 Immunodetection of p66^{C} is eliminated by pre-adsorption with C-terminal SETD2 peptide

TX–DOC extracts from quiesced 10T½ cells were precipitated with TCA and resulting pellets were re-solubilised in urea/SDS-PAGE buffer at room temperature (odd lanes marked with "–") or at 95°C for 10 minutes (even lanes marked with "+"). Samples were resolved by 8% SDS-PAGE and transferred to PVDF in SDS transfer buffer. Five equivalent sections of Ponceau S–stained membrane **(vi)** were immunoblotted with α-SETD2 C-terminal antibody which had been pre-incubated for 30 minutes alone **(i)** or with SETD2 N-terminal peptide **(ii, iii)** or C-terminal peptide **(iv, v)** at either 1 μM **(ii, iv)** or 0.1 μM **(iii, v)** final concentrations.


Peptide Pre-absorption:
None
SETD2 N-term
1 µM
0.1 µM
SETD2 C-term
1 µM
0.1 µM
95°C, 10':
– + – + – + – + – +
(i–v)
IB: α-SETD2 C-term
250
130
95
72
55
36
p290C
*
p140C
p66C
p45C
1 2 3 4 5 6 7 8 9 10
(i)
(ii)
(iii)
(iv)
(v)
(vi)
Ponceau S
250
130
95
72
55
36
1 2 3 4 5 6 7 8 9 10

3.3.4. Characterisation of the minimal conditions required for p66^{C} detection

Optimal immunodetection of p66^{C} only occurred after heating at 95°C in the presence of urea for at least 10 minutes (**Figure 3.6A-i**). A subtle but progressive decrease in electrophoretic mobility was also observed for p66^{C} with an increase in the duration of heating (**Figure 3.6A-i**). Immunodetection of p66^{C} depended reciprocally on the length of time of heating at 95°C and the concentration of urea (**Figure 3.6B**), such that at lower concentrations of urea, longer duration of heating at 95°C was required for comparable signal intensity, *e.g.* with 0.25 M urea, heating for 20 minutes was required to obtain intensity comparable to heating for 10 minutes in 2 M urea (**Figure 3.6B**, compare lane 9 in panel **ii** with lane 7 in panel **v**). Strikingly, p66^{C} was not observed after a similar duration of heating at 70°C – even after heating for 20 minutes in the presence of 4 M urea (**Figure 3.6B-vi**, lane 5).

At the lowest concentration of urea used here (0.25 M), the intensity of the appearance of p66^{C} after ≈15 minutes at 95°C did not correlate with a comparable decrease in intensity of higher molecular weight species (**Figure 3.6C**). Furthermore, loss of virtually all C-terminal immunoforms was observed after heating for 20 minutes in the absence of urea without a corresponding increase in p66^{C} (**Figure 3.6D**, lane 18). One could argue that failure to observe p66^{C} under the latter condition may be due to a requirement of urea to subsequently expose an epitope of a species that is generated initially by thermal denaturation, but which remains conformationally inaccessible in the absence of urea. However, when the sequential order of heat and urea-mediated denaturation was varied, p66^{C} was only detected when extracts were heated at 95°C concomitant with the presence of urea: thermal denaturation prior to urea addition was insufficient for detection (**Figure A.2** in Appendix A).

Whether p66^{C} exists in an aggregated form in the absence of thermo-urea denaturation is unclear, but p66^{C} remained soluble after ultracentrifugation of the TX−DOC extract, which argues against the incorporation of p66^{C} into highly-complexed material in TX−DOC buffer (**Figure A.3** in Appendix A).

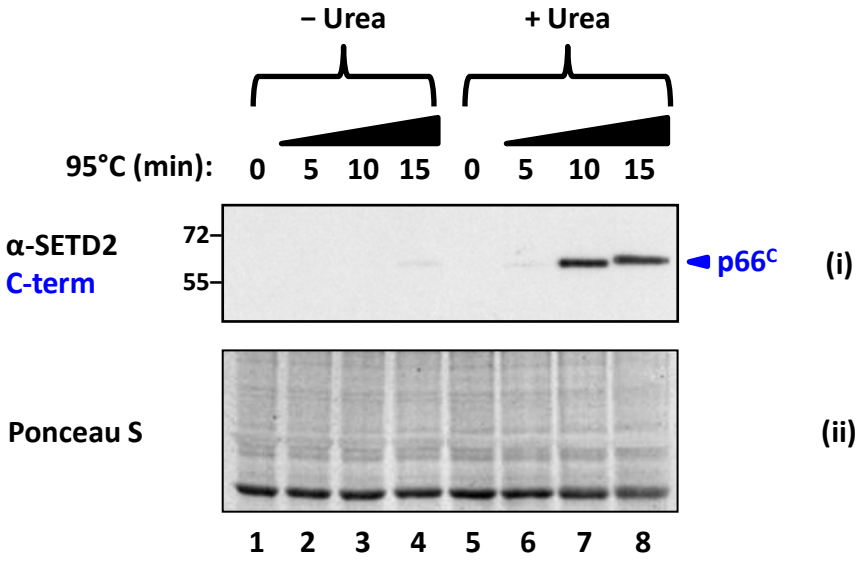
Figure 3.6 Immunodetection of p66^{C} is influenced by [urea], temperature and duration of heating

(A) Optimal p66^{C} detection requires heating at 95°C in the presence of urea: Samples represent a p66^{C}-enriched, high-salt extract described in **Figure 4.1**, prepared from quiesced 10T½ cells. TCA-precipitated extract was resuspended in SDS-PAGE loading buffer in the absence (lanes 1–4) or presence of 4 M urea (lanes 5–8). Samples were incubated at 95°C for the indicated times, before separation by 8% SDS-PAGE. After Ponceau S–staining **(ii)**, the membrane was probed with α-SETD2 C-terminal antibody **(i)**.

(B) Extent of p66^{C} detection is influenced by a reciprocal relationship between [urea] and duration of heating at 95°C: Identical aliquots of TX–DOC extract from quiesced 10T½ cells were adjusted to SDS-PAGE loading buffer supplemented with the indicated concentrations of urea (0–4 M) and incubated on ice (lane 1), at 70°C (lanes 2–5) or at 95°C (lanes 6–9) for the indicated times, before analysis by 8% SDS-PAGE. Membranes were stained with Ponceau S (right panel) and immunoblotted with α-SETD2 C-terminal antibody (left panel)

(C & D) *Continued on page 104.*

(A)


– Urea
+ Urea
95°C (min):
0 5 10 15 0 5 10 15
α-SETD2
C-term
72
55
p66C
(i)
Ponceau S
(ii)
1 2 3 4 5 6 7 8

(B)

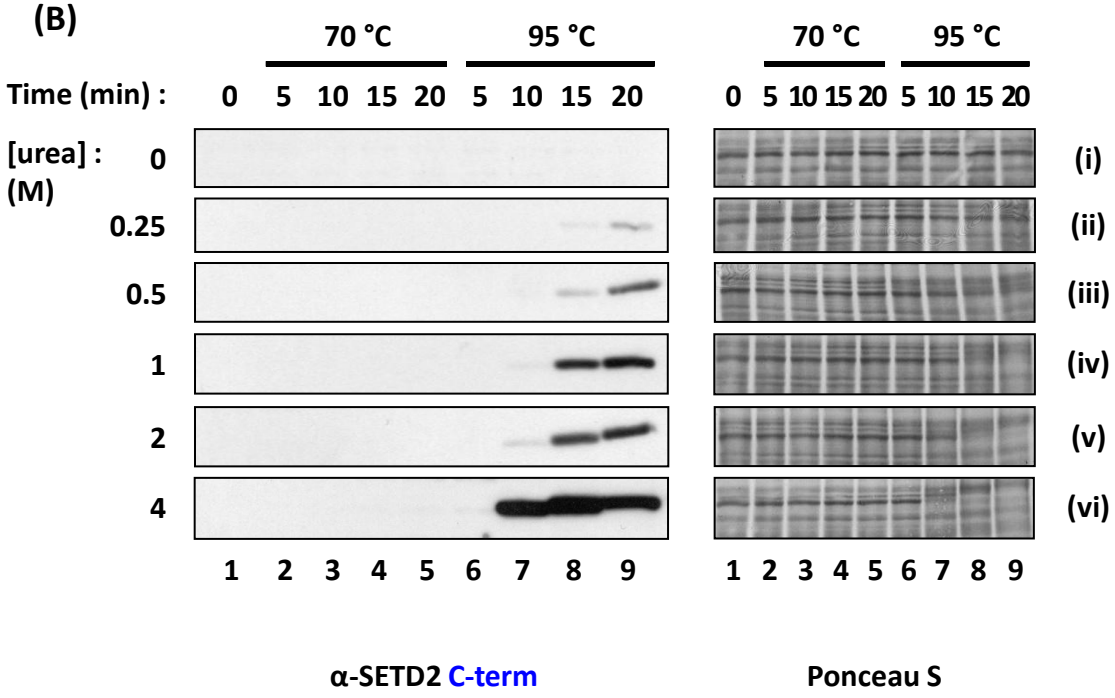
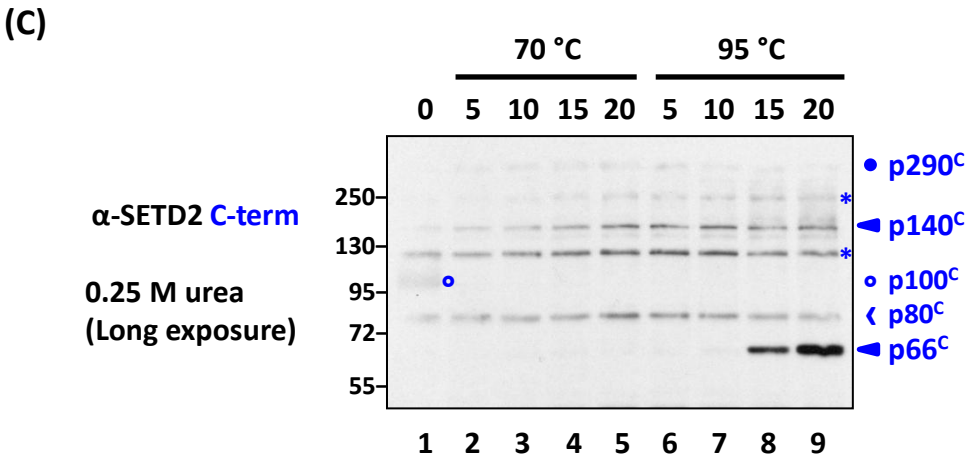
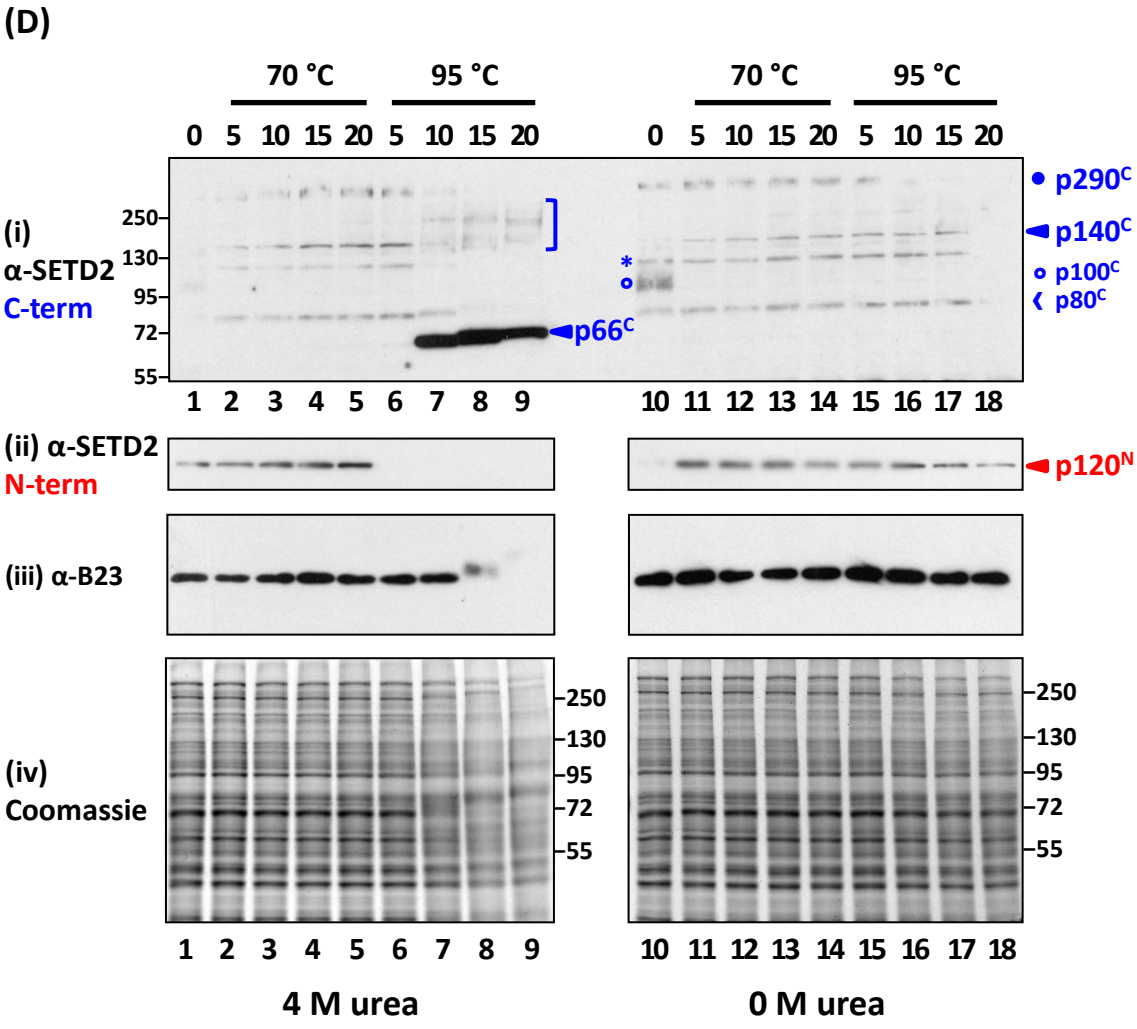

70 °C
95 °C
70 °C
95 °C
Time (min) :
0 5 10 15 20 5 10 15 20
0 5 10 15 20 5 10 15 20
[urea] : (M)
0
0.25
0.5
1
2
4
(i)
(ii)
(iii)
(iv)
(v)
(vi)
1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9
α-SETD2 C-term
Ponceau S

Figure 3.6 Immunodetection of p66^{C} is influenced by [urea], temperature and duration of heating *(continued)*

(C) Detection of p66^{C} after thermo-urea denaturation does not correlate with loss of larger C-terminal immunoforms: Analysis of α-SETD2 C-terminal signals from extracts containing 0.25 M urea, processed according to **(B)**. Immunoblot represents a longer exposure of the entire membrane featured partially in panel **(ii)** of **(B)** previously. The bands marked with asterisks are not reproducible between independent biological experiments.

(D) Protein-specific responses to thermal denaturation in the presence or absence of 4 M urea: Extracts processed according to **(B)** in the presence of 4 M urea (lanes 1–9) or its absence (lanes 10–18). Samples were separated by 8% SDS-PAGE and either transferred to PVDF in Towbin buffer **(i–iii)** or stained with Coomassie **(iv)**. Membranes were probed with α-SETD2 C-terminal **(i)**, α-SETD2 N-terminal **(ii)** or α-B23 (nucleophosmin) **(iii)** antibodies. Immunoblot **(i)** represents an over-exposure of the entire membrane featured partially in panels **(i)** and **(vi)** of **(B)** previously.


(C)
70 °C
95 °C
0 5 10 15 20 5 10 15 20
α-SETD2 C-term
0.25 M urea
(Long exposure)
250
130
95
72
55
p290C
p140C
p100C
p80C
p66C
1 2 3 4 5 6 7 8 9


(D)
70 °C
95 °C
70 °C
95 °C
0 5 10 15 20 5 10 15 20 0 5 10 15 20 5 10 15 20
(i)
α-SETD2
C-term
250
130
95
72
55
p66C
p290C
p140C
p100C
p80C
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
(ii) α-SETD2
N-term
p120N
(iii) α-B23
(iv)
Coomassie
250
130
95
72
55
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
4 M urea
0 M urea

For all the remaining experiments in this thesis, analysis of p66^{C} was performed by heating aliquots of the sample, leaving sufficient unheated sample to separately analyse other α-SETD2-immunoreactive bands, because of their susceptibility to degradation or epitope masking and complications arising from urea-mediated carbamylation.

Pertinent to detection of other putative isoforms, immunodetection of p120^{N} was extremely sensitive to thermo-urea denaturation conditions: virtually all of the signal disappeared after 5 minutes at 95°C with 4 M urea (**Figure 3.6D-ii** and **Figure A.1** in Appendix A). Thus thermo-urea denaturation was avoided for analysis of p290^{N} and p120^{N} proteins.

In this experiment, an additional p100^{C} protein was also detected in the absence of thermal denaturation (**Figure 3.6D-i**). This band was also observed in human MRC5 and 293T cells only in the absence of thermal denaturation, suggesting conservation of the epitope and its properties (**Figure 3.7C**). Its providence is unclear: its appearance did not correlate with degradation of higher molecular weight bands, and it may represent a thermo-labile protein or an epitope that is susceptible to modification during thermal denaturation. The presence of this protein was only appreciated in the late stages of this work, and should be re-evaluated in future studies.

3.3.5. Conservation of α-SETD2-reactive proteins in multiple murine and human cell types

Since the C- and N-terminal SETD2 peptides are 100% and 95% identical, respectively, between humans and mice, similar proteins should theoretically be observed when probing similar human cells. Two types of human cells were analysed for SETD2 immunoforms: MRC5 lung fibroblasts and HEK 293T cells. Whilst the former represents a similar cell type to 10T½ cells, the exact cellular origin of the latter is unclear as 293T cells were derived from transformation of disaggregated kidney tissue with sheared adenovirus DNA (Shaw *et al.*, 2002). *Drosophila* S2 cells were analysed in parallel as a negative control since the *Drosophila* Set2 homolog shares no sequence similarity with the antibody epitopes in the mammalian proteins. The putative SETD2 isoforms detected in 10T½ cells would therefore be predicted to be absent from *Drosophila* S2 extracts. Given the lack of sequence

conservation, observation of a protein displaying similar properties to p66^{C} in *Drosophila* would immediately refute the assertion that p66^{C} could be a *bona fide* SETD2 isoform.

Similar N-terminal reactive species were observed in extracts from human cells but not *Drosophila* S2 cells (**Figure 3.7A**). MRC5 cells appeared to express low levels of p290^{N} whilst 293T cells expressed p250^{N} and p290^{N} more strongly than p120^{N} in comparison to the other mammalian cell types (**Figure 3.7A**).

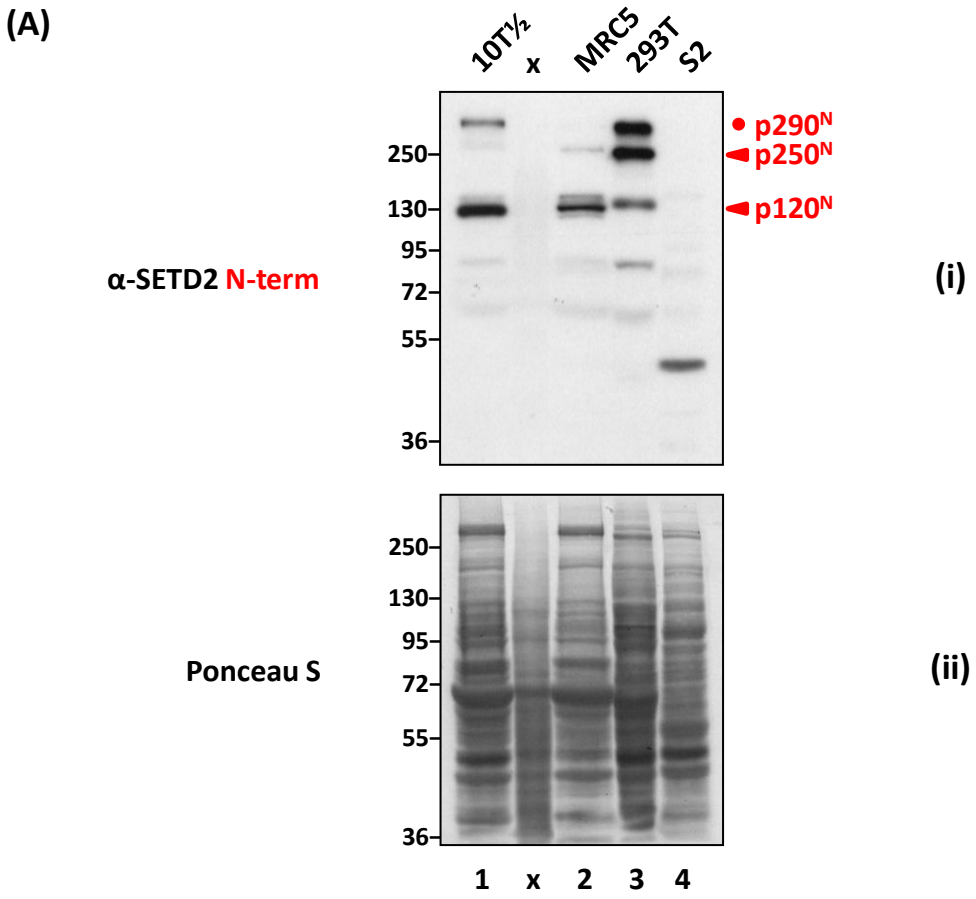
To evaluate the conservation of p66^{C} and C-terminal-reactive proteins, extracts from the four cell types were each solubilised in four different conditions: with/without adjustment to 4 M urea; and incubated at room temperature, or heated at 95°C for 10 minutes. Both human extracts, but not *Drosophila* extracts, contained a protein similar to p66^{C} and the requirement of thermo-urea denaturation for its immunodetection was conserved (**Figure 3.7C**, lanes 4, 8 and 12). The requirement of thermo-urea denaturation for p66^{C} detection by the C-terminal antibody was also observed in murine embryonic stem cells (**Figure 3.7B**, lane 4). Human cells contained p140^{C} shared with 10T½, but also an exclusive p35^{C} species, undetectable in murine fibroblasts (**Figure 3.7C-iii**, lanes 5–12).

Figure 3.7 Conservation of SETD2 immunoforms in multiple mammalian cell types

(A) Analysis of α-SETD2 N-terminal immunoforms in human and *Drosophila* cells: TX–DOC extracts were prepared from quiesced murine 10T½ fibroblasts (lane 1), human MRC5 fibroblasts (lane 2), human HEK 293T cells (lane 3) and *Drosophila* S2 cells (lane 4). Samples were separated by 8% SDS-PAGE and transferred to PVDF in Towbin buffer. Membrane was stained with Ponceau S **(ii)** and probed with α-SETD2 N-terminal antibody **(i)**. The lane marked *x* should be ignored—it represents murine ESC extract that degraded under acidic storage conditions (*cf.* Ponceau S–stain of intact extract in lane 1 of **[**B**]** below).

(B) Conservation of α-SETD2 C-terminal immunoforms in murine ESCs: TX–DOC extracts were prepared from murine ESCs (*CGBP*$^{-/-}$; lanes 1–4) and quiesced 10T½ fibroblasts (lanes 5–8). Samples were precipitated with TCA, resolubilised in urea/SDS-PAGE buffer, and incubated at room temperature (odd lanes marked with "–") or 95°C for 10 minutes (even lanes marked with "+"). Following separation by 8% SDS-PAGE, membranes were stained with Ponceau S (lanes 3, 4, 7 and 8), before probing with α-SETD2 C-terminal antibody (lanes 1, 2, 5 and 6). The asterisk represents an immunoreactive band apparently specific to mESCs.

(C) *Continued on page 110.*


(A)
10T½
x
MRC5
293T
S2
p290N
p250N
p120N
250
130
95
72
55
36
α-SETD2 N-term
(i)
Ponceau S
(ii)
1 x 2 3 4

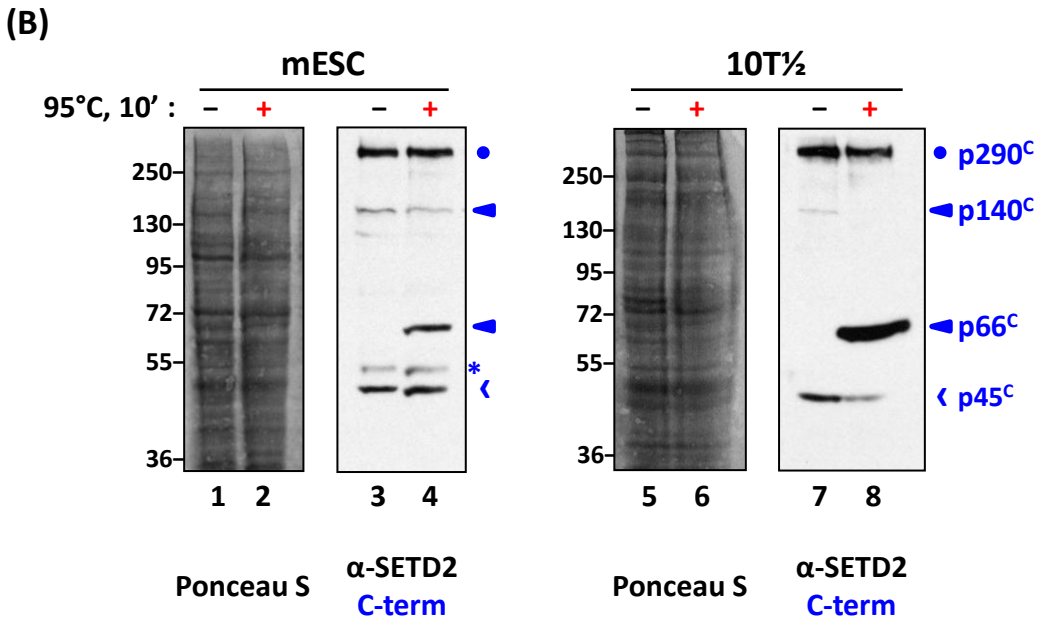
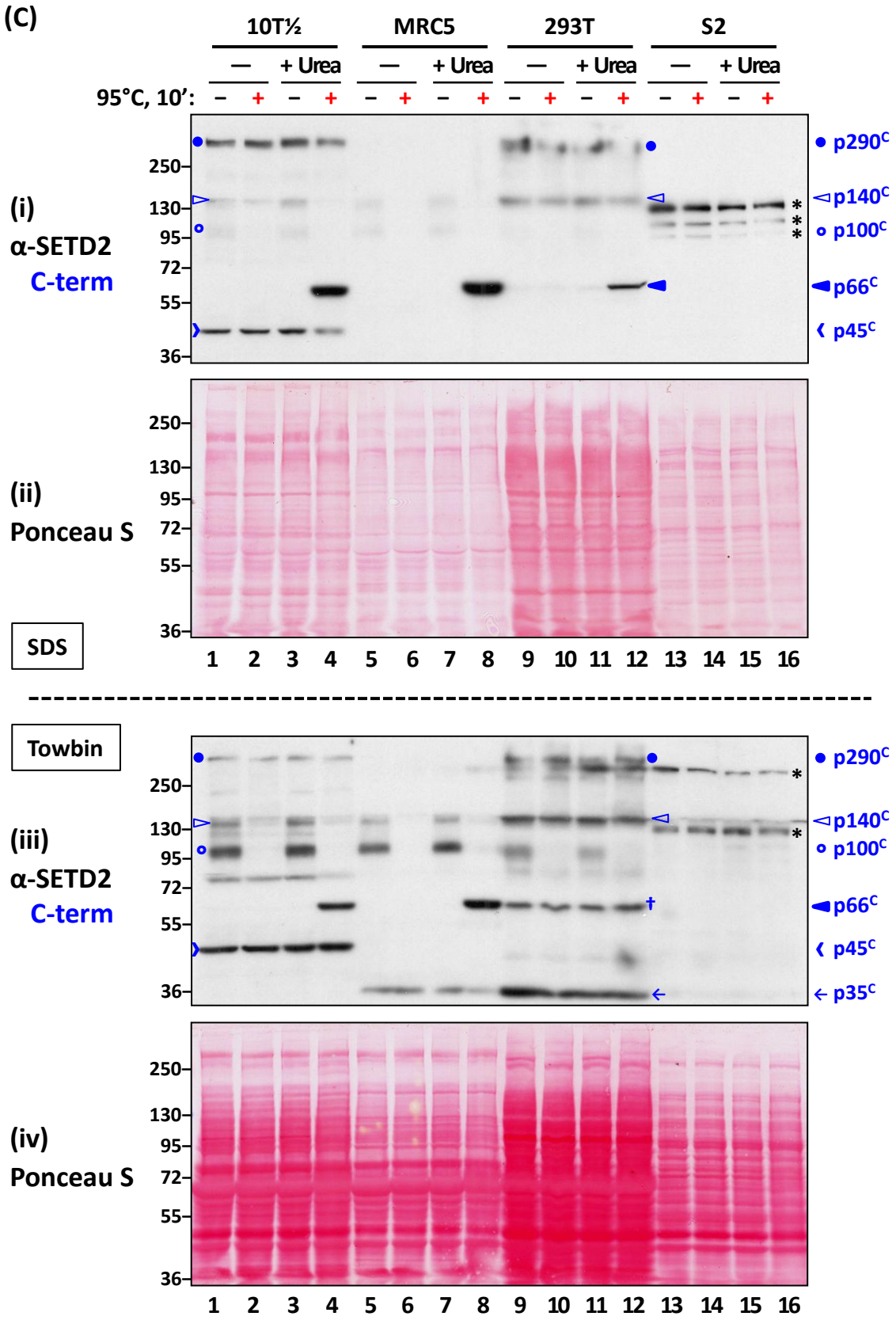

(B)
mESC
10T½
95°C, 10' : – + – + – + – +
p290C
p140C
p66C
p45C
250
130
95
72
55
36
1 2 3 4 5 6 7 8
Ponceau S
α-SETD2
C-term
Ponceau S
α-SETD2
C-term

Figure 3.7 Conservation of SETD2 immunoforms in multiple mammalian cell types *(continued)*

(C) Conservation of C-terminal SETD2 immunoforms and detection characteristics of p66^{C} in human cells: Immunoblot analysis of the effect of 4 M urea and heating on detection of α-SETD2 C-terminal immunoforms in human and *Drosophila* cells. TX–DOC extracts prepared from quiesced murine 10T½ fibroblasts (lanes 1–4), human MRC5 fibroblasts (lanes 5–8), human HEK 293T cells (lanes 9–12), and *Drosophila* S2 cells (lanes 13–16) were adjusted to 1× SDS-PAGE loading buffer in the absence of urea (lanes 1, 2, 5, 6, 9, 10, 13 and 14) or presence of 4 M urea (lanes 3, 4, 7, 8, 11, 12, 15 and 16). Samples were either incubated at room temperature (odd lanes marked with "−") or at 95°C for 10 minutes (even lanes marked with "+") before separation by 8% SDS-PAGE and transfer in either SDS **(i, ii)** or Towbin **(iii, iv)** transfer buffer. Membranes were stained with Ponceau S to assess loading **(ii, iv)** before probing with α-SETD2 C-terminal antibody **(i, iii)**. Asterisks indicate *Drosophila* proteins unrelated to SETD2. The dagger symbol (†) indicates a protein of similar molecular mass to p66^{C} in 293T cells, which appears insensitive to thermal-urea denaturation


(C)
10T½
MRC5
293T
S2
— + Urea
95°C, 10': – + – + – + – + – + – + – + – +
(i)
α-SETD2
C-term
250
130
95
72
55
36
p290C
p140C
p100C
p66C
p45C
(ii)
Ponceau S
SDS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Towbin
(iii)
α-SETD2
C-term
p35C
(iv)
Ponceau S

3.3.6. Transient siRNA transfection reduces p290$^{N/C}$ expression exclusively

Transient siRNA transfection was performed to evaluate the identity of these α-SETD2-reactive proteins. This approach has been successfully used previously to confirm a ≈290 kDa protein detected with crude α-SETD2 sera as the SETD2 protein, responsible for H3K36 trimethylation (Edmunds *et al.*, 2008). Unsurprisingly, this protein is equivalent to p290$^{N/C}$ detected with affinity-purified antibodies (**Figure 3.8**). RNAi analysis was re-evaluated and extended here for the following reasons: (i) the denaturation and SDS-transfer conditions used previously did not permit detection of p120^{N} and p66^{C}; (ii) potential knockdown of proteins detected by affinity-purified antibodies may have been obscured by non-specific detection of unrelated co-migrating proteins by crude antisera; and (iii) acquisition of an additional, commercial α-SETD2 CT21 antibody. Note that results pertaining to this antibody will be discussed later in Chapters 4 and 6.

10T½ cells were transiently transfected with non-targeting siRNA (unrelated to *Setd2* sequence) or siRNAs specific for two 3′ regions of *Setd2*: Setd2(0) targets exon 13 whilst Setd2(3) targets exon 21. TX−DOC extracts were prepared 48 hours after transfection and analysed by immunoblotting under either standard SDS-PAGE denaturation conditions (**Figure 3.8A/C/D**) or after thermo-urea denaturation for p66^{C} analysis (**Figure 3.8B**). Depletion of p290^{N} and p290^{C} occurred but with little reproducible effect on the levels of other immunoreactive forms. The p120^{N} protein was unaffected by transient transfection of 3′ region siRNAs compared to mock and non-targeting controls (**Figure 3.8A-iii**). Transfection of siRNAs targeting exon 3 of *Setd2* only reduced the level of p290^{SETD2} (**Figure 3.8D-i/iv**) with no apparent effect on the levels of p120^{N} or p140^{C}. Minor variations in levels of p120^{N} and p140^{C} correlated with subtle loading variations as judged by Ponceau-staining (**Figure 3.8A**, compare: [**i** versus **ii]** and [**iii** versus. **iv]**; **Figure 3.8D**, compare **ii** versus **iii**) and were not reproducibly observed (**Figure 3.8E**). Thus in summary, under conditions where p290^{SETD2} depletion acts as a positive control for the effectiveness of the *Setd2* siRNA transfection, no other α-SETD2-reactive band was substantially or reproducibly depleted (**Figure 3.8E**).

Figure 3.8 Transient transfection of siRNAs targeting 5' and 3' regions of *Setd2* only depletes p290$^{N/C}$

(A) siRNAs targeting 3′ regions of *Setd2* deplete p290$^{N/C}$ but not p120^{N} or p140^{C}: 10T½ cells were mock-transfected without siRNA (lane 1), or transfected with 50 nM siRNAs: Setd2(0) against exon 13 (lane 2), Setd2(3) against exon 21 (lane 3), or non-targeting control (lane 4). TX−DOC extracts were prepared 48 hours after transfection and separated by conventional SDS-PAGE on large SDS gels. Gels were transferred to PVDF in SDS **(i, ii)** or Towbin **(iii, iv)** transfer buffer. Loading levels were visualised by Ponceau S **(ii**, **iv)** before probing with α-SETD2 C-terminal **(i)** or N-terminal **(iii)** antibody.

(B) siRNAs targeting 3′ regions of *Setd2* do not deplete p66^{C}: Samples were prepared according to **(A)** but precipitated with TCA, solubilised in urea/SDS-PAGE loading buffer and incubated at 95°C for 15 minutes before electrophoresis.

(C) Depletion of H3K36me3 by *Setd2* siRNAs: Acid extracts were prepared from TX-DOC-resistant pellets from experiment **(A)** and resolved by 15% SDS-PAGE for Coomassie-staining to assess total histone levels, or for immunoblotting with α-H3K36me3 or α-H3K27me3 antibodies.

(D & E) *Continued on page 115.*

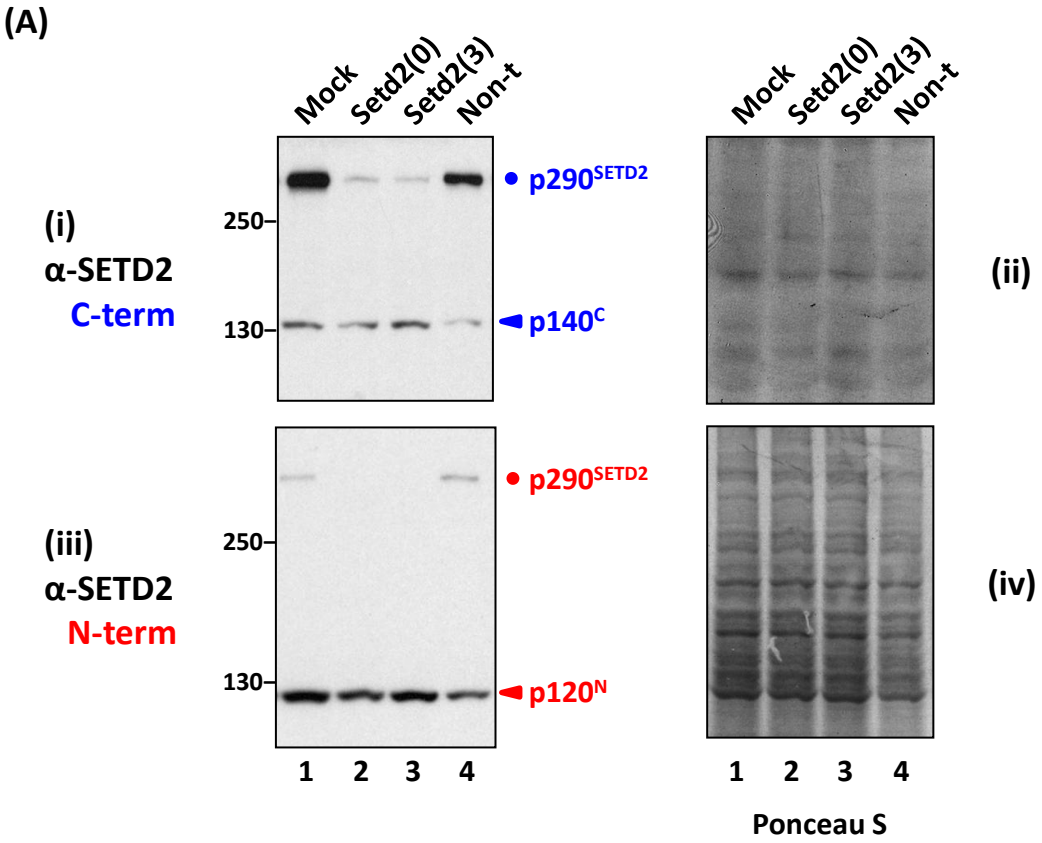
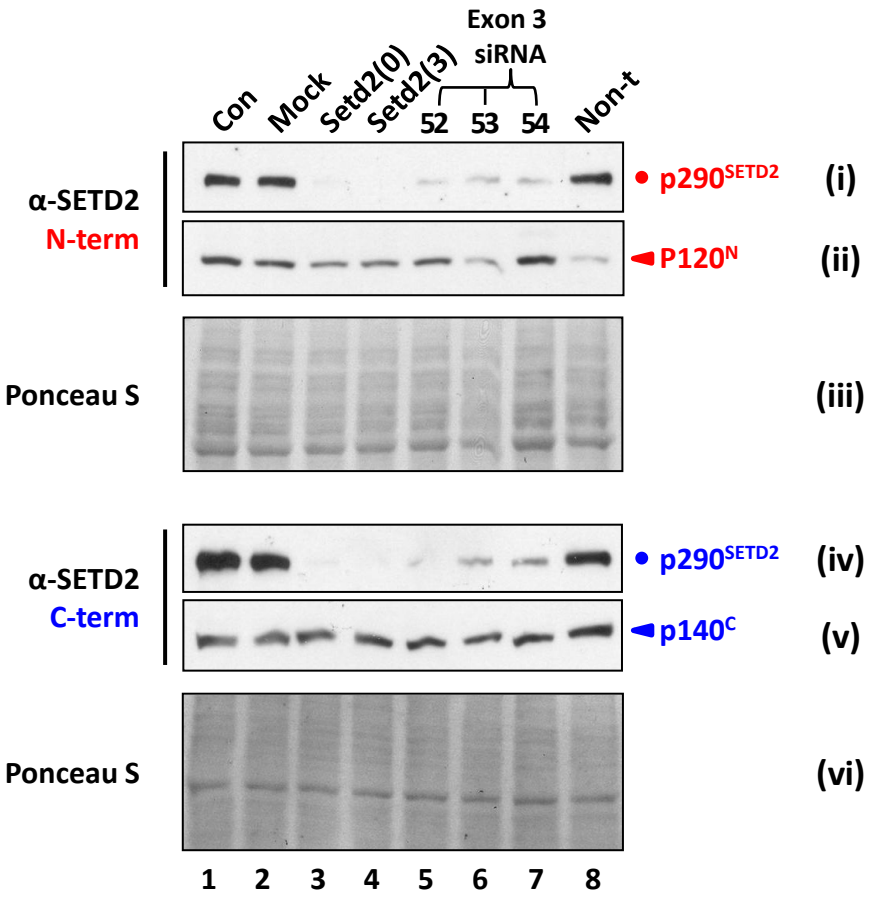

(A)
Mock
Setd2(0)
Setd2(3)
Non-t
(i)
α-SETD2
C-term
250
130
p290SETD2
p140C
(ii)
(iii)
α-SETD2
N-term
p120N
(iv)
1 2 3 4
Ponceau S
(B)
α-SETD2
C-term
p66C
Ponceau S
(C)
α-H3K36me3
α-H3K27me3
Coomassie

Figure 3.8 Transient transfection of siRNAs targeting 5' and 3' regions of *Setd2* only depletes p290$^{N/C}$ *(continued)*

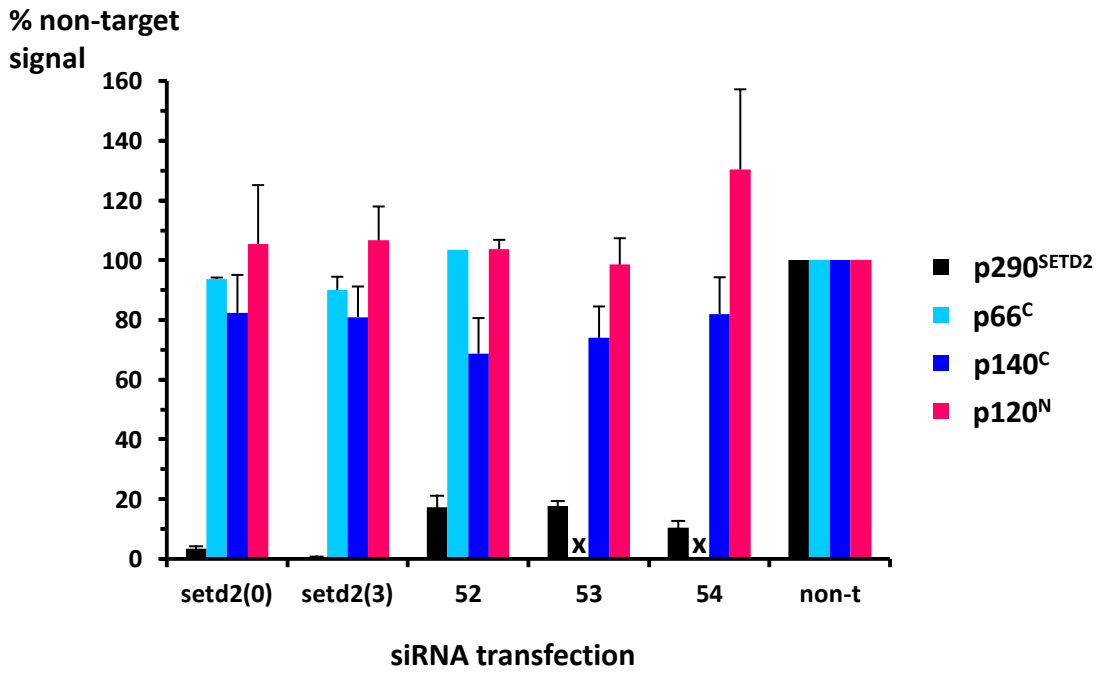
(D) Transfection of 5′ region siRNAs that target exon 3 of *Setd2* reduces p290^{SETD2} but not p120^{N} or p140^{C}: 10T½ cells were left untreated (lane 1), mock-transfected without siRNA (lane 2), or transfected with 50 nM siRNAs: Setd2(0) against exon 13 (lane 3), Setd2(3) against exon 21 (lane 4), and siRNAs against three different regions of exon 3 (lanes 5, 6 and 7) or non-targeting control (lane 8). TX–DOC extracts were prepared 48 hours after transfection and separated by 8% SDS-PAGE. Gels were transferred to PVDF in Towbin **(i–iii)** or SDS **(iv–vi)** transfer buffer. Loading levels were visualised by Ponceau S before probing membrane **(iii)** with α-SETD2 N-terminal antibody **(i**, **ii)** and membrane **(vi)** with C-terminal antibody **(iv, v)**.

(E) Quantified summary of siRNA experiments: Immunoblot signals were quantified by ImageJ and are presented as mean percentage signal relative to non-target controls. Error bars represent standard error of the mean ($n \geq 3$). The 'x' symbols indicate that p66^{C} levels were not analysed.

(D)


Exon 3 siRNA
Con
Mock
Setd2(0)
Setd2(3)
52
53
54
Non-t
α-SETD2 N-term
p290SETD2
(i)
P120N
(ii)
Ponceau S
(iii)
α-SETD2 C-term
p290SETD2
(iv)
p140C
(v)
Ponceau S
(vi)
1 2 3 4 5 6 7 8

(E)


% non-target signal
160
140
120
100
80
60
40
20
0
setd2(0)
setd2(3)
52
53
54
non-t
siRNA transfection
p290SETD2
p66C
p140C
p120N

3.3.7. p290^{SETD2} is relatively labile compared to additional immunoreactive proteins

Since protein depletion by RNAi depends crucially on the rate of protein turnover, the aforementioned RNAi results need to be interpreted within the context of the endogenous stability of these proteins. Protein turnover was therefore investigated by treating 10T½ cells with emetine to inhibit translation. Emetine was selected due to its minimal cross-activation of stress signalling pathways in 10T½ cells, compared to cycloheximide, puromycin or anisomycin (Edwards & Mahadevan, 1992).

The p290^{SETD2} signal detected with both N-terminal and C-terminal antibodies decreased with increasing duration of emetine treatment (**Figures 3.9A-i/iii** and **3.9B-i/iii**), suggesting that in the absence of protein synthesis, p290^{SETD2} was being rapidly degraded. The half-life of p290^{SETD2} was estimated as 3.3 ± 1.6 hours (mean (*M)* ± standard deviation (*s*); $n = 6$).

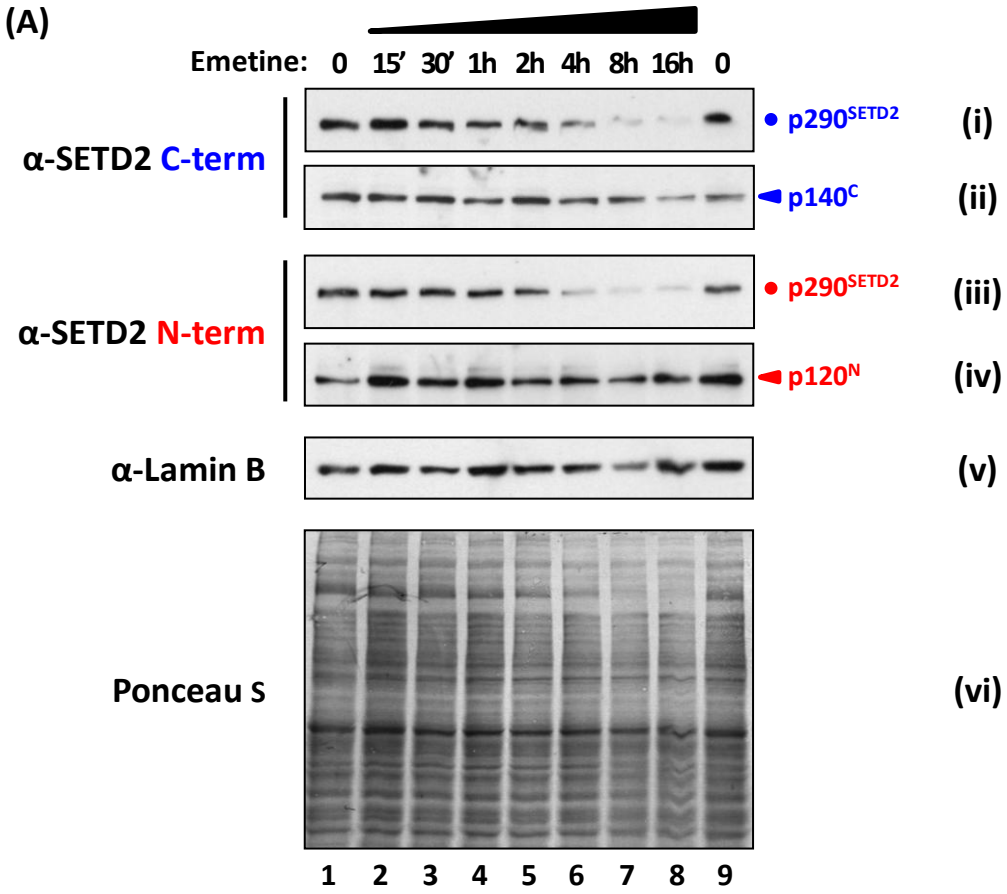
The decrease in p290^{SETD2} intensity could result from: (i) protein degradation; (ii) loss of epitope reactivity, independent of total protein level; or (iii) redistribution of p290^{SETD2} into the TX-DOC-insoluble fraction. The second explanation is unlikely since both p290^{N} and p290^{C} signals responded similarly. The final interpretation was excluded, since virtually no p290^{SETD2} signal could be detected in fractions resulting from further extraction of TX–DOC-resistant material with urea- and SDS/urea-containing buffers (**Figure A.3** in Appendix A). Given emetine predominantly inhibits translation, the first explanation is the most likely rationale, though emetine-induced changes to epitope reactivity towards select proteins cannot be formally excluded. An alternative, but not necessarily mutually-exclusive, interpretation of this experiment is that, independently of (or additionally to) inhibition of protein synthesis, emetine itself stimulated selective degradation of p290^{SETD2} compared to the other putative isoforms. Overall, however, the simplest explanation is that p290^{SETD2} is relatively unstable, being continuously synthesised and degraded—consistent with elimination of virtually all p290^{SETD2} after only 48 hours of RNAi (**Figure 3.8**).

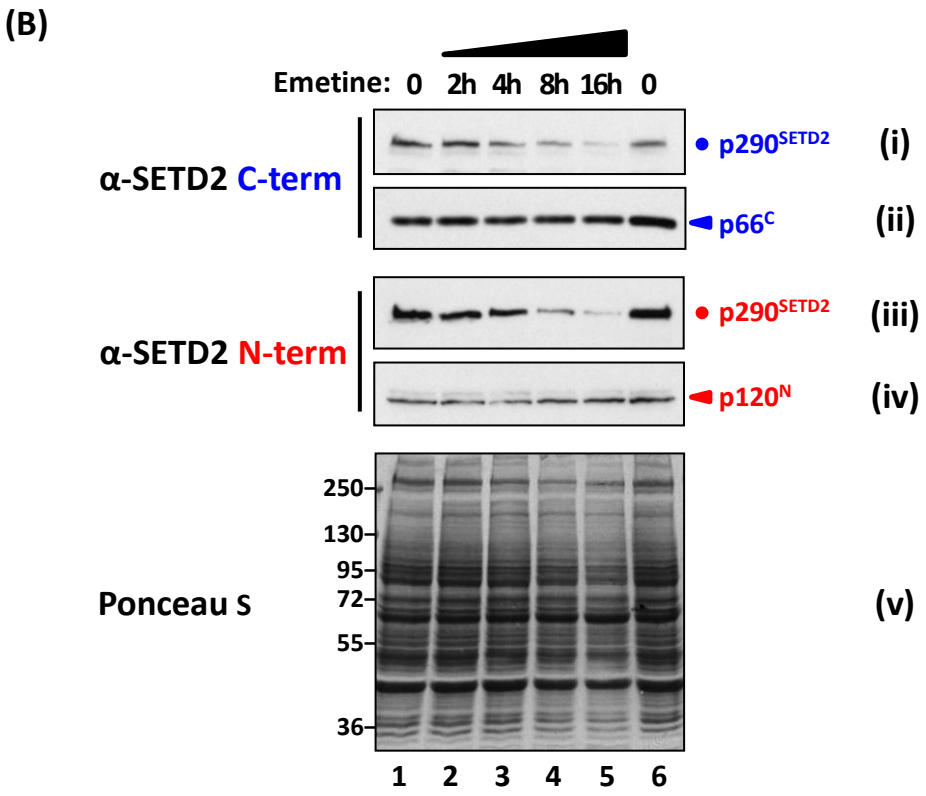
The levels of p120^{N}, p140^{C} and p66^{C}, however, were relatively more stable compared to p290^{SETD2}: similar levels of these proteins existed after 8 and 16 hours of translation inhibition compared to untreated cells (**Figures 3.9A-ii/iv** and **3.9B-ii/iv**).

Figure 3.9 p290^{SETD2} is relatively unstable compared to p120^{N}, p140^{C} and p66^{C}

(A) Protein synthesis inhibition reveals greater instability of p290^{SETD2} compared to p120^{N} and p140^{C}: Quiesced 10T½ cells were left untreated (lanes 1 and 9: cells harvested after 26 or 34 hours of quiescence, respectively) or treated with 10 μg/ml emetine for the indicated times (lanes 2–8: 15 or 30 minutes, or 1–16 hours) before preparing soluble TX–DOC extracts. Samples were analysed by standard 8% SDS-PAGE and transferred in SDS **(i, ii)** or Towbin **(iii–vi)** buffer for subsequent immunoblotting with the indicated antibodies. Ponceau S–stained membrane **(vi)** was probed first with α-SETD2 N-terminal antibody **(iii, iv)** before stripping and reprobing with α-Lamin B antibody **(v)**.

(B) p66^{C} is relatively stable after 16 hours of protein synthesis inhibition: Independent experiment performed as in **(A)** with differences in duration of emetine treatment indicated. Samples for analysis by α-SETD2 C-terminal antibody **(i, ii)** were solubilised in urea/SDS-PAGE buffer and denatured at 95°C for 10 minutes, before separation by 8% SDS-PAGE and western transfer in SDS buffer. Samples for analysis by α-SETD2 N-terminal antibody **(iii, iv)** were analysed as in **(A)** by standard SDS-PAGE conditions and transferred to PVDF in Towbin buffer.


(A)
Emetine: 0 15' 30' 1h 2h 4h 8h 16h 0
α-SETD2 C-term
p290SETD2
(i)
p140C
(ii)
α-SETD2 N-term
p290SETD2
(iii)
p120N
(iv)
α-Lamin B
(v)
Ponceau S
(vi)
1 2 3 4 5 6 7 8 9


(B)
Emetine: 0 2h 4h 8h 16h 0
α-SETD2 C-term
p290SETD2
(i)
p66C
(ii)
α-SETD2 N-term
p290SETD2
(iii)
p120N
(iv)
Ponceau S
250
130
95
72
55
36
(v)
1 2 3 4 5 6

3.4. Discussion

3.4.1. Summary of results

The major objective of this chapter was to establish detection characteristics of putative SETD2 isoforms in murine 10T½ cells by affinity-purified α-SETD2 antibodies, and evaluate their identity by RNAi. The results are summarised as follows:

- Both affinity-purified α-SETD2 antibodies to N-terminal and C-terminal regions specifically detected p290 protein in murine 10T½ fibroblasts (**Figures 3.1, 3.3** and **3.5**). p290 was confirmed as SETD2 by depletion of both p290^{N} and p290^{C} after transient transfection of five different siRNAs, targeting 5′ and 3′ regions of *Setd2* (**Figure 3.8A-i/iii**, **B** and **D-i/iv**). The effectiveness of knockdown was consistent with p290^{SETD2} being relatively labile (half-life of ≈3–4 hours) (**Figure 3.9A-i/iii** and **B-i/iii**).

- The antibodies specifically detected additional proteins of lower molecular mass. The most prominent and reproducibly detected species were an N-terminal-specific p120^{N} protein (**Figure 3.2**) and C-terminal-specific p140^{C} and p66^{C} proteins (**Figure 3.5**).

- p120^{N}, p140^{C} and p66^{C} were not depleted by RNAi (**Figure 3.8E**). These proteins exhibited relatively slow turnover (stable for at least 16 hours) (**Figure 3.9A-ii/iv** and **B-ii/iv**), suggesting that transient siRNA transfection is not an appropriate technique to assess their identity.

- Detection of p66^{C} by the C-terminal antibody in multiple murine and human cell lines required concomitant heating at 95°C in the presence of urea (**Figures 3.4-i**, **3.6**, and **3.7**-**B/C**).

- Similar p290$^{N/C}$, p120^{N}, p140^{C} and p66^{C} forms were observed in mESCs and human cells, although p290^{SETD2} appeared to be absent from MRC5 cells (**Figure 3.7**).

3.4.2. Detection of putative SETD2 isoforms in addition to p290^{SETD2}

Affinity-purified antibodies detected p290^{SETD2} along with additional putative isoforms that are summarised in **Table 3.2**. The p290^{SETD2} protein corresponds to the histone H3K36 methyltransferase of ≈286 kDa predicted from the reference SETD2 model in current protein databases and its immunodetection was verified by RNAi. As discussed in section 1.6.2, additional SETD2 isoforms have been detected by mass spectrometry and immunoblotting. No proteins that migrated greater than 250 kDa were detected, but several smaller isoforms were consistently detected in both murine and human cells. With limited information available from previous reports, the relationship of the putative isoforms identified here with those previously described is unclear. The p130 species detected by mass spectrometry may relate to p140^{C} or p120^{N} (Fujiki *et al.*, 2009). The p35^{C} form observed exclusively in human cells may be related to a ≈33 KDa isoform detected in human platelets by MS (Garcia *et al.*, 2004), though this is difficult to confirm without peptide sequence data. No p66^{C}-like protein has been reported previously. Speculation of function of these putative isoforms is challenging with only the knowledge that they contain a specific N- or C-terminal epitope, and that alternative splicing has the potential to generate proteins with widely different domain composition.

Although peptide competition demonstrates that all of the putative SETD2 isoforms are specifically detected in mouse fibroblasts, this technique does not prove that the immunoforms actually contain SETD2 sequence. The nature of antigen–antibody interactions may permit the complementarity-determining regions to recognise chemically similar epitopes with differences in affinity. Furthermore, mimotopes—chemically dissimilar structures that nevertheless recapitulate a similar epitope configuration or conformation—may bind with equivalent affinity to paratopes intended to be reactive against a chemically distinct peptide. Consequentially, the detection of non-SETD2 proteins containing mimotopes would also be specifically eliminated by peptide competition. This form of cross-reactivity—inherent to the nature of antigen–antibody interactions—cannot be avoided by peptide affinity-purification; and its extent is hard to predict *a priori*. Thus without independent corroboration of identity (for example, by genetics or protein

Table 3.2 – Summary of mammalian α-SETD2 immunoforms detected

Protein	Antibody reactivity	Detected in				Detection comments
		Mouse		Human		
		10T½	ESC	MRC5	293T	
Major immunoforms						
p290^{SETD2}	**C-terminal**	**+**	**+**	**–**	**++**	
	N-terminal	**+**	nd	**–**	**++**	***avoid*** thermo-urea denaturation
p120^{N}	**N-terminal**	**++**	nd	**+**	**+**	***requires*** Towbin transfer ***avoid*** thermo-urea denaturation
p140^{C}	**C-terminal**	**+**	**+**	**+**	**+**	
p66^{C}	**C-terminal**	**++**	**++**	**++**	**+**	***requires*** thermo-urea denaturation
	CT21 *†	**+***	**+†**	**+†**	nd	
Minor immunoforms						
p250^{N}	N-terminal	(+/–)	nd	(+)	**++**	Human band equivalence unclear
p125^{N}	N-terminal	(+/–)	nd	(+)	(+/–)	***avoid*** thermo-urea denaturation
p100^{C}	C-terminal	(+/–)	nd	(+/–)	(+/–)	***avoid*** thermo-urea denaturation
p80^{C}	C-terminal	(+/–)	nd	–	–	
p45^{C}	C-terminal	(+/–)	**+**	–	–	*poor reproducibility*
p35^{C}	C-terminal	**–**	**+**	**+**	**+**	***requires*** Towbin transfer (Human)

Key: Strongly detected [++], detected [+], weakly detected [(+)], variably detected[(+/–)], not detected [–], not determined [nd].

* data presented in Chapter 4.

† data not shown.

Molecular masses (36–250 kDa) were interpolated by established procedures (Weber & Osborn, 1969); p290^{SETD2} was estimated based on gene model predictions of ≈286 kDa SETD2 protein.

sequencing), an immunoreactive band should be cautiously interpreted as related to the antigen to which the antibody was originally raised. For this reason, the term 'immunoform' or α-SETD2-reactive protein (rather than SETD2 isoform/protein) is preferentially used in cases where protein identity is not conclusive.

Both antibodies detect smaller immunoforms in addition to p290^{SETD2}. Whilst the N-terminal antibody only detects one extra protein, p120^{N} (**Figure 3.1D**), the C-terminal antibody detects multiple smaller species: p250^{C}, p140^{C}, p80^{C}, p66^{C} and p45^{C} (**Figure 3.3D**). None of these proteins are substantially depleted once p290^{SETD2} is eliminated by RNAi, suggesting that these smaller proteins are unlikely to derive artefactually by proteolysis from p290^{SETD2} during cell lysis. However, the generation of forms such as p45^{C} and p80^{C} by cleavage of the larger p250^{C} or p140^{C} species (for instance)—by either a physiologically regulated proteolysis or by artefactual degradation during cell extraction— cannot be excluded at present. Future transfection of epitope-tagged cDNA constructs would be useful to assess the extent of SETD2 proteolysis and a potential relationship with these additional smaller immunoforms.

3.4.3. Feasibility of transient siRNA transfection to investigate identity of α-SETD2 immunoforms

All the siRNAs employed in this study substantially reduced the level of p290^{SETD2} after 48 hours of transfection (**Figure 3.8**) and thus validate it as a SETD2 isoform corresponding to the ≈286 kDa protein present in current databases and the previously characterised H3K36 trimethyltransferase (Edmunds *et al.*, 2008).

Although none of the smaller immunoforms were depleted by transient siRNA transfection, the slow turnover of p120^{N}, p140^{C} and p66^{C} *in vivo* suggests that these proteins may be too stable to deplete by this technique. Alternatively, lack of knockdown of smaller proteins could be due to generation from alternatively spliced transcript variants that lack regions that are targeted by siRNAs. Additionally, even if transcripts do contain siRNA binding sites, mRNA-siRNA interactions may be inhibited *in vivo* or regulated by other endogenous mechanisms.

Nevertheless, the successful RNAi-mediated depletion of p290^{SETD2} confirms that the additional immunoreactive proteins do not derive by artefactual proteolysis from p290^{SETD2} during sample preparation. This observation is especially important in refuting the possibility that thermo-urea denaturation generates p66^{C} from p290^{SETD2} by degradation. A biologically relevant derivation from regulated proteolysis of p290^{SETD2} is also unlikely, though cannot be absolutely excluded: due to their slow turnover, these proteins could derive from cleavage of p290^{SETD2} before siRNA-mediated knock-down.

The loss of global H3K36me3 after depletion of p290^{SETD2} suggests that the putative isoforms are unlikely to act as H3K36 methyltransferases that contribute substantially to global H3K36me3 levels (Edmunds *et al.*, 2008). However, a putative KMT role in generation of minute proportions of H3K36me3 or in generation of H3K36me1/2 or methylation of non-histone substrates is still possible.

3.4.4. Stability of p290^{SETD2}

The stability data here are hard to interpret in light of recently published proteomic data for murine NIH 3T3 fibroblasts, which determined SETD2 half-life as 10.75 hours (Schwanhausser *et al.*, 2011). This study employed a stable isotope labelling with amino acids in culture (SILAC)-based mass spectrometry approach, which has the advantage of more accurate quantification, but has the disadvantage that it is unable to assign the filiation of identical peptides, which may derive from distinct isoforms with different stabilities. Aside from the fact that stabilities may differ between different cell types (though the difference between 3T3 versus 10T½ fibroblasts may be small), the difference in value may thus derive from the inability of mass spectrometry to assign peptides to discrete SETD2 isoforms with distinct half-lives. Despite the difference (and the unknown extent to which the quantitative difference is actually significant), the proteomics study reveals that SETD2 falls within the top 8 percent of unstable proteins in mouse fibroblasts (n = 5001 proteins; median = 46/47.9 hours). Curiously, similar to murine p290^{SETD2}, Set2 in yeast is also highly labile ($t_{½}$ ≈ 2 minutes; global mean ≈ 43 minutes) (Belle *et al.*, 2006; Youdell *et al.*, 2008), which is consistent with many transcription factors (Ghaemmaghami

et al., 2003). Moderately rapid turnover may thus represent a conserved feature of SETD2 in eukaryotes.

3.4.5. Detection of p66^{C} after extensive thermal denaturation with urea

3.4.5.1. *Urea mediates physical and chemical changes to protein structure*

Urea produces both chemical and physical changes to protein structure, which may rationalise the requirement of such denaturation conditions for p66^{C} detection.

Although use of urea as a chaotropic protein denaturant has been recognised for over a century, the exact mechanism of denaturation is still contentious (Bennion & Daggett, 2003; Lim *et al.*, 2009). It is likely to strongly denature protein tertiary and secondary structure by: (i) disruption of hydrogen bonding of water molecules around proteins, interfering with the hydrophobic effect [FF model]; or (ii) solubilisation of polar groups and the peptide backbone during unfolding [SSKS model] (Bennion & Daggett, 2003).

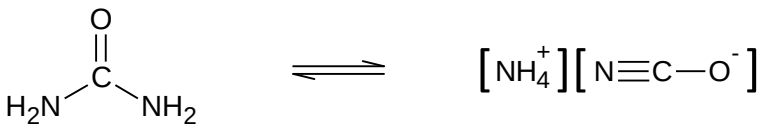
Urea can also modify proteins chemically by carbamylation (**Figure 3.10**). Urea is in equilibrium with ammonia and isocyanic acid. Since the conjugate acid of cyanate, isocyanic acid, reacts with deprotonated, primary amines to form carbamides ($HNCO + RNH_2 \rightarrow RNHC(O)NH_2$), the decomposition of urea can lead to non-enzymatic carbamylation ($CONH_2$-addition) of ε-amino group of lysines (to produce homocitrulline), arginines and the α-amino group at the N-terminus. Additionally, carbamylation affects the sulphydryl group of free cysteines, and isocyanic acid can also react similarly with tyrosine hydroxyls (Smyth, 1967). Carbamylation increases protein mass (≈4 Da per modification) and removes one positive charge per lysine modification. A long-established technical dogma in the field of proteomics is that protein extracts containing urea should never be heated, especially over 37°C, as heating accelerates the decomposition of urea into cyanate and thus promotes the generation of carbamylation-mediated artefacts (McCarthy *et al.*, 2003).

Two global consequences of thermo-urea denaturation are apparent from Coomassie-staining: Firstly, a decrease in stain intensity (**Figure 3.6B**, **D-iv**), which may result from:

Figure 3.10 The decomposition of urea causes carbamylation of proteins

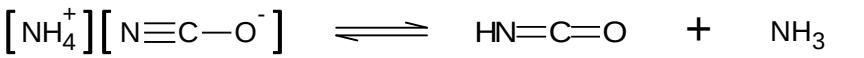
Aqueous urea exists in equilibrium with ammonium cyanate **(A)**, which inter-converts to isocyanic acid and ammonia **(B)**. Isocyanic acid reacts with nucleophiles, such as the ε-amino group of lysine, producing a carbamylated product—homocitrulline in the case of lysine **(C)**. Since the N-carbamyl group (outlined in red) is relatively stable, lysine acts as a sink for the depletion of isocyanic acid, drawing the equilibria of reactions **(A)** and **(B)** forward.

(A)



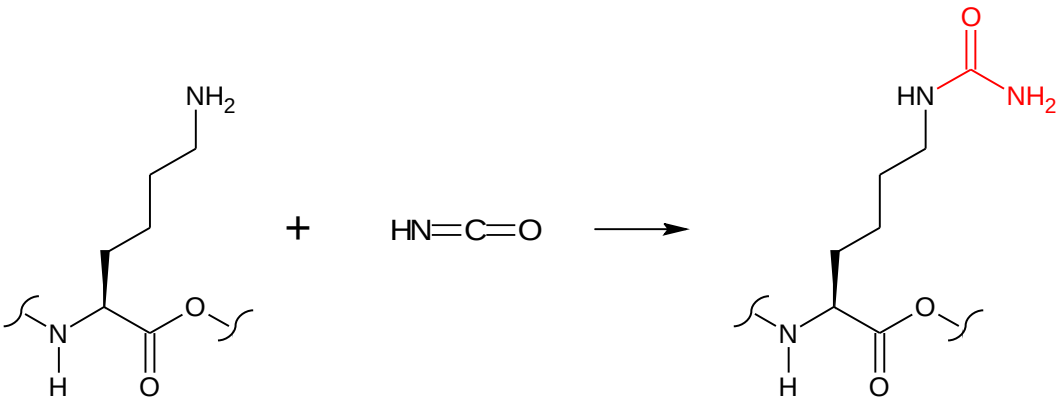
urea **ammonium cyanate**

(B)



ammonium cyanate **isocyanic acid** **ammonia**

(C)



lysine + **isocyanic acid** → **ε-*N*-carbamyl-lysine (homocitrulline)**

(i) carbamylation-derived inhibition of Coomassie-binding; (ii) protein degradation; or (iii) protein aggregation preventing entry into the resolving gel. Reduction in protein staining after increasing duration of thermal denaturation is likely explained on a global scale by chemical cleavage in SDS (Kowit & Maloney, 1982) and scission of aspartyl-proline bonds, which is accelerated at 95°C (Burgess, 2009). Secondly, thermo-urea denaturation shifts certain proteins towards higher molecular mass (**Figure 3.6B**, **D-iv**). This may result from: (i) carbamylation (Marton & Stefansson, 1986); (ii) hydrolysis of post-translational modifications; or (iii) unfolding of residual secondary structure.

The consequences for immunoblotting are complex and protein-specific: migration of nucleophosmin is substantially retarded and signal lost gradually (**Figure 3.6D-iii**); detection of p120^{N} is eliminated within 5 minutes (**Figure 3.6D-ii**); yet p66^{C} is only strongly detected after 10 minutes (**Figure 3.6D-ii**).

Although larger C-terminal signals are readily lost when extracts are heated at 95°C with 4 M urea (**Figure 3.6D-i**), this is likely explained by proteolysis and epitope modification.At lower concentrations of urea (0.25 M), p66^{C} appears without loss of signal of comparable intensity from larger C-terminal proteins (**Figure 3.6C**). Importantly, RNAi-mediated elimination of p290^{SETD2} demonstrates p66^{C} is not derived therefrom by artefactual proteolysis (**Figure 3.8**). In fact, no other protein detected with the C-terminal antibody after SDS-PAGE is comparable in intensity to that of p66^{C}.

Reproducible solubilisation of certain proteins (histones and HMGN1, for example) in SDS-PAGE loading buffer requires additional 4 M urea (Barratt *et al.*, 1994b). The presence of urea alone, however, is insufficient for p66^{C} detection (**Figure 3.6:A–D**). Additionally, urea is not sufficient to expose p66^{C} after thermal denaturation (**Figure A.2**) indicating that concomitant thermal and urea-mediated denaturation is required.

A mimotope may be generated artefactually by chemical modification (*e.g.*carbamylation): an epitope, unreactive under conventional denaturation conditions, may be converted by thermo-urea denaturation into a mimotope—irrespective of whether the epitope derives from a *bona fide* SETD2 isoform or an unrelated protein. Although carbamylation would be

extensive after boiling in 4 M urea for 10–15 minutes (McCarthy *et al.*, 2003), this possibility will be excluded later by the observation of p66^{C} under conditions where carbamylation-mediated epitope conversion is excluded (see section 4.3.4).

In contrast to generating a reactive epitope, thermo-urea denaturation may hydrolyse post-translational modifications (within or adjacent to the epitope) that could occlude paratope binding under conventional denaturation conditions. The C-terminal SETD2 epitope (RGTKRPDDRYDTPTSKKKVR) contains residues with a high propensity for post-translational modification, such as phosphorylation, acetylation, methylation and ubiquitylation. Furthermore, antibody binding may be influenced by a three/four-residue cluster of phosphorylated serine/tyrosine that has been detected immediately upstream of the epitope in mouse and human SETD2 by mass spectrometry (Dephoure *et al.*, 2008; Gauci *et al.*, 2009; Mayya *et al.*, 2009). For example, if p66^{C} corresponds to a SETD2 isoform and these phosphorylated residues occlude recognition by the C-terminal antibody, heating in urea may promote their hydrolysis and thus permit binding of the antibody.

An alternative (and not necessarily mutually exclusive) scenario is that the p66^{C} epitope may be conformationally inaccessible due to a localised structure that is resistant to denaturation unless heated in the presence of urea. SDS-resistant secondary structure and complex formation is not uncommon, as exemplified by: (i) SDS-resistant complexes of HUB1 (Luders *et al.*, 2003); (ii) iNOS dimers, which resist denaturation by boiling in Laemmli buffer, 4 M urea and 6 M guanidine hydrochloride (Kolodziejski *et al.*, 2003); (iii) SDS-resistant Aβ amyloid oligomers (Chang & Abraham, 1996); (iv) SNAP-SNARE complexes mediated by coiled-coil motifs, which resist denaturation in standard Laemmli buffer upto 95°C (Yang *et al.*, 1999); and (v) α-synuclein oligomers (Cappai *et al.*, 2005). An extreme case is the protease–tetrabrachion complex from *Staphylothermus marinus,* which requires 70% (*v*/*v*) sulphuric acid or fuming trifluoromethanesulphonic acid for complete denaturation whilst partial dissociation of its protease subunit occurs only at 130°C in 1% (*w*/*v*) SDS or 6 M guanidinium hydrochloride for complete denaturation at neutral pH. Its resistance to denaturation is explained by extremely stable coiled-coil structures (Peters *et al.*, 1996). Surprisingly, however (and relevant to observations in the next chapter), the

protease can be dissociated from the tetrabrachion complex at room temperature when incubated below pH 4.5 in 0.1% (*w*/*v*) SDS (Peters *et al.*, 1995).

3.4.5.2. *Apparent rarity of proteins exhibiting detection characteristics of p66^{C}*

The detection of p66^{C} only afer thermo-urea denaturation in four different cell types suggests conservation of its structure or epitope between human and murine sources. However, extensive literature searches revealed only a couple of examples of proteins that partially exhibit a similar behaviour to p66^{C} towards denaturation with urea and heat.

Firstly, increased detection of [URE3] prion in *S. cerevisiae* with α-Ure2p antibody critically requires boiling samples for 5 minutes in 8 M urea (Ripaud *et al.*, 2004). However, this response occurs exclusively for insoluble [URE3], an aggregated form distinct from amyloid fibrils, since the soluble population is detectable by α-Ure2p in the absence of heating in urea (Ripaud *et al.*, 2004). Parallels are hard to draw with p66^{C}, however, since it appears that virtually all p66^{C} requires thermo-denaturation for detection, and this requirement is independent of sedimentation behaviour (**Figures 4.3B** and **A.3**), suggesting protein aggregation akin to [URE3] is unlikely to be responsible in the case of p66^{C}.

Secondly, glycan-based epitopes contained within complement molecule C1q and specific members of the collagen family are detected only by concanavalin A and α-collagen antibody, respectively, after heating at 100°C in the presence of 6 M urea for at least 5 minutes (Marton & Stefansson, 1986). These experiments were performed with purified collagen mixtures, and the increase in immunoblot signal does not correspond to any band stainable by Coomassie or India ink. Analogous to p66^{C}, the immunodetection of the was dependent on urea concentration and duration of heating (Marton & Stefansson, 1986). The authors suggest that the relevant epitopes remain concealed in a structure that is incompletely denatured by urea and SDS without such heating.[1] The structural properties

[1]One potential caveat of this study, however, is that the antibodies were originally raised against collagen protein that had been extracted after lengthy exposure to high concentrations of urea, and heated at 100°C for 2 minutes in urea. Antibodies may therefore have been generated unknowingly against a carbamylated immunogen.

responsible were not characterised, but appear to be specific to certain types of collagen type.

Since the aforementioned studies only partially approximate the requirement of thermo-urea denaturation for p66^{C} immunodetection, this behaviour of p66^{C} represents a substantially novel finding with little precedent as far as we are aware. The denaturation requirement may have functional relevance with regards to the interaction of C-terminal fragments of SETD2 with Huntingtin, CIC and Ataxin-1, which are all polyglutamine-containing proteins that are susceptible to aggregation and amyloid formation *in vivo* during neurodegeneration. Relevant to p66^{C} behaviour, the C-terminal SETD2 epitope overlaps a predicted coiled-coil upstream of the QP-rich region. Coiled-coils and glutamine-rich regions have been strongly implicated in prion-related aggregate formation and—pertinent to SETD2 function—the formation of and *in vivo* behaviour of Huntingtin aggregates (Fiumara *et al.*, 2010). As discussed above, coiled-coils often form stable structures that resist conventional SDS-PAGE denaturation procedures. With respect to this, both human and murine SETD2 contain strongly predicted ($P \approx 0.8$) coiled-coil regions in the C-terminus, just overlapping the C-terminal antibody epitope.

3.4.5.3. *Consequences of the denaturation requirement for identification of p66^{C}*

The requirement of thermal denaturation in the presence of urea for detection of p66^{C} hinders characterisation by standard proteomic techniques—mass spectrometry and N-terminal sequencing—due to artefactual carbamylation. This modification produces a heterogenous population of proteins, with changes in mass and charge to an unpredictable extent, which makes peptide identification by mass spectrometry near impossible. The modifications further alter cleavage sites for protease activity, which is required to produce appropriately-sized and predictable peptides for analysis. N-terminal carbamylation also immediately inhibits sequencing by Edman degradation.

3.5. Conclusions

In conclusion, affinity-purified antibodies detected the p290^{SETD2} methyltransferase in addition to several smaller immunoforms. Evaluation of the latter proteins as putative SETD2 isoforms by RNAi was inconclusive, because the proteins appear to turnover too slowly to be amenable to depletion by transient siRNA transfection. The next chapter explores means to facilitate their physical identification, whilst Chapter 5 investigates the transcriptional products of the murine *Setd2* gene to assess how transcriptional diversity may relate to the putative isoforms identified here.

Chapter 4 Generation of a p66-enriched fraction for two-dimensional gel electrophoresis and proteomic analysis

4.1. Introduction

Verification of the immunoforms detected in Chapter 3 as putative SETD2 isoforms requires independent corroboration. The most direct method to ascertain the identity of these molecules is to obtain peptide sequence information—either by mass spectrometry or N-terminal sequencing. Both of these techniques require isolation of the protein of interest. To this end, subcellular fractionation was employed to enrich for proteins of interest and further characterise their properties.

The requirement of thermo-urea denaturation for immunodetection of $p66^{C}$ poses a problem for proteomic identification: in the absence of such denaturation, carbamylation is avoided but without detection of $p66^{C}$. Note that rigorous identification of unknown protein requires certainty that the protein sample is in fact taken forward for analysis. Currently, it is uncertain whether $p66^{C}$ even enters the polyacrylamide gel in the absence of denaturation.

The first section of this chapter aims to further investigate the identity, provenance and properties of putative SETD2 isoforms (along with the confirmed $p290^{SETD2}$ form) by various subcellular fractionation techniques. The second section uses this fractionation information in combination with two-dimensional electrophoresis to show p66 is detected by two independent α-SETD2 antibodies, and to facilitate proteomic identification by mass spectrometry.

4.2. Objectives

The objectives of this chapter are:

- To characterise subcellular fractionation properties of $p290^{SETD2}$ and putative SETD2 isoforms.
- To generate a $p66^{C}$-enriched subcellular fraction.
- To develop a method to detect and physically identify $p66^{C}$ in the absence of thermo-urea denaturation.
- To characterise the $p66^{C}$-enriched fraction by proteomic methods.

4.3. Results

4.3.1. Differential extraction of α-SETD2-reactive proteins by nuclease digestion and high-salt extraction

To further characterise α-SETD2-reactive proteins and facilitate their identification by proteomic methods, several types of sequential extraction were tested to generate fractions enriched in specific immunoforms. This chapter describes the two most effective methods, which are additionally informative about differences in subcellular properties of α-SETD2-reactive forms.

Murine 10T½ fibroblasts were differentially fractionated by hypotonic lysis and sequential extraction of insoluble material by nuclease digestion and high-salt extraction (**Figure 4.1**). This method was derived from classical nuclear fractionation studies that progressively extract soluble cytoplasmic and nuclear components to leave an insoluble nuclear matrix (Rampersaud, 2003). This method generates four fractions that differ in protein composition as a result of differences in subcellular localisation, interactions and inherent protein characteristics. Although each fraction contains a complex mixture of protein, certain fractions are predominantly enriched for discernible species: the high-salt

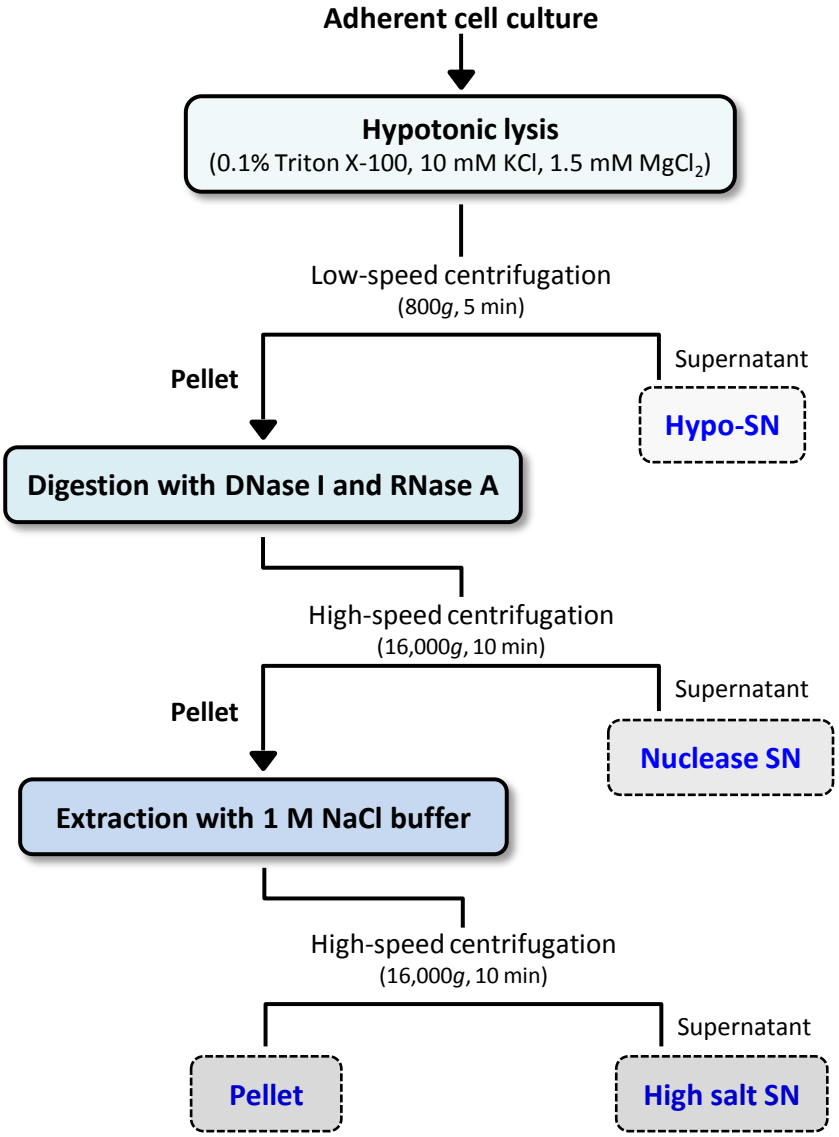


Figure 4.1 Subcellular fractionation by hypotonic lysis, nuclease digestion and high salt extraction

See section 2.7.3 for complete details of method.

extract was enriched in core histones and histone H1 (**Figure 4.2A**, lanes 5 and 6) whilst the final pellet was largely devoid of histones but enriched in the cytoskeletal protein vimentin (**Figure 4.2A**, lanes 7 and 8) and nuclear matrix component lamin B (**Figure 4.2B-v**, lanes 7 and 8). Each extract was incubated at room temperature or at 95°C for 10 minutes prior to analysis; the most conspicuous changes in Coomassie-stainable protein occurred for histones, which appeared to migrate faster and become less well resolved (**Figure 4.2A**, lane 5 versus 6).

For SETD2, $p290^{SETD2}$ was predominantly enriched in the high-salt extract (**Figure 4.2B-i/iii**, lanes 5 and 6), though a small proportion was extracted in the nuclease extract and final pellet (**Figure 4.2**, lanes 3 and 7). Virtually all $p120^{N}$ was extracted in the initial hypotonic supernatant (Hypo-SN) and its detection was severely impaired after heating (**Figure 4.2B-ii**, lane 1 versus 2). Approximately equal amounts of $p140^{C}$ fractionated in the nuclease-digest and high-salt extracts (**Figure 4.2B-iii**, lanes 3–6). $p80^{C}$ fractionated in Hypo-SN and nuclease-digest fractions (**Figure 4.2B-iii**, lanes 1–4), and electrophoretic retardation occurred after thermal denaturation. Consistent with previous results, $p66^{C}$ was only detected in extracts denatured at 95°C: a minority was extracted in Hypo-SN (**Figure 4.2B-iii**, lane 2), but the remainder was substantially more resistant to extraction with a proportion released by high-salt and the majority remaining in the final pellet (**Figure 4.2B-iv**, lanes 6 and 8).

For $p66^{C}$ that extracted by high-salt, virtually all existed in a complexed form, which sedimented after ultracentrifugation at 100,000*g* for one hour (**Figure 4.3B**, lane 4). The distinct enrichment of $p66^{C}$ in the P^{100} fraction is striking, given most Coomassie-stainable protein, RNAPII, $p140^{C}$, and the majority of $p290^{SETD2}$ remained soluble in the S^{100} fraction (**Figure 4.3-A** and **B**, lanes 1 and 2). Nuclease-salt fractionation thus highlighted the complexed nature of $p66^{C}$ two-fold: firstly, that virtually all $p66^{C}$ extracted by high-salt sediments at 100,000*g*; secondly, that the salt-resistant $p66^{C}$ co-sediments at 16,000*g* with insoluble material, comprising components of the cytoskeleton and nuclear matrix.

Figure 4.2 Subcellular fractionation of 10T½ cells by sequential nuclease digestion and high-salt extraction reveals distinct fractionation of SETD2 immunoforms

(A) Total protein content of subcellular fractionations: Quiesced 10T½ fibroblasts were lysed in a hypotonic buffer and centrifuged at 800*g* to prepare the Hypo-SN supernatant (lanes 1 and 2). The insoluble pellet containing crude nuclei was digested with DNase I and RNase A to yield a nuclease (NAse)–digest fraction (lanes 3 and 4), and the nuclease-resistant material was extracted with a buffer containing 1 M NaCl, to yield a high-salt supernatant (lanes 5 and 6), and a pellet enriched for elements of the cytoskeleton and nuclear substructure (lanes 7 and 8). Supernatants were precipitated with TCA and all samples solubilised in urea/SDS-PAGE loading buffer before incubation at room temperature (odd lanes marked with "–") or at 95°C for 10 minutes (even lanes marked with "+"). Equivalent proportions of each subcellular fraction (5% of a 145 mm plate) were separated by 15% SDS-PAGE and protein content was visualised by Coomassie staining. The asterisk marks soybean trypsin inhibitor present in extraction buffers.

(B) Differential distribution of SETD2 immunoforms after subcellular fractionation: Extracts prepared as in **(A)** were separated by 8% SDS-PAGE, transferred to PVDF in Towbin **(i–iv)** or SDS **(v)** buffer, and probed using α-SETD2 N-terminal **(i, ii)**; α-SETD2 C-terminal **(iii, iv)**; and α-Lamin B **(v)** antibodies. Panel **(iv)** is a lighter exposure of panel **(iii)**.

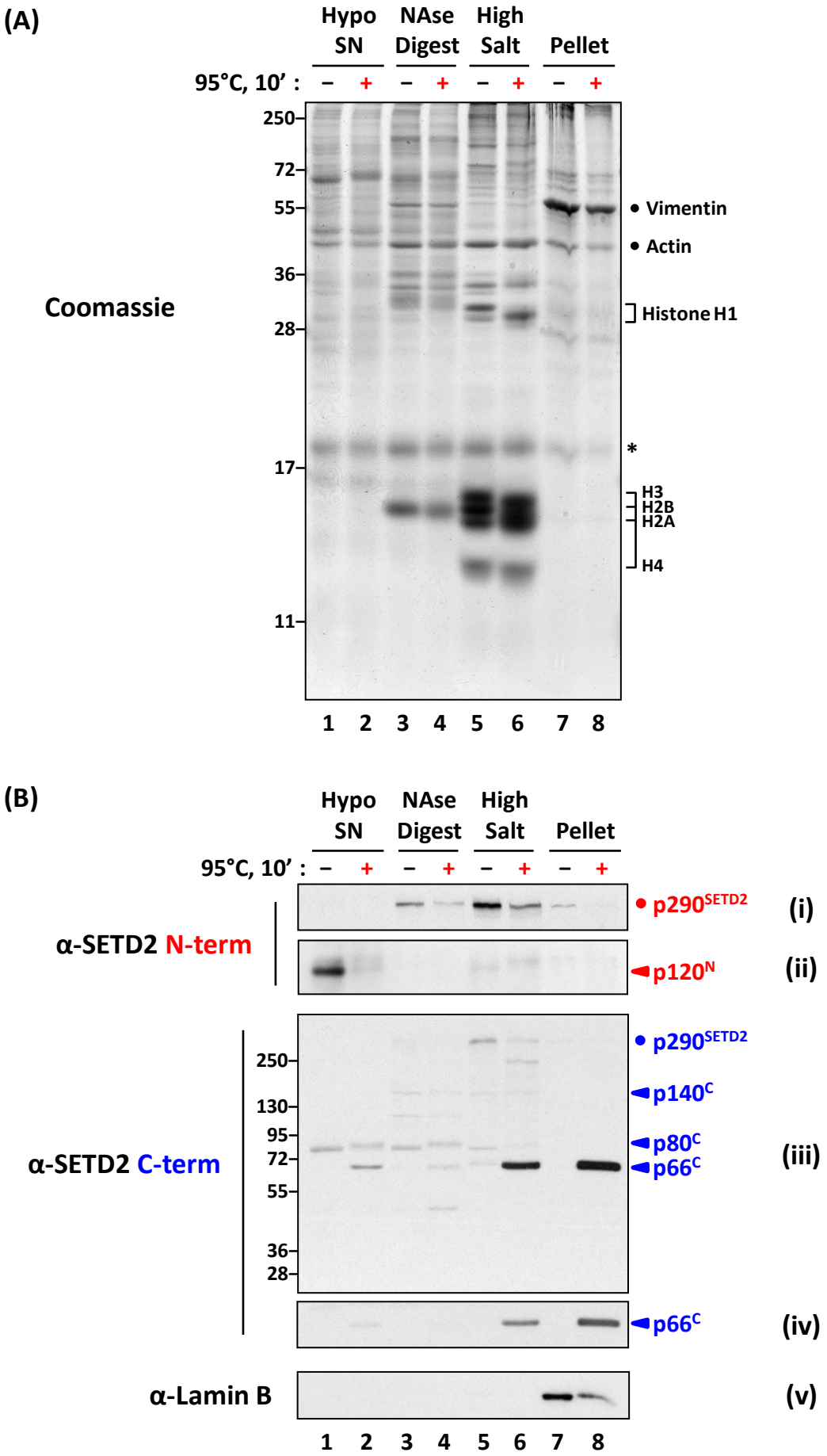
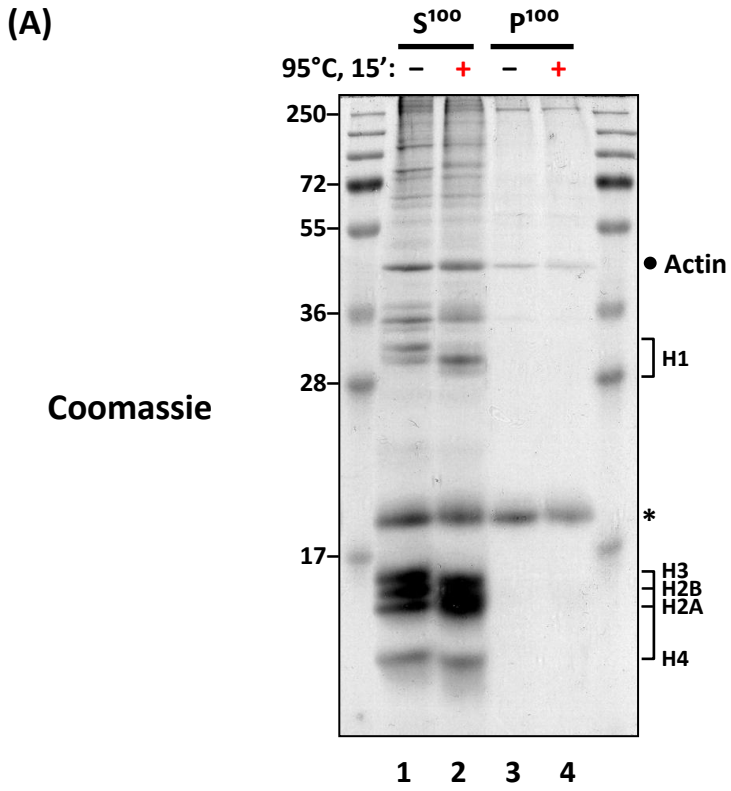
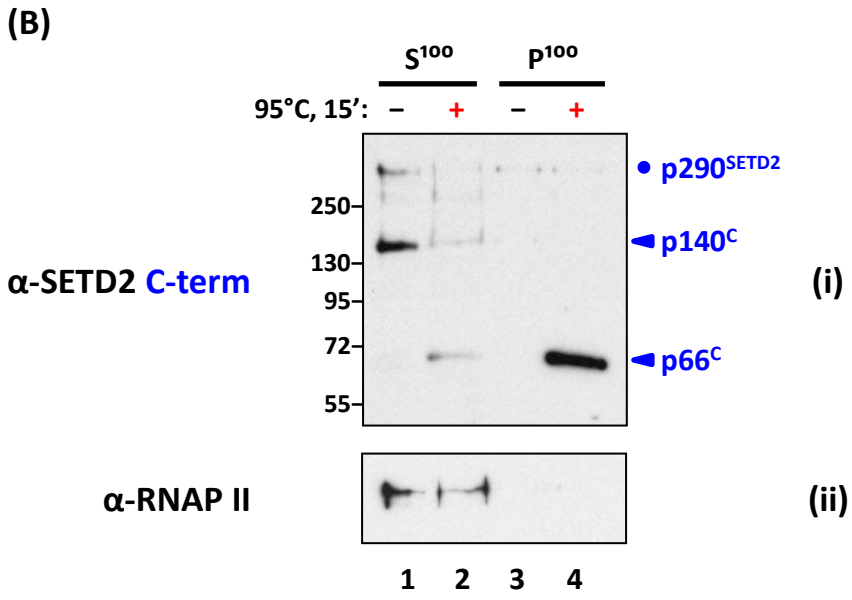

(A)
Hypo SN
NAse Digest
High Salt
Pellet
95°C, 10' : – + – + – + – +
250
72
55
36
28
17
11
Coomassie
• Vimentin
• Actin
Histone H1
*
H3
H2B
H2A
H4
1 2 3 4 5 6 7 8
(B)
Hypo SN
NAse Digest
High Salt
Pellet
95°C, 10' : – + – + – + – +
α-SETD2 N-term
• p290SETD2
(i)
p120N
(ii)
α-SETD2 C-term
250
130
95
72
55
36
28
• p290SETD2
p140C
p80C
p66C
(iii)
p66C
(iv)
α-Lamin B
(v)
1 2 3 4 5 6 7 8

Figure 4.3 Virtually all p66^{C} that is extracted by high-salt buffer exists in a complexed form that sediments by ultracentrifugation

(A) Total protein content of complexed (P^{100}) and non-complexed (S^{100}) high-salt extract: The high-salt supernatant from quiescent 10T½ cells (**Figure 4.2**) was centrifuged at 100,000*g* for 1 hour at 4°C to yield S^{100} supernatant ('non-complexed') and P^{100} pellet ('complexed') fractions. The supernatant was precipitated with TCA and both fractions were re-solubilised in urea/SDS-PAGE loading buffer. After incubation at room temperature (lanes 1 and 3) or heating at 95°C for 15 minutes (lanes 2 and 4), equal proportions of extracts were separated by 15% SDS-PAGE for subsequent Coomassie-staining. The asterisk indicates soybean trypsin inhibitor present in extraction buffers.

(B) Sedimentation of p66^{C} in high-salt extract: extracts prepared in **(A)** were separated by 8% SDS-PAGE and transferred to PVDF in SDS buffer. Membranes were probed with α-SETD2 C-terminal **(i)** or α-RNAPII **(ii)** antibodies.


(A)
S100
P100
95°C, 15': – + – +
250
72
55
36
28
17
Actin
H1
*
H3
H2B
H2A
H4
Coomassie
1 2 3 4


(B)
S100
P100
95°C, 15': – + – +
250
130
95
72
55
p290SETD2
p140C
p66C
α-SETD2 C-term
(i)
α-RNAP II
(ii)
1 2 3 4

Although this method can yield fractions that are characterised by exclusive enrichment of $p120^{N}$, $p66^{C}$ and $p140^{C}/p290^{SETD2}$, it is limited in generating a $p66^{C}$-enriched fraction, since $p66^{C}$ is divided between two fractions, with the majority of $p66^{C}$ retained in the insoluble pellet. The $p66^{C}$ retained in the final pellet resists complete extraction by detergent and guanidine hydrochloride (*cf.* **Figure B.1**; data not shown). A different fractionation procedure was therefore developed based on differential detergent extraction and sedimentation to generate a soluble fraction containing the majority of $p66^{C}$ separated from other immunoforms.

4.3.2. Generation of a $p66^{C}$-enriched fraction by differential detergent extraction and centrifugation

Combinations of ionic and non-ionic detergents and different salt concentrations were screened to maximise $p66^{C}$ extraction from crude nuclei prepared by low-speed centrifugation after hypotonic lysis. Two buffers were found ideal for virtually complete extraction of $p66^{C}$ from crude nuclei: whilst TX–DOC buffer co-extracted all $p66^{C}$ with approximately half of $p290^{SETD2}$ and $p140^{C}$ into the TD fraction (**Figure 4.4A**, lane 2), Triton-extraction buffer (TX–DOC lacking 1% (*w*/*v*) sodium deoxycholate) released virtually all $p66^{C}$ in the TX fraction with minor contamination of $p290^{SETD2}$ and $p140^{C}$ (**Figure 4.4A**, lane 4). To assess whether the detergent-solubilised $p66^{C}$ was complexed and amenable to further purification, the TX and TD fractions were centrifuged at 100,000*g* for one hour. Ultracentrifugation revealed that the majority of $p66^{C}$ from the TX fraction sedimented in the TX-P^{100} fraction (**Figure 4.4B**, lane 9), suggesting it existed in a complexed form under these conditions. Virtually all $p66^{C}$ from the TD fraction, however, remained soluble whereas $p290^{SETD2}$ and $p140^{C}$ distributed between both supernatant and pellet (**Figure 4.4B**, lanes 5–6). Ultracentrifugation of the Hypo-SN fraction revealed that the tiny amount of $p66^{C}$ that was extracted also predominantly sedimented (**Figure 4.4B**, lane 3).

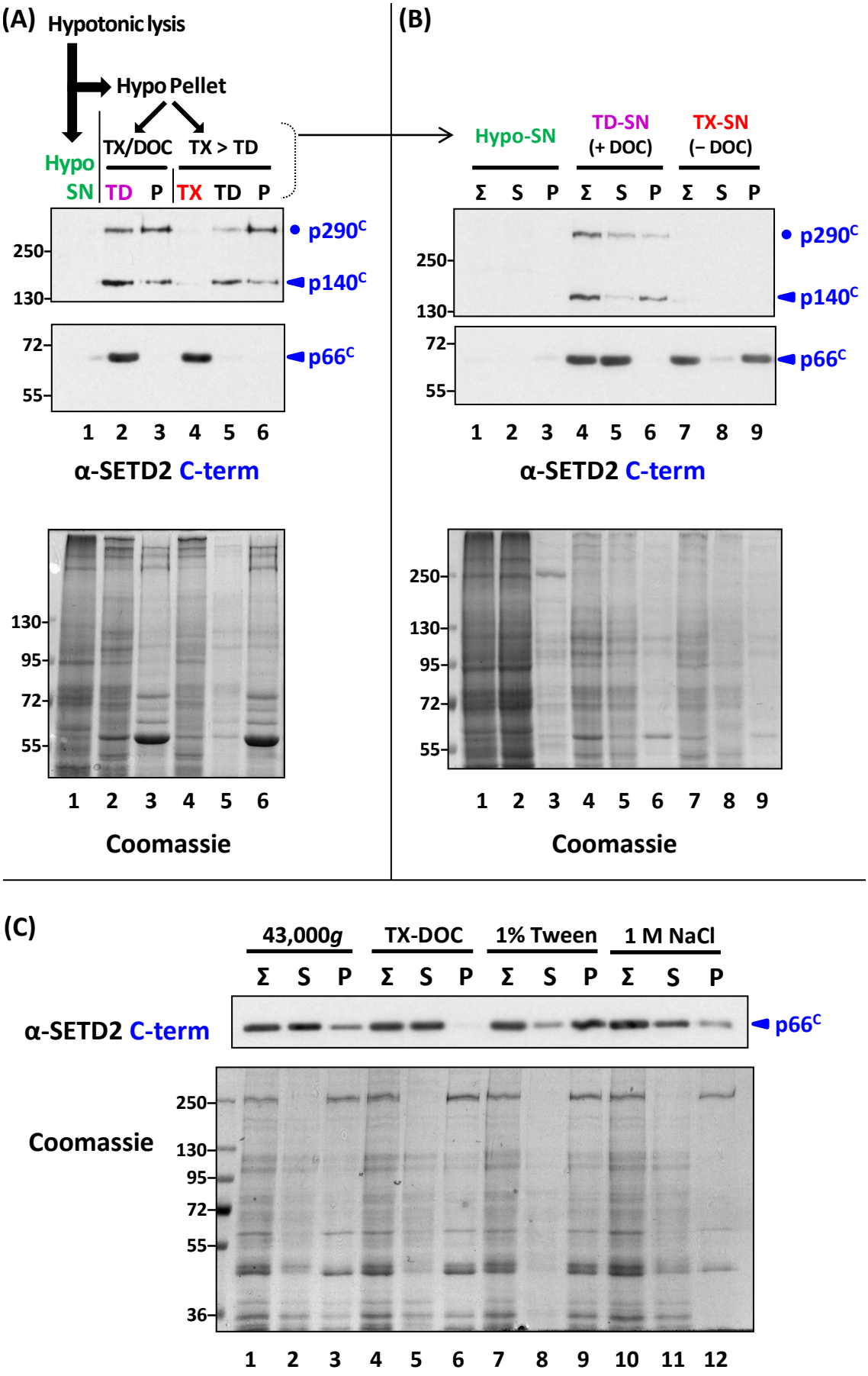
Based on the observation that $p66^{C}$ did not sediment after ultracentrifugation of TD extracts (**Figure 4.4B**, lanes 5–6) or 'whole cell' TX–DOC extracts (**Figure A.3** in Appendix A), $p66^{C}$ was further extracted from the TX-P^{100} fraction by resolubilisation in TX–DOC

Figure 4.4 Characterisation of the detergent extraction and sedimentation properties of C-terminal SETD2 immunoforms

(A) Sodium deoxycholate/Triton X-100–containing buffers differentially extract p66^{C} and p290^{SETD2}/p140^{C} from crude nuclei: Quiesced 10T½ cells were lysed by hypotonic lysis and low-speed centrifugation to generate a hypotonic supernatant (lane 1). The pellet was processed by two methods: (i) extracted with TX–DOC buffer and centrifuged at 16,000*g* for 10 minutes to yield a TD-supernatant (lane 2) and TD-pellet (lane 3); or (ii) extracted sequentially with Triton-extraction buffer and centrifugation as above to yield a TX-supernatant (lane 4) and a pellet, which was further extracted by re-extraction in TX–DOC buffer and centrifugation, producing a TD-supernatant (lane 5) and TD-pellet (lane 6). Supernatants were precipitated with TCA and pellets solubilised in urea/SDS-PAGE loading buffer before heating at 95°C for 10 minutes. Equivalent proportions of each fraction were separated by 8% SDS-PAGE and then gels were either stained with Coomassie or transferred to PVDF in SDS buffer for subsequent immunoblotting with α-SETD2 C-terminal antibody.

(B) Sedimentation of p66^{C} in Triton X-100–containing buffer that lacks sodium deoxycholate: The hypotonic and p66^{C}-enriched supernatants from **(A)** were centrifuged at 100,000*g* for 1 hour at 4°C, and then extracts were precipitated, resolubilised and denatured as in **(A)**. Equivalent proportions of total (Σ, non-centrifuged) extracts (lanes 1, 4 and 7), ultracentrifugation supernatants (lanes 2, 5 and 8) and pellets (lanes 3, 6 and 9) were resolved by 8% SDS-PAGE and gels were processed as in **(A)**.

(C) Deoxycholate converts complexed p66^{C} into a soluble non-complexed form: Equal aliquots of the TX-P^{100} pellet (lane 9 in panel **B**) were solubilised in Triton-extraction buffer (lanes 1–3), TX–DOC buffer (lanes 4–6), Triton-extraction buffer containing 1% (*v*/*v*) Tween-20 (lanes 7–9), or Triton-extraction buffer containing 1 M NaCl (lanes 10–12). Re-solubilised extracts were centrifuged for 1 hour at 4°C at either 43,000*g* (lanes 2 and 3) or 100,000*g* (lanes 5, 6, 8, 9, 11 and 12). Samples were precipitated, resolubilised and denatured as in **(A).** Equal proportions of total (Σ, non-centrifuged) extracts (lanes 1, 4, 7 and 10), centrifuged supernatants (lanes 2, 5, 8 and 11) and pellets (lanes 3, 6, 9 and 12) were resolved by 8% SDS-PAGE and gels were processed as in **(A)**.


(A)
Hypotonic lysis
Hypo Pellet
Hypo SN
TX/DOC
TX > TD
TD P TX TD P
250
130
72
55
p290C
p140C
p66C
1 2 3 4 5 6
α-SETD2 C-term
130
95
72
55
1 2 3 4 5 6
Coomassie
(B)
Hypo-SN
TD-SN (+ DOC)
TX-SN (– DOC)
Σ S P Σ S P Σ S P
250
130
72
55
p290C
p140C
p66C
1 2 3 4 5 6 7 8 9
α-SETD2 C-term
250
130
95
72
55
1 2 3 4 5 6 7 8 9
Coomassie
(C)
43,000g
TX-DOC
1% Tween
1 M NaCl
Σ S P Σ S P Σ S P Σ S P
α-SETD2 C-term
p66C
Coomassie
250
130
95
72
55
36
1 2 3 4 5 6 7 8 9 10 11 12

buffer and re-ultracentrifugation. This treatment moved virtually all p66^{C} into the soluble supernatant (**Figure 4.4C**, lane 5), indicating that 1% (*w*/*v*) sodium deoxycholate is critical in converting p66^{C} from a complexed sedimentable species into a non-complexed form. Other attempts to further purify p66^{C} through use of lower centrifugation speed or extraction with Triton-containing buffer supplemented with 1% (*v*/*v*) Tween or 1 M NaCl did not aid generation of a maximally p66^{C}-enriched fraction (**Figure 4.4C**).

The unique fractionation properties of p66^{C} compared to other α-SETD2-reactive proteins permitted generation of a p66^{C}-enriched fraction by a differential detergent extraction and centrifugation (DDC) method (outlined in **Figure 4.5**). The extent of p66^{C} fractionation and total Coomassie-stained protein were monitored at each stage of the method (**Figure 4.6A**), demonstrating p66^{C}-enrichment in the TD-S^{100} fraction (**Figure 4.6A-ii**, lane 8). Partial purification of p66^{C} was accompanied with progressive reduction in Coomassie-stainable protein (**Figure 4.6B**). In this particular purification, loss of p66^{C} in TX-P and TD-P^{100} fractions was minor, but a substantial amount was retained in the TX-S^{100} fraction (**Figure 4.6A-ii**, lane 6). The distribution of p66^{C} between TX-S^{100} and TD-S^{100} fractions appeared to fluctuate between experiments for unclear reasons, but in all cases, the p66^{C}-enrichment consistently skewed towards TD-S^{100} (*cf*. **Figure 4.4B**, **Figure 4.6A-ii** and **Figure 4.7A-ii**). Nevertheless, this procedure successfully generates a p66^{C}-enriched TD-S^{100} fraction that is substantially depleted of total Coomassie-stainable protein (**Figure 4.6A-i**, lane 8), which is amenable for downstream two-dimensional gel electrophoresis and proteomic techniques.

Upon completion, the DDC procedure generates five fractions: (i) Hypo-SN, which is essentially a crude cytoplasmic extract, (ii) TX-S^{100}, (iii) TD-S^{100}, (iv) TD-P^{100} and (v) TX-P. The most protein-abundant fractions are Hypo-SN, TX-S^{100} and TX-P (**Figure 4.7A-ix**, lanes 1, 2 and 5). The TX-P fraction is enriched for cytoskeletal elements, such as actin and vimentin (**Figure 4.7A-ix**, lane 5), along with nuclear proteins: hypo- and hyper-phosphorylated RNAPII, lamin B, nucleophosmin and histones (**Figure 4.7A: v–viii,** lane 5). A small proportion of nucleophosmin fractionated in the Hypo-SN, TX-S^{100} and TD-S^{100} fractions (**Figure 4.7A-vii,** lanes 1, 2 and 3). The TX-P fraction also contained virtually all p140^{C} and p290^{SETD2} (**Figure 4.7A-iii/iv,** lane 5).

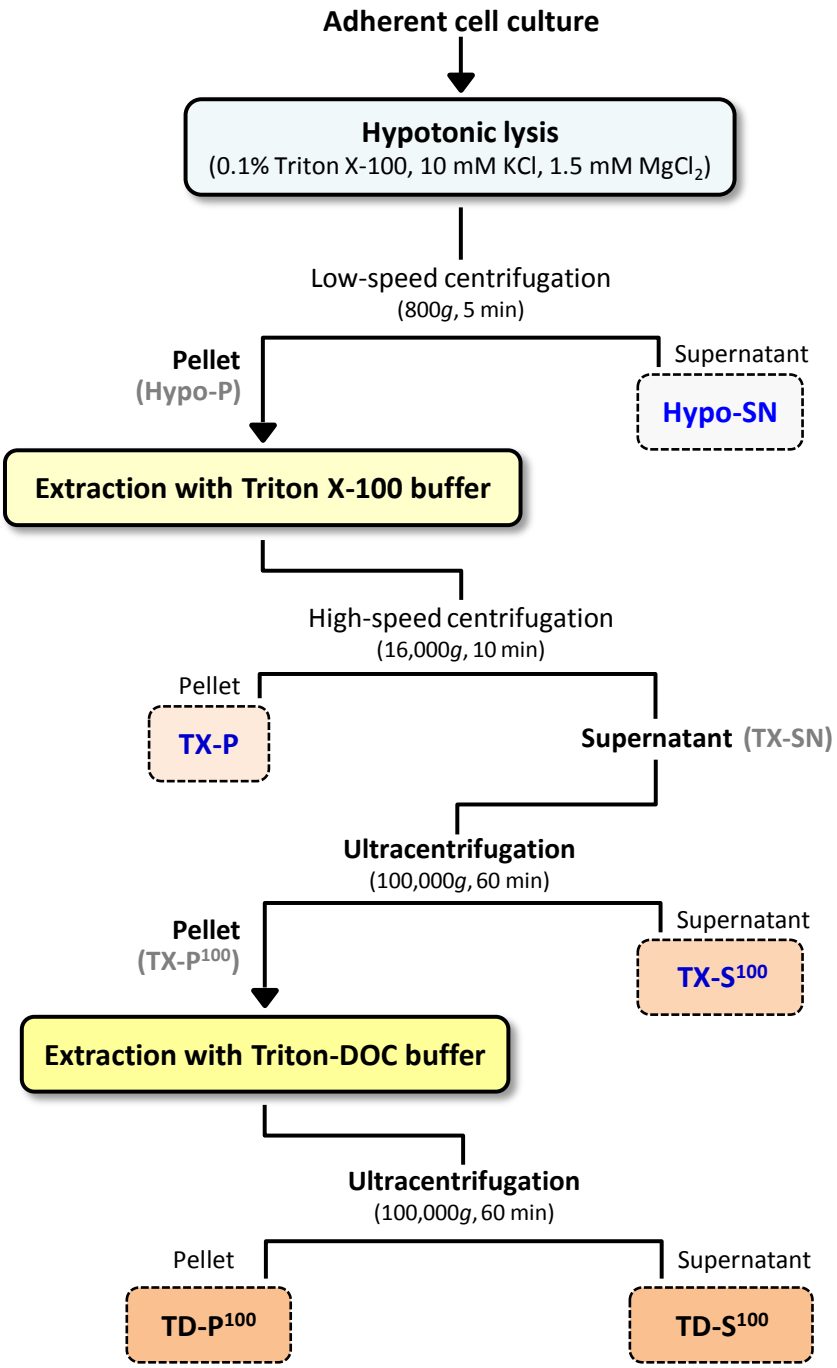


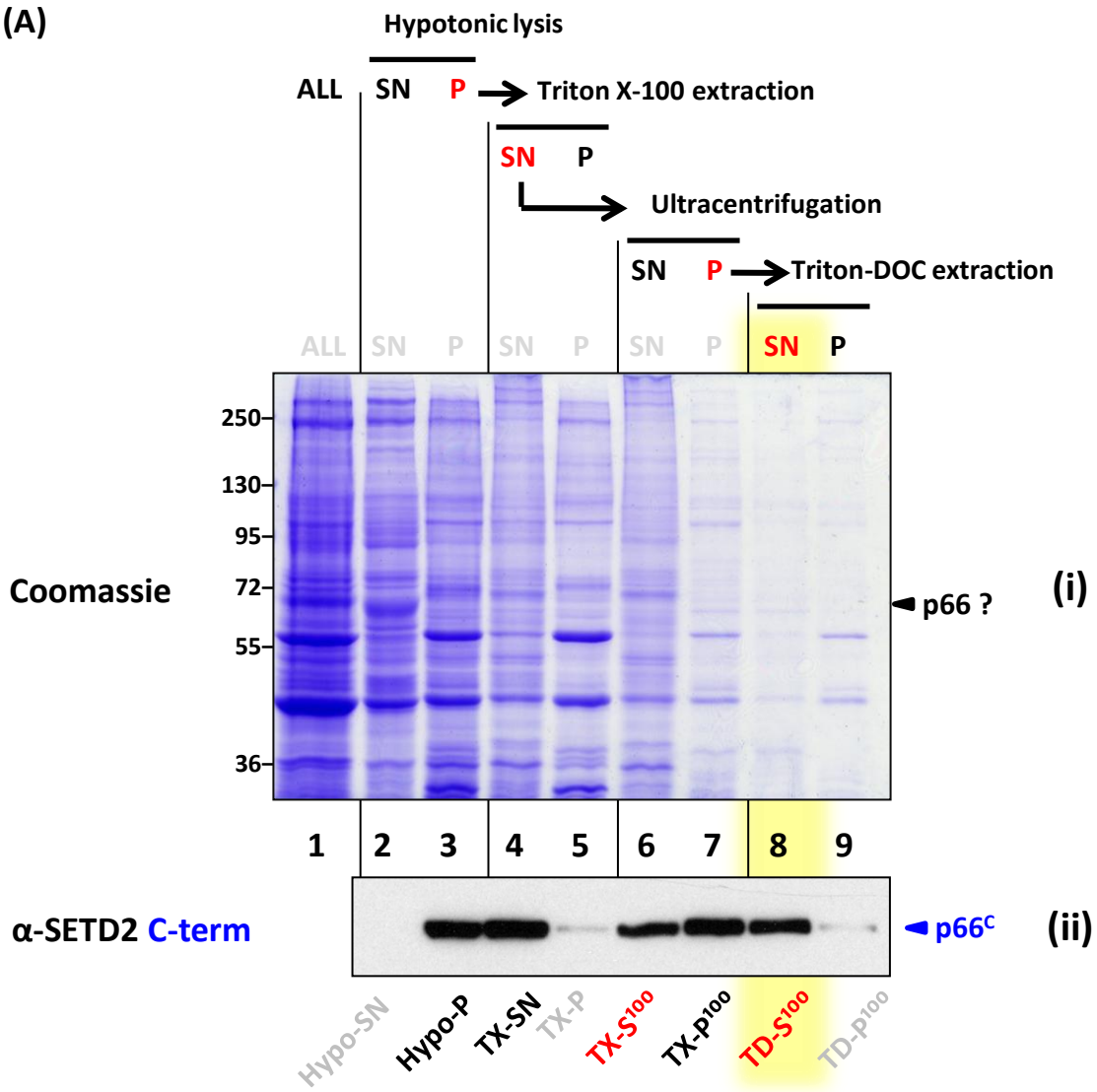
Figure 4.2 Subcellular fractionation by differential detergent extraction and centrifugation (DDC)

See section 2.7.4 for complete details of fractionation method

Figure 4.6 Generation of a p66^{C}-enriched fraction by differential detergent extraction and centrifugation

(A) Enrichment of p66^{C} in TX-S^{100} and TD-S^{100} fractions. Quiesced 10T½ cells were fractionated by the DDC protocol and equivalent proportions of fractions produced at each stage of the method were analysed. Extracts were precipitated with TCA, resolubilised in urea/SDS-PAGE loading buffer and incubated at room temperature **(i)** or at 95°C for 10 minutes **(ii)**. Samples were separated by 8% SDS-PAGE and then either stained with Coomassie **(i)** or transferred to PVDF in SDS buffer for immunoblotting with α-SETD2 C-terminal antibody **(ii)**. The final p66^{C}-containing fractions are highlighted in red. Across the procedure, p66^{C}-depleted and p66^{C}-enriched fractions are annotated below panel **(ii)** in grey and black, respectively.

(B) Overview of total protein reduction during DDC fractionation of 10T½ cells: Equivalent proportions of p66^{C}-enriched fractions from **(A-i)** were separated by 8% SDS-PAGE and stained with Coomassie.


(A)
Hypotonic lysis
ALL
SN
P
Triton X-100 extraction
SN
P
Ultracentrifugation
SN
P
Triton-DOC extraction
ALL SN P SN P SN P SN P
250
130
95
72
55
36
Coomassie
p66 ?
(i)
1 2 3 4 5 6 7 8 9
α-SETD2 C-term
p66C
(ii)
Hypo-SN
Hypo-P
TX-SN
TX-P
TX-S100
TX-P100
TD-S100
TD-P100

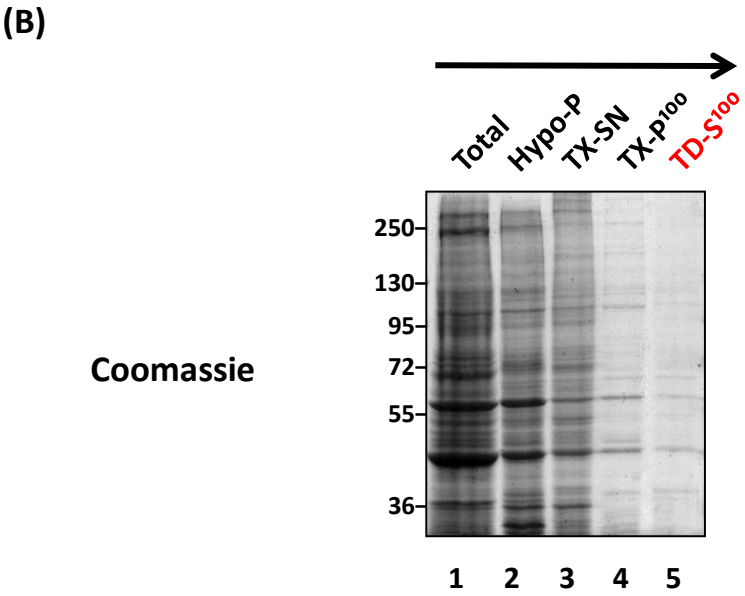
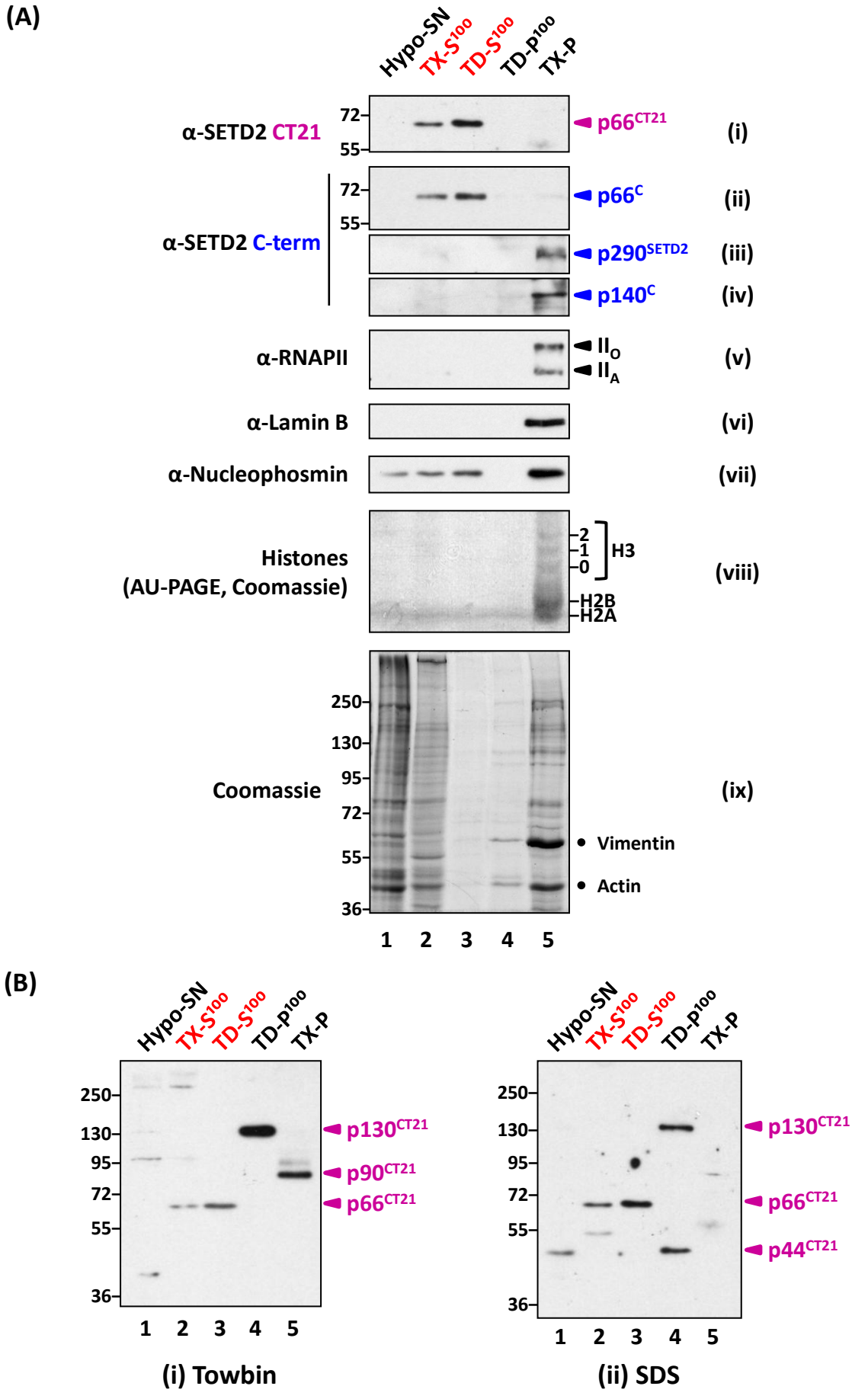

(B)
Total
Hypo-P
TX-SN
TX-P100
TD-S100
250
130
95
72
55
36
Coomassie
1 2 3 4 5

Figure 4.7 Fractionation of SETD2 immunoforms and other proteins in DDC-derived extracts

(A) p66^{C} and p66^{CT21} co-fractionate in exclusive DDC extracts: Quiesced 10T½ cells were fractionated by the DDC method (**Figure 4.5**) and extracts were precipitated with TCA. Samples for SDS-PAGE were resolubilised in urea/SDS-PAGE loading buffer at room temperature **(i, iii–vii, ix)** or heated at 95°C for 10 minutes **(ii)**. Equivalent proportions of each fraction were separated by 8% SDS-PAGE **(i–vii, ix)** or 15% AU-PAGE **(viii)** and either stained with Coomassie **(viii, ix)** or immunoblotted with the following antibodies: α-SETD2 CT21 **(i)**; α-SETD2 C-terminal **(ii–iv)**; α-RNAPII **(v)**; α-Lamin B **(vi)**; or α-Nucleophosmin **(vii)**. Two forms of RNAPII, II$_O$ and II$_A$ are detected, corresponding to hyperphosphorylated and hypophosphorylated polymerase, respectively.

(B) α-SETD2 CT21 antibody detects additional proteins in discrete DDC fractions: Equivalent proportions of extracts prepared as in **(A-i)** were separated by 8% SDS-PAGE and transferred to PVDF in Towbin **(i)** or SDS **(ii)** transfer buffers, for immunoblotting with α-SETD2 CT21 antibody.


(A)
Hypo-SN
TX-S100
TD-S100
TD-P100
TX-P
α-SETD2 CT21
72
55
p66CT21
(i)
α-SETD2 C-term
72
55
p66C
(ii)
p290SETD2
(iii)
p140C
(iv)
α-RNAPII
IIO
IIA
(v)
α-Lamin B
(vi)
α-Nucleophosmin
(vii)
Histones
(AU-PAGE, Coomassie)
2
1
0
H3
H2B
H2A
(viii)
Coomassie
250
130
95
72
55
36
Vimentin
Actin
(ix)
1 2 3 4 5
(B)
Hypo-SN
TX-S100
TD-S100
TD-P100
TX-P
250
130
95
72
55
36
p130CT21
p90CT21
p66CT21
1 2 3 4 5
(i) Towbin
Hypo-SN
TX-S100
TD-S100
TD-P100
TX-P
250
130
95
72
55
36
p130CT21
p66CT21
p44CT21
1 2 3 4 5
(ii) SDS

For $p66^{C}$, the majority distributes in TD-S^{100} with a smaller proportion in the TX-S^{100} fraction (**Figure 4.7A-ii**, lanes 2–3).

4.3.3. p66 is detected by two C-terminal region α-SETD2 antibodies

During the development of the DDC procedure, we acquired an additional affinity-purified α-SETD2 antibody: the α-CT21 antibody (Everest Biotech), which was raised against a peptide corresponding to the near C-terminus of SETD2 (**Figure 1.2A**). In the five DDC fractions, the α-CT21 antibody detected several proteins with characteristic fractionation properties (**Figure 4.7B**).

Most significantly, a protein of ≈67 kDa was detected exclusively in TX-S^{100} and TD-S^{100} fractions with greater enrichment in the latter (**Figure 4.7B-i/ii**, lanes 2 and 3). This protein thus shares identical DDC fractionation characteristics with $p66^{C}$. Furthermore, when analysed by nuclease-salt extraction, both proteins co-fractionated in high-salt and salt-resistant material (*cf.* **Figure B.1** in Appendix B). Since there is no significant difference in the calculated molecular masses of $p66^{C}$ and the ≈67 kDa protein ($p = 0.01$; Student's *t*-test), the latter will be named $p66^{CT21}$.

Additionally, α-CT21 antibody detected $p130^{CT21}$ exclusively in the TD-P^{100} fraction (**Figure 4.7B**, lane 4), indicating incorporation into a complex resistant to both Triton X-100 and sodium deoxycholate. Towbin transfer promoted detection of $p90^{CT21}$ in the TX-P fraction (**Figure 4.7B-i**, lane 5) whereas SDS transfer facilitated detection of $p44^{CT21}$, which was present in both Hypo-SN and TD-P^{100} fractions (**Figure 4.7B-ii**, lanes 1 and 4). Curiously, no obvious protein corresponding to $p290^{SETD2}$ was detected. The only signals above 250 kDa fractionated in the Hypo-SN and TX-S^{100} fractions (**Figure 4.7B-i,** lanes 1 and 2), inconsistent with fractionation of $p290^{SETD2}$ in TX-P (**Figure 4.7A-iii**, lane 5).

Without peptide competition data, however, the specificity of these bands is uncertain. None of the α-CT21-reactive proteins were substantially or reproducibly depleted by RNAi (*cf.* **Figure B.2** in Appendix B). However, detection of ectopically expressed constructs

demonstrated that the α-CT21 antibody does indeed recognise SETD2 sequence (*cf.* **Figure 6.2** in Chapter 6).

Do $p66^{C}$ and $p66^{CT21}$ represent the same protein? Why not just use the CT21 antibody to detect p66 for proteomic analysis? Ideally, analysis of p66 with in-house antisera is more resourceful and avoids batch-to-batch variability of commercial antibodies, but most importantly, only the α-C-terminal reagent detects a protein with such an unusual denaturation requirement. Despite this characteristic being a hindrance for proteomic identification by conventional techniques, $p66^{C}$ is more interesting as a protein displaying this characteristic. Since current data do not confirm equivalence of $p66^{C}$ and $p66^{CT21}$ (although they suggest it), it would be unwise to monitor further p66 purification and characteristics by α-CT21 antibody alone, since such a protein may not co-fractionate with $p66^{C}$. Since the thermo-urea denaturation characteristics of $p66^{C}$ is of such rarity (as far as we are aware), identification of methods to further analyse this protein, specifically, would be useful—not only for $p66^{C}$ identification, but also for future studies aiming to understand the biochemical basis behind thermo-urea denaturation and whether it may be important in SETD2 biology.

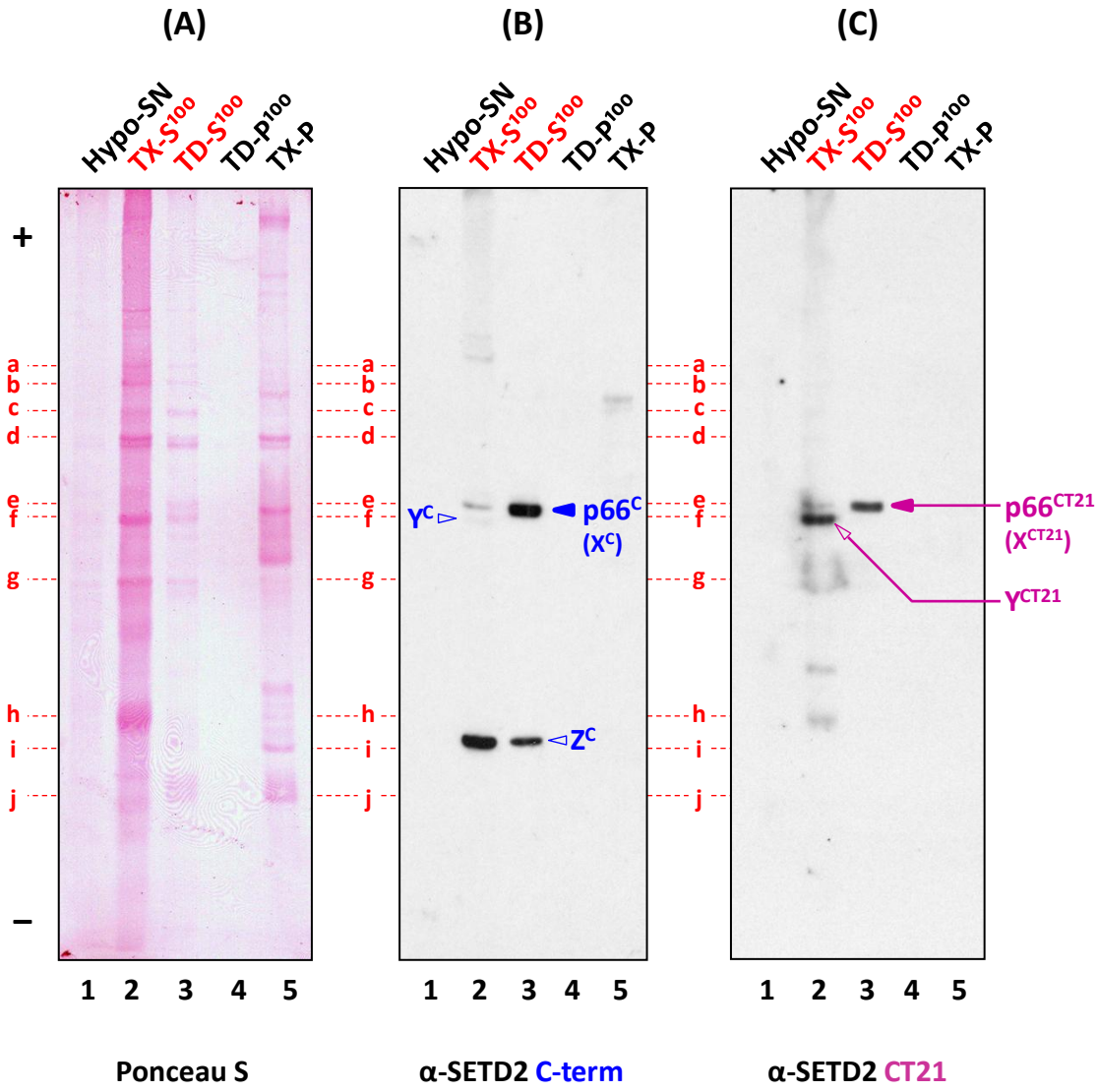
4.3.4. AU-PAGE permits identification of $p66^{C}$ by mass spectrometry in the absence of thermo-urea denaturation

Acetic acid–urea polyacrylamide gel electrophoresis (AU-PAGE) separates proteins according to their charge and mass through differences in charge-dependent electrophoretic mobility and polyacrylamide gel–mediated sieving, respectively (Panyim & Chalkley, 1969). AU-PAGE has been extensively used to characterise charge-isoforms of histones that differ in phosphorylation and acetylation (Barratt *et al.*, 1994a; Hazzalin & Mahadevan, 2005), and has also been employed for analysis of ribosomal proteins and high-mobility group proteins (Glover, 1982; Prymakowska-Bosak *et al.*, 2001).

The five DDC fractions were resolved by AU-PAGE and immunoblotted with α-SETD2 antibodies (**Figure 4.8**). Since no convenient molecular weight markers exist for AU-PAGE,

Figure 4.8 Two C-terminal-region SETD2 antibodies detect similar proteins in p66$^{C/CT21}$-enriched fractions after separation by AU-PAGE

Quiesced 10T½ cells were sequentially extracted by the DDC method, and the five resulting fractions in pellet form were precipitated with TCA and dissolved in AU-PAGE loading buffer at room temperature. Equivalent proportions of each fraction were separated by 10% AU-PAGE and transferred to PVDF membrane. After Ponceau S staining **(A)** to visualise total protein content, membranes were immunoblotted with α-SETD2 C-terminal **(B)** or α-SETD2 CT21 **(C)** antibodies. Since no convenient molecular weight markers exist for AU-PAGE, the major Ponceau S–stained proteins are annotated alphabetically according to their distance from the anode (+).


(A)
(B)
(C)
Hypo-SN
TX-S100
TD-S100
TD-P100
TX-P
+
−
a
b
c
d
e
f
g
h
i
j
γC
p66C
(XC)
ZC
p66CT21
(XCT21)
γCT21
1 2 3 4 5
Ponceau S
α-SETD2 C-term
α-SETD2 CT21

protein migration is designated arbitrarily by reference to conspicuous Ponceau S–stained proteins, which are labelled alphabetically *a–j* according to distance from the anode (**Figure 4.8A**). The α-C-terminal antibody detected protein X^{C}, which migrated at ≈*e* and fractionated predominantly in TD-S^{100} (**Figure 4.8B**, lane 3). A small amount of X^{C} was detected in the TX-S^{100} fraction, which also contained a faint band, Y^{C}, which migrated slightly further at ≈*f* (**Figure 4.8B**, lane 2). Additionally, the α-C-terminal antibody strongly detected Z^{C}, a faster migrating species at ≈*i* which distributed between TX-S^{100} and TD-S^{100} fractions (**Figure 4.8B**, lane 2). The α-CT21 antibody detected protein X^{CT21} at ≈*e* that predominantly fractionated in TD-S^{100} with a small amount present in TX-S^{100} (**Figure 4.8C**, lanes 2 and 3) similar to X^{C}. This antibody also strongly detected Y^{CT21} at ≈*f*, which was present exclusively in TX-S^{100}, similar to Y^{C} (**Figure 4.8C**, lane 2).

Both X^{C} and Z^{C} mirror the fractionation of $p66^{C/CT21}$ by SDS-PAGE. However, only protein X appears to be co-detected by both α-C-terminal and α-CT21 antibodies. Given co-detection and recapitulation of $p66^{C/CT21}$ fractionation as observed by SDS-PAGE, protein $X^{C/CT21}$ likely represents $p66^{C/CT21}$. Confirmation of the molecular weight of protein $X^{C/CT21}$ will be analysed by two-dimensional gels in section 4.3.5.

Protein Z^{C} is unlikely to represent $p66^{C/CT21}$ detected by SDS-PAGE since α-CT21 antibody did not detect any protein migrating at ≈*i*. Considering $p66^{C}$ is the only protein detected in TX-S^{100} and TD-S^{100} fractions by α-C-terminal antibody after SDS-PAGE, Z^{C} may plausibly represent a degraded form of X^{C}.

The identity of the Y^{C} and Y^{CT21} proteins are presently unclear, but their co-migration and co-fractionation suggest possible identity. Considering AU-PAGE separates by charge and size, these proteins may represent charge isoforms of X^{C} and X^{CT21}, respectively.

The ability to apparently detect the $p66^{C}$ protein with α-C-terminal antibody without recourse to thermo-urea denaturation is important, since AU-PAGE analysis avoids carbamylation and thus permits downstream proteomic identification. But does the apparent $p66^{C}$ protein detected after AU-PAGE (*i.e.* X^{C}) actually correspond to a ≈66 kDa protein?

4.3.5. Two-dimensional AU/SDS-PAGE confirms p66^{C} detection in the absence of thermo-urea denaturation

Similar to p66^{C} and p66^{CT21} detected after SDS-PAGE, the co-fractionation and co-migration by AU-PAGE of X^{C} and X^{CT21} proteins detected by α-C-terminal and α-CT21 antibodies, respectively, suggested that these proteins were likely related to p66$^{C/CT21}$. To identify the molecular mass of X^{C} and X^{CT21} (and thus assess correspondence to p66$^{C/CT21}$), the TD-S^{100} fraction was analysed by two-dimensional AU/SDS-gel electrophoresis, employing AU-PAGE in the first dimension, followed by SDS-PAGE in the second dimension. The whole TD-S^{100} fraction from an entire confluent plate of 10T½ cells (≈5 × 10^{6} cells) was required to reliably visualise proteins by Coomassie-stain after resolution by 2D-AU/SDS-PAGE (**Figure 4.9A-iii**). Two points are worth noting: Firstly, since this technique only analysed a specific segment of the first-dimension AU gel (**Figure 4.9A-i**: spanning proteins *a–g*, and containing X^{C} and X^{CT21}), most proteins below ≈45 kDa and above ≈130 kDa are excluded from the 2D gel. Secondly, because samples are initially solubilised without SDS, the complete range of SDS-solubilised proteins visualised on one-dimensional SDS gels (**Figure 4.9A-iii**) is unlikely to be present on 2D gels.

For 2D-AU/SDS-PAGE analysis, the α-C-terminal antibody detected a single spot at ≈67 kDa (**Figure 4.9B-iii**) that aligned with X^{C} detected after AU-PAGE (**Figure 4.9B-i**) and p66^{C} after SDS-PAGE (**Figure 4.9B-ii**), indicating that X^{C} corresponds to p66^{C}. Similarly, the α-CT21 antibody also detected a single spot at ≈67 kDa (**Figure 4.9C-iii**) that aligned with X^{CT21} detected after AU-PAGE (**Figure 4.9C-i**) and p66^{CT21} after SDS-PAGE (**Figure 4.9C-ii**), indicating that X^{CT21} corresponds to p66^{CT21}. Furthermore, both antibodies detect a faint 'comet-tail' at the anodic side of the main immunoreactive spot (**Figure 4.9**: **B** and **C**, panel **iii**). This is not a general property of Coomassie-stainable proteins in the TD-S^{100} fraction (**Figure 4.9A-iii**), and thus represents another specific characteristic shared by p66^{C} and p66^{CT21}.

Note, however, that while the exposures of C-terminal versus CT21 signals are comparable between experiments, the exposures within each immunoblot analysis for the different gel systems are not, because 2D-AU/SDS-PAGE required ≈20-fold greater protein load for

Figure 4.9 Two-dimensional AU/SDS-PAGE confirms detection of $p66^{C}$ and $p66^{CT21}$ in the absence of thermo-urea denaturation

(A) The TD-S^{100} fraction from a 145 mm dish of quiesced 10T½ cells was separated according to three systems: **(i)** one-dimensional resolution by 10% AU-PAGE following solubilisation (of 10% total) in AU-PAGE loading buffer at room temperature; **(ii)** one-dimensional resolution by 10% SDS-PAGE following solubilisation (of 5% total) in urea/SDS-PAGE buffer and incubation at room temperature; and **(iii)** two-dimensional AU/SDS-PAGE analysis by 10% AU- and 10% SDS-gels: the TD-S^{100} fraction derived from an entire 145 mm dish (≈5×10^6 cells) was solubilised in AU-PAGE sample buffer at room temperature. Following electrophoresis, gels were stained with Coomassie. Since no molecular weight markers exist for AU-PAGE, the major Coomassie-stained proteins separated by 1D-AU-PAGE are annotated alphabetically according to their distance from the anode (+). Arrowheads indicate 'landmark' spots that can be faintly observed by Ponceau S–staining of PVDF after western transfer. *X* represents the putative location of $p66^{C/CT21}$.

(B & C) *Continued on page 158.*

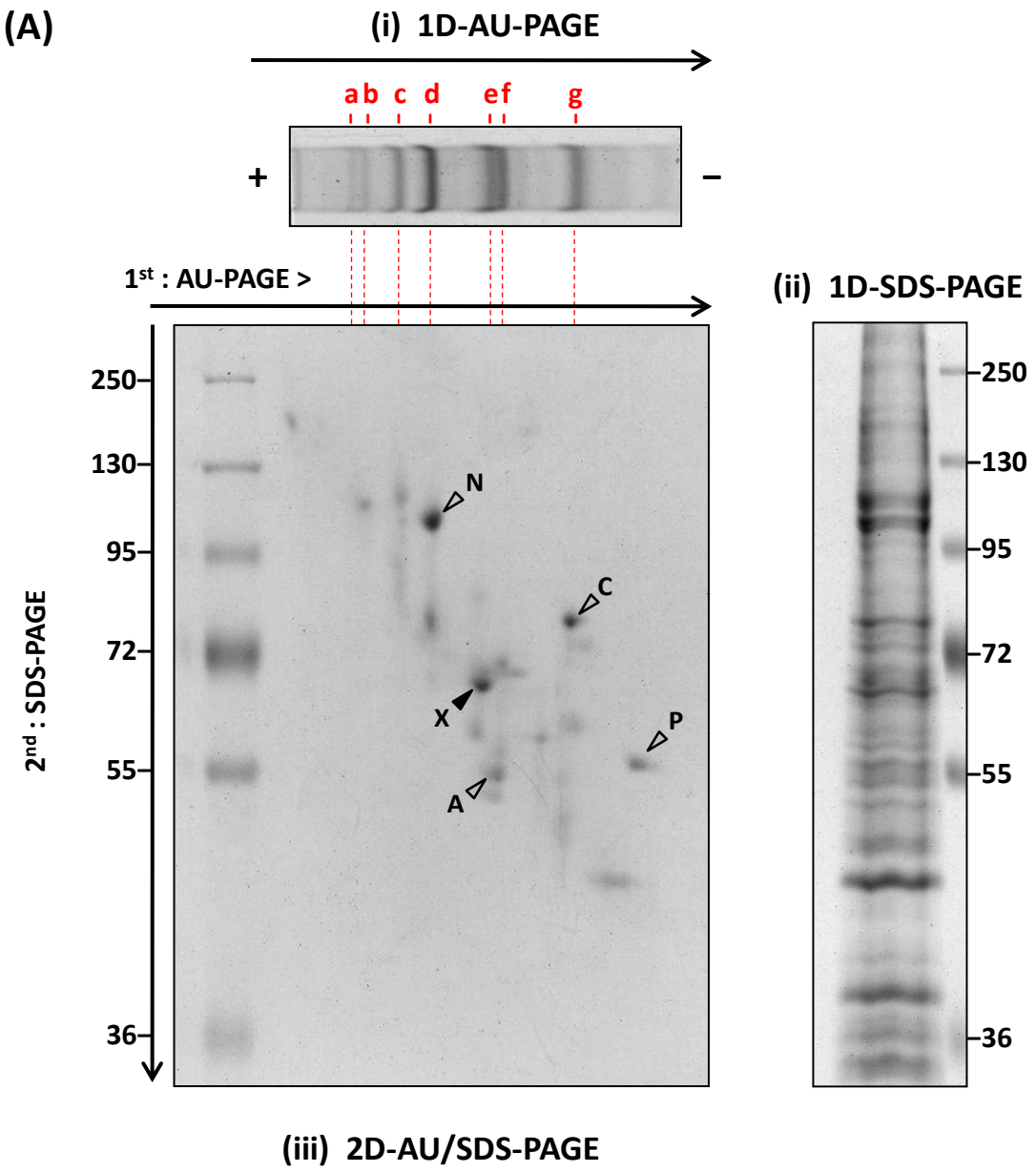
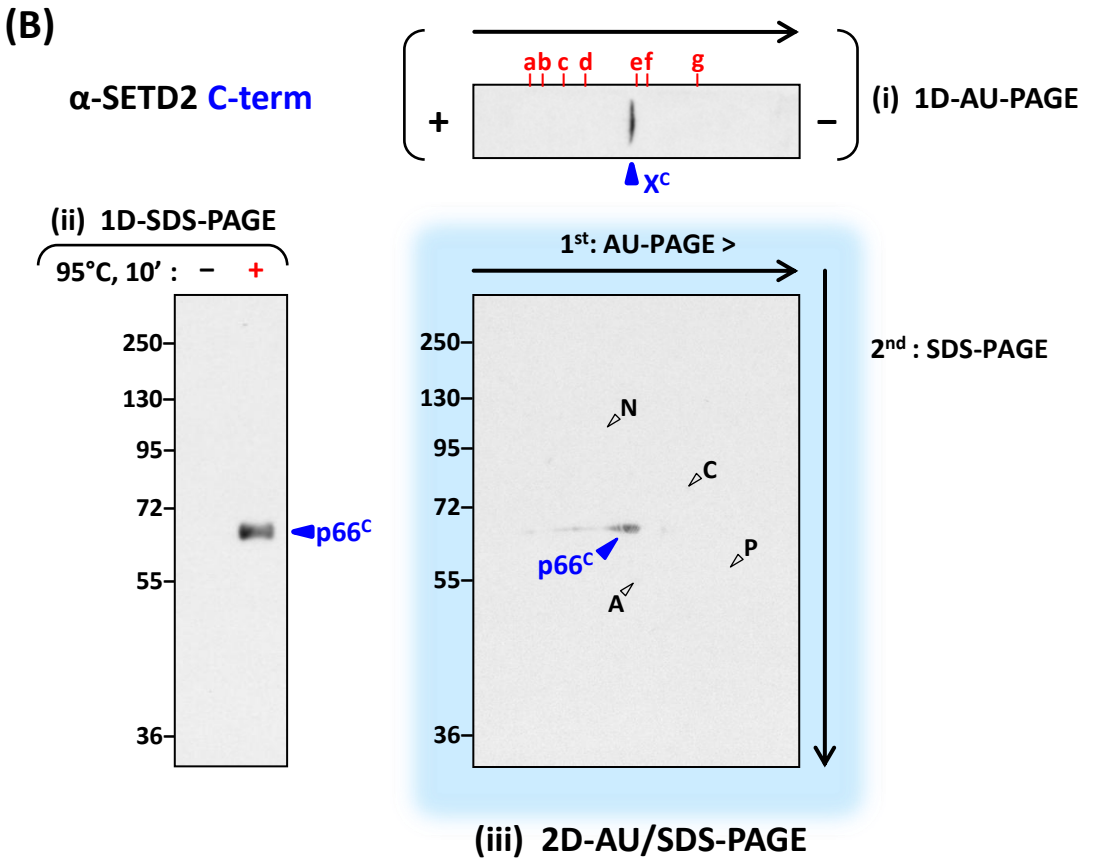
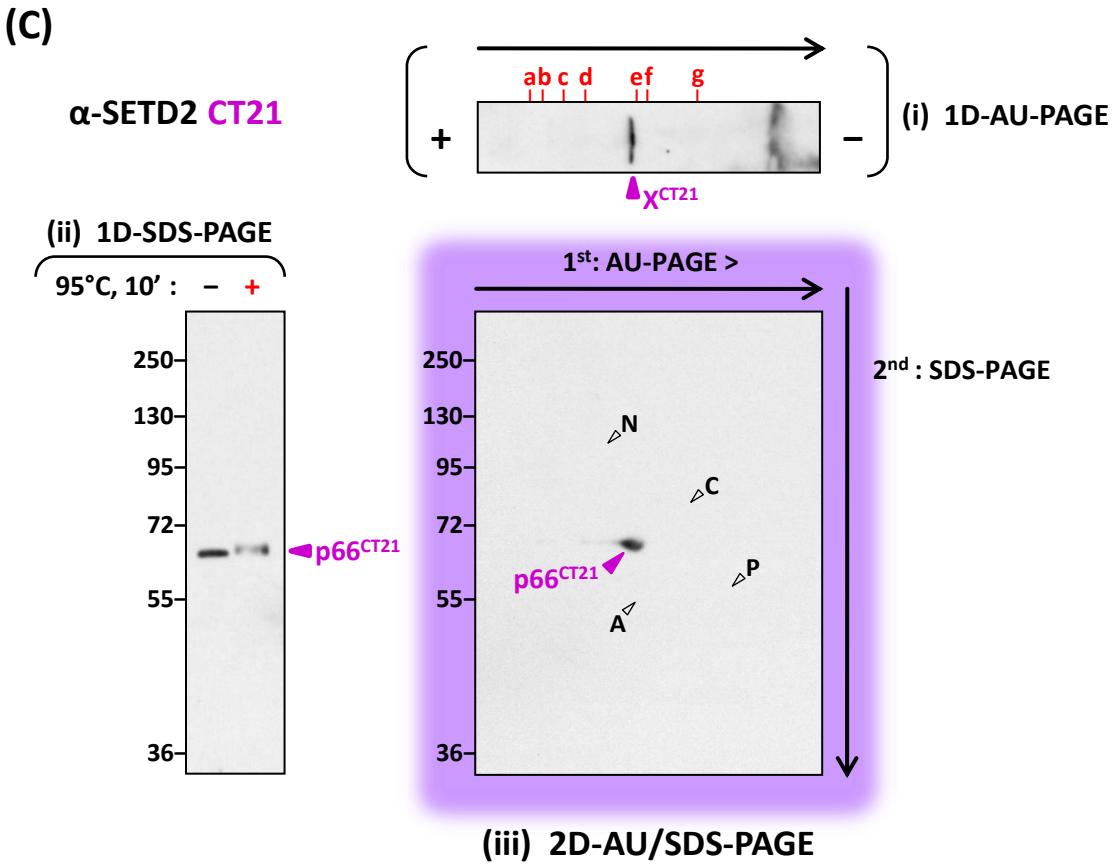

(A)
(i) 1D-AU-PAGE
a b c d e f g
+
–
1st : AU-PAGE >
(ii) 1D-SDS-PAGE
2nd : SDS-PAGE
250
130
95
72
55
36
N
C
X
P
A
(iii) 2D-AU/SDS-PAGE

Figure 4.9 Two-dimensional AU/SDS-PAGE confirms that p66^{C} can be detected by two C-terminal SETD2 antibodies in the absence of thermo-urea denaturation *(continued)*

(B) The p66^{C}-like protein detected after AU-PAGE is p66^{C}: TD-S^{100} extract from DDC fractionation of quiesced 10T½ fibroblasts was separated by the three different electrophoresis systems as described in **(A)** but with samples for one-dimensional SDS-PAGE **(i)** incubated at room temperature or 95°C for 10 minutes as indicated. Gels were transferred to PVDF in AU **(i)** or Towbin **(ii, iii)** transfer buffers. All membranes were immunoblotted with α-SETD2 C-terminal antibody. Open arrowheads indicate locations of 'landmark' spots visible by Ponceau S stain. Signal exposures are not comparable between membranes.

(C) The p66^{CT21}-like protein detected after AU-PAGE is p66^{CT21}: Replicate TD-S^{100} samples were separated by one-dimensional and two-dimensional electrophoresis as described in **(B)**. Membranes were immunoblotted with α-SETD2 CT21 antibody. Open arrowheads indicate locations of 'landmark' spots visible by Ponceau S stain. Signal exposures are not comparable between membranes.


(B)
α-SETD2 C-term
a b c d e f g
+
–
(i) 1D-AU-PAGE
XC
(ii) 1D-SDS-PAGE
95°C, 10' : – +
250
130
95
72
55
36
p66C
1st: AU-PAGE >
2nd : SDS-PAGE
N
C
P
A
(iii) 2D-AU/SDS-PAGE


(C)
α-SETD2 CT21
a b c d e f g
+
–
(i) 1D-AU-PAGE
XCT21
(ii) 1D-SDS-PAGE
95°C, 10' : – +
250
130
95
72
55
36
p66CT21
1st: AU-PAGE >
2nd : SDS-PAGE
N
C
P
A
(iii) 2D-AU/SDS-PAGE

detection; the exposures for 2D-PAGE and AU-PAGE are comparable, but those for SDS-PAGE are considerably shorter.

Two-dimensional analysis thus confirms that the X^{C} and X^{CT21} proteins detected after AU-PAGE correspond to $p66^{C}$ and $p66^{CT21}$ proteins previously observed by SDS-PAGE. The co-migration of $p66^{C}$ and $p66^{CT21}$ by two different electrophoresis systems further supports the assertion that they likely represent the same—or at least highly-related—protein(s), suggesting the existence of a putative C-terminal-specific SETD2 isoform. Furthermore, this technique confirms that $p66^{C}$ can be detected without thermo-urea denaturation and localised within the gel, under conditions amenable for downstream proteomics.

The 2D-AU/SDS-PAGE system demonstrates that, in the absence of thermo-urea denaturation, $p66^{C}$ definitively exists within the gel and its location can be pinpointed. Previously, the location of $p66^{C}$ on an SDS gel could not be rigorously determined, because it was only observable after thermo-urea denaturation—a condition that causes carbamylation and thus renders identification by mass spectrometry or N-terminal sequencing impossible. The 2D-AU/SDS-PAGE system permits rigorous localisation of $p66^{C}$ at finer resolution in the absence of thermo-urea denaturation, and thus under conditions amenable for proteomic identification.

4.3.6. Mass spectrometry identifies CKAP4 as the predominant constituent of the $p66^{C/CT21}$-enriched spot

A critical initial step for protein identification from two-dimensional gels is definitive confirmation of the exact location of the protein of interest. Proteins transferred from the current 2D-gel system to PVDF membrane stained poorly with Ponceau S: even with a 10-fold increase in protein load, only the 'landmark' spots (A/C/N/P) that were detected on Coomassie-stained gels were visible. Increasing sample load improved staining but led to a decrease in spot resolution.

The Coomassie-stained profile contains a cluster of proteins that reflected potential locations of $p66^{C/CT21}$, estimated from the location of the $p66^{C/CT21}$-immunoblot signals

relative to the landmark spots visible on PVDF (**Figure 4.10B**). However, differential swelling of gels that are processed for Coomassie-staining and those that are transferred to PVDF, can lead to errors in judgment in pinpointing the exact location of the protein corresponding to the immunoblot signal—especially with an uncharacterised protein of unknown abundance. Furthermore, even when handled with utmost care, gels are susceptible to distortions in the horizontal dimension; thus obscuring the exact location of $p66^{C/CT21}$ between Ponceau-stained membranes and Coomassie-stained gels.

Repeated preparation of two-dimensional gels produced a reproducible cluster of spots that permitted prediction of a candidate spot containing $p66^{C/CT21}$ (**Figure 4.10B**). To confirm $p66^{C/CT21}$ location and thus ensure the correct gel segment was excised for downstream MS analysis, two identical aliquots of TD-S^{100} fraction were resolved in parallel lanes on the same 2D-AU/SDS-gel, and after Coomassie-staining (**Figure 4.10A**), the candidate $p66^{C}$-enriched spot was excised for MS analysis (**Figure 4.10C**, lane 2).

After excision of the candidate $p66^{C/CT21}$-enriched spot (**Figure 4.10C**, lane 2), confirmation that this gel segment contained $p66^{C/CT21}$ was obtained by immunoblotting the Coomassie-stained gel with α-SETD2 C-terminal antibody. This revealed that $p66^{C}$ was detected in the intact 2D-gel but absent in the gel from which the candidate spot had been excised (**Figure 4.10D**, lane 1 versus 2), confirming that the correct spot corresponding to $p66^{C}$ had also been analysed by MS.

Curiously, immunoblotting the Coomassie-stained gel with α-SETD2 C-terminal antibody revealed another protein migrating at ≈85 kDa, which was detected at similar intensity to $p66^{C}$ (**Figure 4.10D**). The identity of this protein is unclear but may represent cross-contamination of $p80^{C}$ from the initial hypotonic supernatant (**Figure 4.2B-iii**, lane 1). Alternatively, extensive incubation with Coomassie stain and destaining solution before transfer to PVDF may have exposed additional immuno-reactive epitopes. Nevertheless, this result is not representative of immunoblots prepared by the standard 2D-gel method (**Figure 4.9B**) and since the experiment in **Figure 4.10** was only performed once, I currently reserve judgment on the significance of this additional ≈85 kDa species.

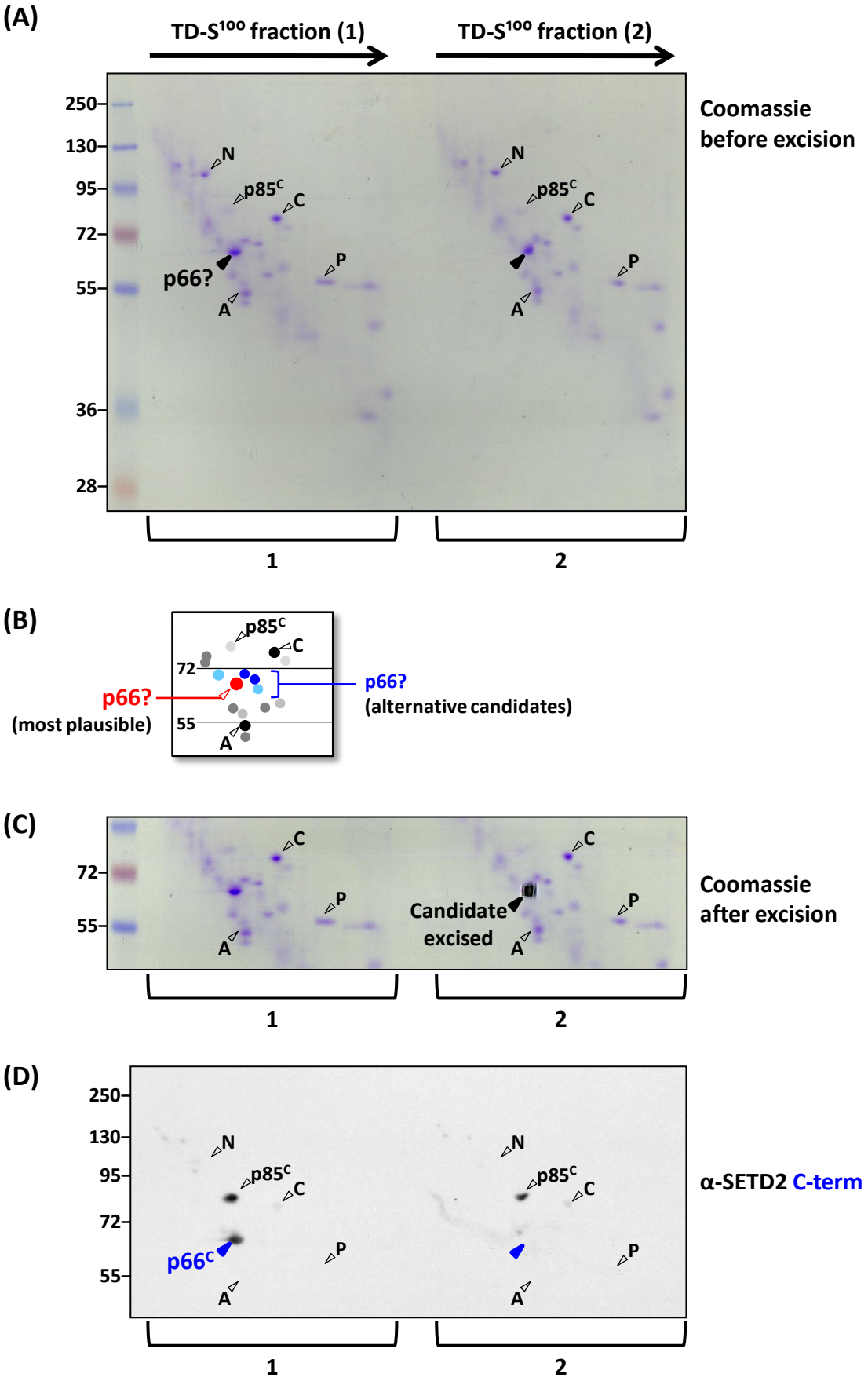
Figure 4.10 Localisation of p66$^{C/CT21}$ on a Coomassie-stained 2D-AU/SDS-gel

(A) Putative location of p66^{C} on Coomassie-stained 2D-AU/SDS-polyacrylamide gel of TD-S^{100} fraction: Two replicate TD-S^{100} extracts each deriving from DDC fractionation of an entire 145 mm plate of quiesced 10T½ cells were separated by 2D-AU/SDS-PAGE and the resulting gel was stained with Coomassie. The arrowheads indicate 'landmark' spots and the predicted location of p66$^{C/CT21}$ based on comparison of multiple Coomassie-stained gels against Ponceau S–stained membranes immunoblotted with α-SETD2 C-terminal antibody.

(B) Schematic representation of protein spots near predicted location of p66^{C}: spot shading approximates Coomassie-stain intensity. The red spot represents the most likely location of p66$^{C/CT21}$ but cyan and blue spots represented additional plausible candidates.

(C) Excision of p66$^{C/CT21}$ candidate spot from 2D-gel: Coomassie-stained gel from **(A)** after excision of the most plausible p66$^{C/CT21}$ spot from lane 2.

(D) Gel excision eliminates p66^{C}: The whole 2D gel was completely destained and transferred to PVDF in Towbin buffer. Immunoblotting using α-SETD2 C-terminal antibody revealed loss of p66^{C} in lane 2 at the position of gel excision.


(A)
TD-S100 fraction (1)
TD-S100 fraction (2)
250
130
95
72
55
36
28
N
p85C
C
P
A
p66?
Coomassie
before excision
1
2
(B)
p85C
C
72
55
A
p66?
(most plausible)
p66?
(alternative candidates)
(C)
72
55
C
P
A
Candidate
excised
Coomassie
after excision
1
2
(D)
250
130
95
72
55
N
p85C
C
P
A
p66C
α-SETD2 C-term
1
2

The $p66^{C/CT21}$-enriched gel fragment, in addition to landmark spots (**Figure 4.10A**, labelled A/C/N/P), were digested *in situ* with trypsin and proteins identified by MALDI–TOF MS (matrix-assisted laser desorption ionisation–time-of-flight mass spectrometry). The landmark spots are major constituents of the TD-S^{100} fraction and corresponded to abundant cellular proteins (detailed in **Table 4.1** and **Figure 4.11C**).

The major constituent of $p66^{C/CT21}$-enriched gel segments from two independent biological replicates was identified as cytoskeletal-associated protein 4 (CKAP4) (**Table 4.1**). No peptides corresponding to SETD2 sequence were detected in the elution or unbound fractions of C18-purified tryptic peptides.

Nano–ultra-performance liquid chromatography (UPLC)–MS/MS, which provides greater peptide resolution than MALDI–TOF, was used to identify chymotryptic peptides from a third biological sample. This technique identified CKAP4 as the major component and ATAD3 as a secondary constituent of the $p66^{C/CT21}$-enriched gel segment. No peptides were identified as SETD2 using alternative MASCOT searches with available protein and EST databases, and variable post-translational modifications.

Figure 4.11 Mass spectrometry–based identification of select proteins in the $p66^{C/CT21}$-enriched TD-S^{100} fraction resolved by two dimensional electrophoresis

The Coomassie-stained two-dimensional AU/SDS-PAGE gel **(C)** and corresponding one-dimensional gels **(A & B)** are reproduced from **Figure 4.9A** with annotation of proteins identified by mass spectrometry. Nucleolin and CKAP4 were identified from comparable spots from two and three independent biological experiments, respectively.

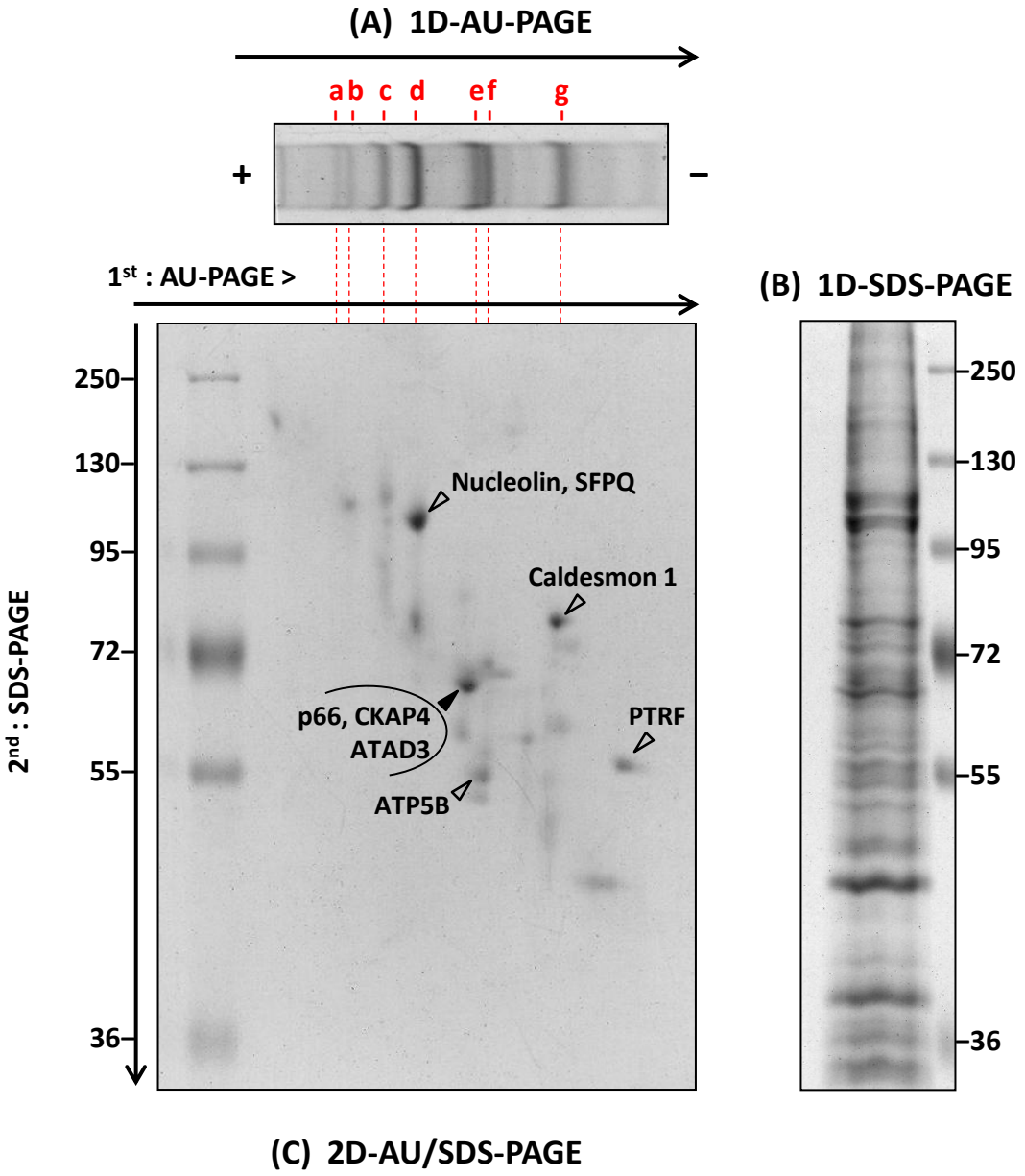

(A) 1D-AU-PAGE
a b c d e f g
+
–
1st : AU-PAGE >
(B) 1D-SDS-PAGE
2nd : SDS-PAGE
250
130
95
72
55
36
Nucleolin, SFPQ
Caldesmon 1
p66, CKAP4
ATAD3
PTRF
ATP5B
250
130
95
72
55
36
(C) 2D-AU/SDS-PAGE

Table 4.1 – Proteins identified in TD-S^{100} fraction by mass spectrometry

Gel spot	Protein	Description	MS technique (replicate)	Unique peptides	Protein copy no. (molecules/cell)*
X (contains p66$^{C/CT21}$)	**CKAP4** Cytoskeletal-associated protein 4	Integral membrane protein of the rough ER involved in sheet versus tubule formation (Schweizer *et al.*, 1995; Shibata *et al.*, 2010) and restricts lateral diffusion of translocon-ribosome complexes (Nikonov *et al.*, 2007). May connect ER membrane directly with microtubules (Klopfenstein *et al.*, 1998; Farah *et al.*, 2005).	MALDI (1) MALDI (1):FT MALDI (2) Nano	6 14 4 20	367,000
	ATAD3 ATPase family AAA domain–containing protein 3	ATPase protein present in mitochondria; involved in organisation of mitochondrial genome; anti-apoptotic function (He *et al.*, 2007; Fang *et al.*, 2010).	Nano	1	36,000
C	**Caldesmon 1**	Binds to calmodulin and regulates actin and myosin (Wang, 2001).	MALDI	7	42,000
N	**Nucleolin**	Associated with chromatin and pre-ribosomal particles; rRNA transcription, ribosome formation; also RNAPII transcription complex also containing CDK9 and SUPT5H (Parada & Roeder, 1999). Also found in cytoplasmic mRNP granules (Jonson *et al.*, 2007).	MALDI (1) MALDI (2)	4 3	881,000
	SFPQ Splicing factor, proline- and glutamine-rich	Binds to intronic pyrimidine-rich sequence and required for spliceosome formation (Gozani *et al.*, 1994). May also regulate transcription and DNA damage repair by non-homologous end joining (Akhmedov & Lopez, 2000; Mathur *et al.*, 2001).	MALDI	1	454,000
A	**ATP5B** Mitochondrial ATP synthase subunit β	Peripheral membrane protein of inner mitochondrial membrane; component of F_o of ATP synthase (Neckelmann *et al.*, 1989).	MALDI	5	1,230,000
P	**PTRF** Polymerase I and transcript release factor	Required for RNAPI transcription (Jansa *et al.*, 1998; Jansa *et al.*, 2001), but is also important for formation and organisation of caveolae (Hill *et al.*, 2008; Liu *et al.*, 2008).	MALDI	2	421,000

Proteins identified by nano-UPLC-MS/MS ($p < 0.05$). All identifications by MALDI-TOF-MS/MS based on >99.9% confidence interval (total ion). Abbreviations: FT (flow-through from C18 peptide purification); MALDI (Matrix-assisted laser desorption ionisation); mRNP (messenger ribonucleoprotein); Nano-UPLC-MS/MS (Nano-ultra-performance liquid chromatography tandem mass spectrometry); RNAPII (RNA polymerase II); TOF-MS/MS (time-of-flight tandem MS).

* average values taken from (Schwanhausser *et al.*, 2011).

4.4. Discussion

4.4.1. Summary of results

- $p290^{SETD2}$ and putative SETD2 isoforms have characteristic fractionation properties after nuclease-salt extraction (**Figures 4.2 and 4.3**) or differential detergent extraction (**Figures 4.4** and **4.7**).

- Generation of a $p66^{C/CT21}$-enriched fraction for proteomic analysis and further characterisation (**Figure 4.6**).

- AU-PAGE permits detection of $p66^{C}$ by α-SETD2 C-terminal antibody without requirement for thermo-urea denaturation (**Figure 4.8**).

- Two independent C-terminal α-SETD2 antibodies detect ≈66 kDa proteins, $p66^{C}$ and $p66^{CT21}$, which co-align by 2D-AU/SDS-PAGE (**Figure 4.9**) and co-distribute in restricted subcellular fractions produced by two different sequential extraction methods (**Figures 4.7** and **4.8**; **Figure B.1** in Appendix B).

- CKAP4 is the major protein co-migrating with $p66^{C/CT21}$ by 2D-AU/SDS-PAGE, but the equivalence of the proteins is uncertain (**Figure 4.11**).

4.4.2. Subcellular fractionation of $p290^{SETD2}$ and putative SETD2 isoforms

Sequential extraction differentially fractionates proteins by exploiting differences in subcellular localisation, protein-protein interactions and inherent physical characteristics (pI, mass, solubility and hydrophobicity). Differential extraction indicates that proteins must differ in the aforementioned properties, and is thus suggestive of functional differences. The resistance of $p290^{SETD2}$ to extraction in the initial hypotonic supernatant and by 1% (*v*/*v*) Triton X-100 suggests interactions with subcellular material: its predominant co-fractionation with histones in the high-salt extract is consistent with a nuclear localisation and chromatin-binding role for a histone H3K36-methyltransferase (Edmunds *et al.*, 2008) and the ability of high salt concentration to disrupt electrostatic

interactions between DNA and DNA-binding proteins (Luger *et al.*, 1997). Population heterogeneity is suggested by distribution of a protein into more than one fraction, which may indicate differences in function: perhaps $p290^{SETD2}$ interacts differently with RNAPII during transcription elongation or at different stages of RNA splicing, which lead to differences in extractability; perhaps the different fractions represent similar interactions with RNAPII but from different chromatin environments? Curiously, in both extraction procedures, $p290^{SETD2}$ and $p140^{C}$ co-fractionate: perhaps they share similar structural features and interactions? Fractionation of $p120^{N}$ in the Hypo-SN, however, is more characteristics of cytoplasmic or membrane proteins, which are solubilised relatively easily by 0.1% (*v*/*v*) Triton X-100. One could speculate that $p140^{C}$ may contain C-terminal domains that are implicated in various interactions of SETD2 with protein and DNA (**Figure 1.2B**) whereas $p120^{N}$ may lack such domains and interactions if it represents an exclusively N-terminal-specific isoform. However, without verification of these immunoforms as *bona fide* SETD2 isoforms, and without understanding their structure and domain composition, virtually nothing can be concluded definitively about how their subcellular fractionation may relate to SETD2 function and proposed interactions.

$p66^{C/CT21}$ fractionates as if it were a component of the nuclear substructure or cytoskeleton after nuclease-salt fractionation. However, its complete extraction from a crude nuclear fraction by Triton-buffer, which fails to extract lamin B, vimentin and actin, indicates association with different subcellular structures. Triton X-100 is frequently employed to solubilise peripheral and integral membrane proteins: indeed, several $TD\text{-}S^{100}$ proteins (ATP5B, CKAP4, PTRF) are membrane-associated and are associated with mitochondria, rough ER and caveolae, respectively. Association with both membrane vesicles and microtubules is a property reminiscent of Huntingtin in its role in trafficking of endocytic vesicles (DiFiglia *et al.*, 1995; Hoffner *et al.*, 2002; Caviston *et al.*, 2007; Smith *et al.*, 2009), which may be relevant to the extractability of $p66^{C/CT21}$ and Huntingtin's proposed interaction with C-terminal domains of SETD2 (Faber *et al.*, 1998; Kaltenbach *et al.*, 2007), which are bounded by the C-terminal and CT21 epitopes, and thus putatively present in $p66^{C/CT21}$. Furthermore, proteins reactive towards α-SETD2 antibodies have been reported in purified neuroendocrine secretory vesicles, although their identity as SETD2 proteins

was not directly confirmed (Wegrzyn *et al.*, 2010). The Triton-extraction buffer also contains 5 mM EDTA, which through Mg^{2+} chelation, also likely contributes to destabilisation of interactions between proteins and nucleic acids. However, nuclease digestion and high-salt extraction as processes themselves may affect subsequent extractability, causing proteins to aggregate in the nucleus. Indeed evidence that nuclease/salt extraction may contribute towards changing nuclear interactions comes from the inability of Triton-extraction buffer to fully extract $p66^{C/CT21}$ from the pellet remaining after nuclease-salt fractionation (*cf.* **Figure B.1**).

The exact nature of the $p66^{C/CT21}$-enriched TD-S^{100} fraction is unclear—it contains nuclear proteins involved in ribosomal RNA transcription, such as nucleolin and PTRF, and a spliceosome factor, SPFQ, but also components of the rough ER (CKAP4) and mitochondria (ATP5B and ATAD3). However, nucleolin, for example, is also part of cytoplasmic messenger ribonucleoprotein granules containing untranslated protein whereas PTRF is an abundant constituent of caveolae (Foster *et al.*, 2003; Jonson *et al.*, 2007; Hill *et al.*, 2008). The extraction conditions may be sufficient to solubilise intranuclear compartments, however it is unclear the proportion of these molecules that are extracted into this specific fraction compared to the other four fractions of the DDC protocol. How subcellular fractions relate to organelles or other internal structures in previous studies is unclear due to differences in extraction conditions (temperature, buffer composition and cell type) and is complicated by certain proteins and structures distributing across multiple fractions (Hancock, 2000). More rigorous characterisation of subcellular localisation would require analysis of well-characterised markers for organelles (Mahadevan *et al.*, 1988) and independent analysis by immunofluorescence, immuno–electron microscopy or live-cell imaging.

Despite the above limitations, differential fractionation provides important diagnostic characteristics for endogenous proteins, which are useful for future experiments: for example, the differences in fractionation between endogenous $p290^{SETD2}$ and $p66^{C/CT21}$ will be used to assess the identity and behaviour of an ectopically expressed novel SETD2 isoform in Chapter 6. The separation of virtually all $p66^{C}$ into an exclusive fraction devoid

of other proteins recognised by C- and N-terminal antibodies further refutes the possibility that its observation after thermo-urea denaturation results from artefactual degradation from larger immunoforms.

4.4.3. The complexed nature of p66$^{C/CT21}$

The sedimentation of p66$^{C/CT21}$ after ultracentrifugation at 100,000*g* indicated incorporation into complexed or aggregated material. This is observed after centrifugation of three distinct extracts, which differ widely in Triton X-100 and salt concentration: (i) in high-salt buffer (1 M NaCl) (**Figure 4.3B**, lane 4); (ii) in hypotonic lysis buffer (0.1% (*v*/*v*) Triton X-100, 10 mM KCl, 1.5 mM $MgCl_2$) (**Figure 4.4B**, lane 3); and (iii) in Triton-extraction buffer (1% (*v*/*v*) Triton X-100, 100 mM NaCl and 5 mM EDTA) (**Figure 4.4B**, lane 9). It is unclear whether such complexes are comparable in nature, and whether they are all physiological or result artefactually from extraction conditions. Nevertheless, the observation of complexed p66$^{C/CT21}$ under three different conditions indicates its propensity for complex formation.

The complex may be homomeric or interact with other proteins, but is unlikely to involve histones, RNAPII or lamin B. The high-salt complex is especially intriguing given the majority of Coomassie-stainable protein does not co-sediment, indicating that complex formation or aggregation is not a general property conferred by the buffer to cellular proteins. Furthermore, selective incorporation into complexes is suggested by sensitivity to deoxycholate, which changes sedimentation behaviour of select proteins within the TX-P^{100} fraction. Intriguingly, the final C-terminal 127 residues of SETD2 display homotypic interactions in yeast two-hybrid assays, suggestive of dimerisation, independent of conserved WW domain residues and DNA-binding (Faber *et al.*, 1998). Interestingly, the yeast homolog migrates as an apparent homodimer by gel filtration (Strahl *et al.*, 2002; Schaft *et al.*, 2003).

Thermo-urea denaturation is not correlated with incorporation into complexed material that is defined by sedimentation properties: p66^{C} does not sediment in TX−DOC buffer yet still requires thermo-urea denaturation for detection (*cf.* **Figure A.3** in Appendix A).

However, this does not exclude the possibility that p66$^{C/CT21}$ may still exist in smaller complexes that do not sediment by the above conditions, and that denaturation is required for their disassembly and entry into the gel system

Another α-SETD2-reactive protein, p130^{CT21} was also incorporated into complexed material, but unlike p66$^{C/CT21}$, was not dissociated by addition of 1% (*w*/*v*) sodium deoxycholate (**Figure 4.7B**).

SETD2 has been detected by MS in murine spinal cord aggregates, resulting from solubilisation with 0.5% (*v*/*v*) Nonidet P-40, 0.25% (*w*/*v*) SDS and 0.5% (*w*/*v*) sodium deoxycholate, and ultracentrifugation (100,000*g* for 30 minutes) (Shaw *et al.*, 2008), which represents a similar—though not strictly comparable—detergent extraction procedure.

If p66$^{C/CT21}$ truly represents a SETD2 isoform, its complexed nature and epitope-detection properties may be relevant to functional aspects of Huntingtin biology, based on the following observations. Firstly, the C-terminal SETD2 epitope is located within a predicted coiled-coil region that overlaps with the downstream glutamine/proline-rich domain (**Figure 1.2A**). Secondly, aggregation-prone prions and amyloidogenic proteins frequently contain Q/N-rich domains that overlap with predicted coiled-coil motifs (Fiumara *et al.*, 2010). Thirdly, the aggregation of Huntingtin absolutely requires both its coiled-coil and polyglutamine segment (Fiumara *et al.*, 2010). Fourthly, both Huntingtin and C-terminal SETD2 epitopes co-localise *in vivo* in aggregates in HD brain samples (Passani *et al.*, 2000). Finally, the N-terminal coiled-coil–poly-Q region of Huntingtin and C-terminal regions of SETD2 interact directly *in vivo* and *in vitro* by a variety of techniques (discussed in section 1.5.5). This intriguing, yet speculative, relationship between the coiled-coil–glutamine-rich region in SETD2 (predicted to occur in p66$^{C/CT21}$) and Huntingtin protein aggregation deserves further investigation in future.

4.4.4. Proteomic identification of p66$^{C/CT21}$

The detection of p66^{CT21} did not require thermo-urea denaturation (*cf.* **Figure 4.9B-ii**). This could mean that the protein is different to p66^{C}. However, it is unknown whether the

thermo-urea denaturation requirement for $p66^{C}$ detection is a global or localised property. Given both antibodies were intentionally raised against SETD2, and that both $p66^{CT21}$ and $p66^{C}$ fractionate identically under multiple extraction conditions and co-fractionate by 2D-AU/SDS-PAGE, both proteins must: (i) be derived completely from crude nuclei; (ii) be extracted by Triton-extraction buffer in their entirety, and (iii) in such an extract, exist in a complexed form to the same extent; which can (iv) sediment by ultracentrifugation; and (v) be completely dissembled into a non-complexed form by sodium deoxycholate; and (vi) migrate with identical mass and charge-mass properties, during separation by two distinct electrophoretic systems. Given that both these proteins share all of the above properties, I believe it is reasonable to predict they represent the same protein. Since coincidental co-migration and co-fractionation of unrelated proteins is entirely possible, however, the equivalence of $p66^{C}$ and $p66^{CT21}$ still requires further support by co-immunoprecipitation or physical identification.

AU-PAGE and 2D-AU/SDS-PAGE provide systems for analysing a proportion of $p66^{C/CT21}$ (even if not representing the total population) under conditions that enable exact localisation of the protein in the gel, and avoidance of carbamylation. Lysine carbamylation requires deprotonated, uncharged amino groups. The acidic conditions of AU-PAGE, however, protonate amino groups (pH below pKa values of amino acid side-chains), and thus minimise potential carbamylation. Furthermore, acidic conditions drive the equilibrium in favour of urea versus isocyanic acid. These conditions also refute the possibility that detection of $p66^{C}$ after thermo-urea denaturation by SDS-PAGE is due to the C-terminal antibody recognising a carbamylation-derived mimotope.

One explanation for the ability to detect $p66^{C}$ by AU-PAGE without thermal denaturation is that there is an endogenous epitope modification, that occludes recognition by the antibody, that is labile when extracts are acidified or heated (in the presence of urea). Relevant to the composition of the C-terminal epitope (RGTKRPDDRYDTPTSKKKVR), tyrosine sulphation and phosphorylation of lysine and arginine residues are notable examples of acid-labile post-translational modification (Steiner *et al.*, 1980; Stone *et al.*,

2009). Additionally, there may exist other, uncharacterised, modifications to the amino acid side-chains within the C-terminal epitope.

It is unclear, however, the extent to which the population of $p66^{C}$ molecules detectable by AU-PAGE correspond to the population of $p66^{C}$ detected after thermal-urea denaturation and SDS-PAGE. Furthermore, since AU-PAGE separates protein isoforms that differ in charge (for instance, differentially acetylated or phosphorylated histones), it is also unclear whether the two SETD2 antibodies are truly recognising the same protein: the two bands (marked X and Y in **Figure 4.8**), which differ in electrophoretic mobility by AU-PAGE, may equally represent two distinct proteins that cross-react with the antibody, or differently modified forms of the same protein.

At the position of $p66^{C/CT21}$, MS detected CKAP4. However, these are unlikely to be equivalent proteins, because the immunoblot signals of $p66^{C}$ and $p85^{C}$ are comparable in intensity (**Figure 4.10D**, lane 1) yet there is no Coomassie-stained spot for $p85^{C}$ of comparable intensity to the spot that contains $p66^{C}$ and CKAP4 (**Figure 4.10A**, lane 1). This would suggest that $p66^{C}$ in isolation would be undetectable by Coomassie staining (at this loading level), and that the majority of Coomassie-stainable protein at the position of $p66^{C}$ is predominantly contributed by CKAP4 (consistent with the greater abundance of CKAP4 determined from proteomic studies as discussed below). Further evidence that argues that $p66^{C/CT21}$ is not CKAP4 comes from the observation that $p66^{C/CT21}$ displays identical electrophoretic migration by reducing and non-reducing SDS-PAGE (data not shown) yet CKAP4 migrates as dimeric and trimeric forms by non-reducing SDS-PAGE (Schweizer *et al.*, 1993). Furthermore, CKAP4 is not expressed in MRC5 cells, unlike $p66^{C}$ (Wesierska-Gadek *et al.*, 2007).

Many factors could explain the lack of detection of SETD2 sequence at the position of $p66^{C}$, such as the inability to generate MS-compatible peptide fragments, unpredictable post-translational modifications, missed peptide cleavages, and potential lack of representative peptide sequence in sequence databases (for example, if $p66^{C}$ derives from an uncharacterised alternatively spliced transcript variant). Most significantly, however,

SETD2 appears to be present in the cell at vanishingly small quantities. A recent MS approach determined that SETD2 is present at 228 average molecules/cell (within the bottom 1.5% percent-rank; median 16,000) (Schwanhausser *et al.*, 2011). This reflects an abundance that is approximately three orders of magnitude below that of CKAP4. Considering that this MS study did not distinguish peptide derivation from specific isoforms, the abundance given is the summation of $p290^{SETD2}$ plus all additional putative isoforms contributing peptide sequence detectable by MS. Thus if $p66^{C/CT21}$ did represent a SETD2 isoform, the number of molecules of $p66^{C/CT21}$ per cell would thus constitute an even smaller proportion of the value reported for total SETD2 molecules. Yeast Set2, like most transcription factors, is also present at very small quantities (217 molecules; within the 0.05% percent-rank of data-set; median ≈2300) (Ghaemmaghami *et al.*, 2003). The low abundance of SETD2 within cells may thus hinder its ability to be detected from insufficiently enriched samples. To bring SETD2 into the range of sensitivity of MS, further purification will be required. The TD-S^{100} fraction may serve as a useful starting point for this process.

4.5. Conclusion

The results of this chapter provide novel, diagnostic criteria for $p66^{C/CT21}$, $p290^{SETD2}$ and other immunoforms—especially for $p66^{C/CT21}$ in terms of its co-detection by two α-SETD2 C-terminal antibodies, co-migration on 2D-gels, and co-fractionation under multiple extraction conditions. These properties will be important to evaluate whether putative protein products of alternative transcripts directly correspond to the endogenous $p66^{C/CT21}$ protein (presented in Chapter 6).

Given the apparent low abundance of $p66^{C/CT21}$, identification by mass spectrometry from current cellular extracts may be challenging. The low-abundance of SETD2 protein reported previously in mouse fibroblasts means further biochemical purification would likely be necessary to identify $p66^{C/CT21}$ (Schwanhausser *et al.*, 2011). Even so, there is always the possibility that the inherent structure or sequence properties of $p66^{C/CT21}$ may further hinder its identification by MS even if purification to homogeneity is possible.

Therefore, as another approach to identify these putative SETD2 isoforms, the transcriptional output of murine *Setd2* gene was investigated, based on the plausibility of isoform derivation from transcript variants. The next chapter aims to characterise the transcriptional products of the murine *Setd2* gene and assess how transcriptional diversity may relate to the putative isoforms identified here.

Chapter 5 Identification of alternative forms of murine *Setd2* transcripts

5.1. Introduction

The preceding chapter suggests putative SETD2 protein isoforms may exist additional to a verified p290^{SETD2} protein responsible for H3K36 trimethylation. Since transcriptional diversity has substantial potential to encode alternative isoforms, this chapter aims to further understand the RNA profile that derives from the murine *Setd2* gene in 10T½ fibroblasts. In summary, multiple transcript variants were identified, defined by mutually exclusive, alternative splicing patterns, which have potential to generate both N-terminal- and C-terminal-specific SETD2 protein isoforms.

5.1.1. Challenges associated with defining transcript variants within complex transcriptomes

Diversity in transcript structure can be generated by alternative promoters, splicing, polyadenylation and RNA editing (Carninci *et al.*, 2005). Transcripts of high abundance, that are relatively small and possess few alternative variants, may be cloned and sequenced in their entirety with confidence (Bohmann *et al.*, 1987; Christy *et al.*, 1988). Such strategy is more challenging for large, multi-exon genes, especially if transcripts are rare and structurally diverse. For many mammalian genes, a definitive understanding of complete transcript structure is therefore limited; most transcripts are represented by models, inferred from concatenation of consistent expressed sequence tags (ESTs), partial complementary DNA (cDNAs) and computational models (Eyras *et al.*, 2004; Thierry-Mieg & Thierry-Mieg, 2006). Although flanking boundaries can be cloned by rapid-amplification-of-cDNA-ends (RACE) (Frohman *et al.*, 1988) and transcription start sites derived from 5′ cap-analysis-of-gene-expression (Barrera *et al.*, 2008), the remaining transcript structure from which such 5′/3′-segments derive remains invisible. Since direct sequencing of individual large RNA transcripts in their entirety is unfeasible at present,

full-length cDNA cloning therefore remains essential for empirical verification of complete transcript structure (Das *et al.*, 2001).

Nevertheless, knowledge of alternative fragments of transcripts is still useful; indeed, techniques like RT-PCR and RNA-sequencing allow concurrent analysis of multiple, mutually exclusive, splicing events, which may not be detected by cloning and sequencing of individual cDNAs, such as large, low abundance or non-polyadenylated transcripts. RNA-sequencing has the advantage of genome-wide analysis, but is currently limited by length of sequence reads and difficulty in aligning certain sequences across splice junctions. RT-PCR allows a directed approach, but is spatially restricted by chosen primers. The latter strategy has been employed here to investigate the prevalence of alternative splicing of *Setd2* in 10T½ fibroblasts.

5.1.2. Functional consequences of alternative transcript generation

Genome-wide bioinformatic studies predict that over 95% of human genes produce at least one alternatively spliced form (Pan *et al.*, 2008; Wang *et al.*, 2008a). The functional significance of most splicing variation is unclear, however. One view proposes that the majority of alternative splicing (AS) is simply noise due to the inherent stochastic nature of splice-site recognition and spliceosome function (Melamud & Moult, 2009a; Melamud & Moult, 2009b). However, the existence of such noise in gene expression does not necessarily imply functional insignificance (Raser & O'Shea, 2005). Alternative transcripts may produce functionally distinct protein isoforms: events may range from differential retention of a single intron in *fosB* that produces antagonistic transcription factors (Mumberg *et al.*, 1991), to the formidable complexity of *dscam* splicing in *Drosophila* that theoretically generates over 38,000 different axonal guidance receptors (Schmucker *et al.*, 2000). Alternatively, transcript variants may encode identical proteins, but UTR variation may specify different rates of turnover or translation, subcellular localisation, or recognition by micro-RNAs, for example. In parallel, AS may insert premature termination codons upstream of exon-junctions, triggering nonsense-mediated decay (NMD) (Moore & Proudfoot, 2009).

5.1.3. Current understanding of mammalian *Setd2* transcripts

Previous reports conflict on the size of major *SETD2* transcripts in human tissues: one study reported a ≈9.0 kb transcript detected in various adult and fetal tissues by northern blotting with a cDNA probe spanning exons 18–21. However, in virtually all tissues, smearing and indistinct bands of lower mass are apparent at approximately 1.5, 2.5, 4 and 5 kb (Passani *et al.*, 2000). Another study identified a predominant ≈7.5 kb expressed in various tissues using northern blotting with a probe spanning exons 16–21 (Rega *et al.*, 2001). Again, smears to lower molecular mass are present in heart, skeletal muscle and pancreas (Rega *et al.*, 2001). The discrepancy in transcript size between these studies is unclear, especially since the probes are derived from similar 3' regions.

Depending on EST alignment algorithms and stringency of analysis, different programs predict different transcript profiles. For example, Ensembl currently lists five alternative transcripts (build 61). Four alternative events are currently annotated: inclusion of alternative exon 12b, and partial intron retention events after exons 9, 10 and 18 (**Figure 5.1C**). These variations would cause significant disruption to the predicted ORF in-frame with $p290^{SETD2}$ and thus predict substantial truncation of protein structure.

5.2. Objectives

The objective of this chapter was to characterise transcript diversity of the *Setd2* gene in C3H 10T½ fibroblasts.

5.3. Terminology

I have used the following notation to abbreviate splicing patterns within transcripts:

12–14 a transcript where exon 12 splices directly to exon 14 (exon 13 is excluded).

12»14 abbreviates 12–13–14, a transcript where exon 12 splices to exon 13, which splices to exon 14.

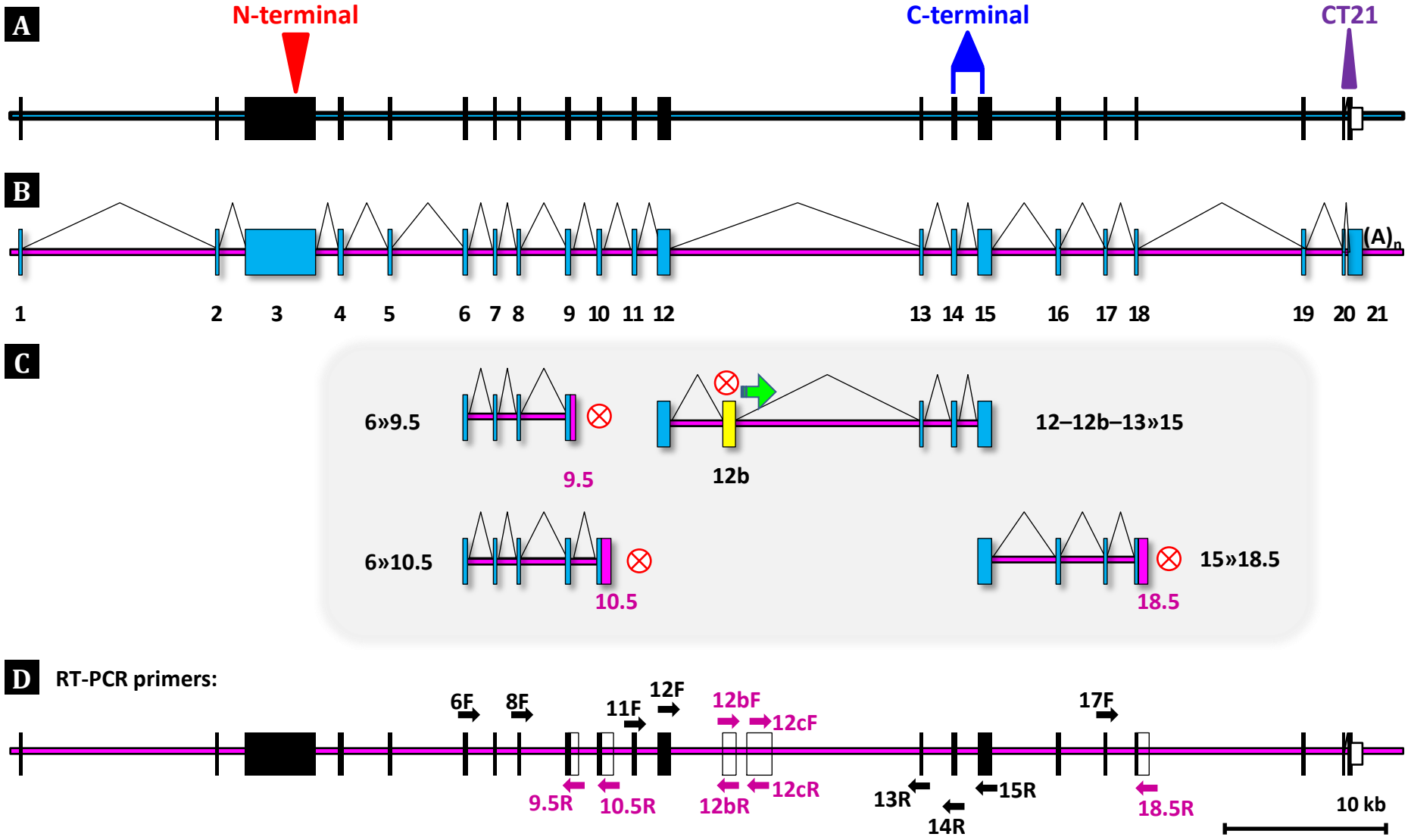
Figure 5.1 Structure of the murine *Setd2* gene

(A) Genomic structure of the murine *Setd2* locus (Chromosome 9: 110,435,101–110,521,137 forward strand, to scale)**.** Exon contributions to antibody epitopes are indicated.

(B) Exon–intron structure of the canonical *Setd2* transcript (RefSeq NM_001081340.2).

(C) Alternatively spliced *Setd2* transcripts based on EST evidence (Ensembl, build 61)**.** Pink regions indicate 'intronic' sequence. Green arrows and red crossed-circles indicate predicted sites of initiation and termination of translation, respectively.

(D) Location of RT-PCR primers: forward primers above; reverse primers below; primers within canonical exons (black); primers within 'intronic' regions (purple). Dashed boxes (not to scale) highlight regions included in alternatively spliced transcripts characterised in this chapter.


A
N-terminal
C-terminal
CT21
B
(A)n
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
C
6»9.5
9.5
12b
12–12b–13»15
6»10.5
10.5
15»18.5
18.5
D
RT-PCR primers:
6F
8F
11F
12F
12bF
12cF
17F
9.5R
10.5R
12bR
12cR
13R
14R
15R
18.5R
10 kb

8–9.5 exon 8 splices to exon 9, which continues within the 9–10 intron (instead of splicing to exon 10).

5.4. Results

5.4.1. Multiple *Setd2* transcripts are detected by northern blotting using 3' region probes

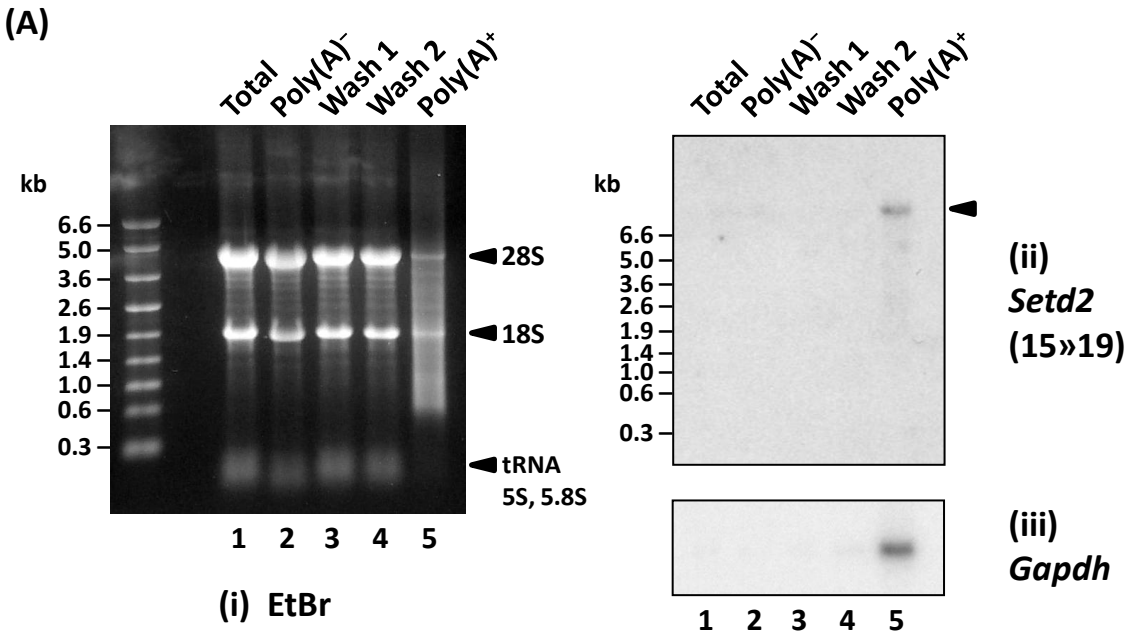
Northern blotting was employed to investigate the profile of full-length *Setd2* transcripts in murine 10T½ fibroblasts. Initially, difficulties in obtaining specific *Setd2* signals were encountered: numerous probes generated from purified RT–PCR products all heavily cross-reacted with ribosomal RNA (rRNA) under highly stringent hybridisation conditions. Such cross-reactivity was avoided with probes prepared from restriction fragments of cloned murine *Setd2*, but specific signals from total RNA remained indiscernible (**Figure 5.2-ii,** lane 1). Three strategies to improve detection of *Setd2* transcripts were therefore employed: (i) poly(A)-selection to permit analysis of greater amounts of mRNA; (ii) emetine treatment of 10T½ cells to stabilise labile transcripts (discussed in (Edwards & Mahadevan, 1992)); and (iii) parallel analysis of 293T cells, which had been previously reported to express high levels of *SETD2* (Rega *et al.*, 2001).

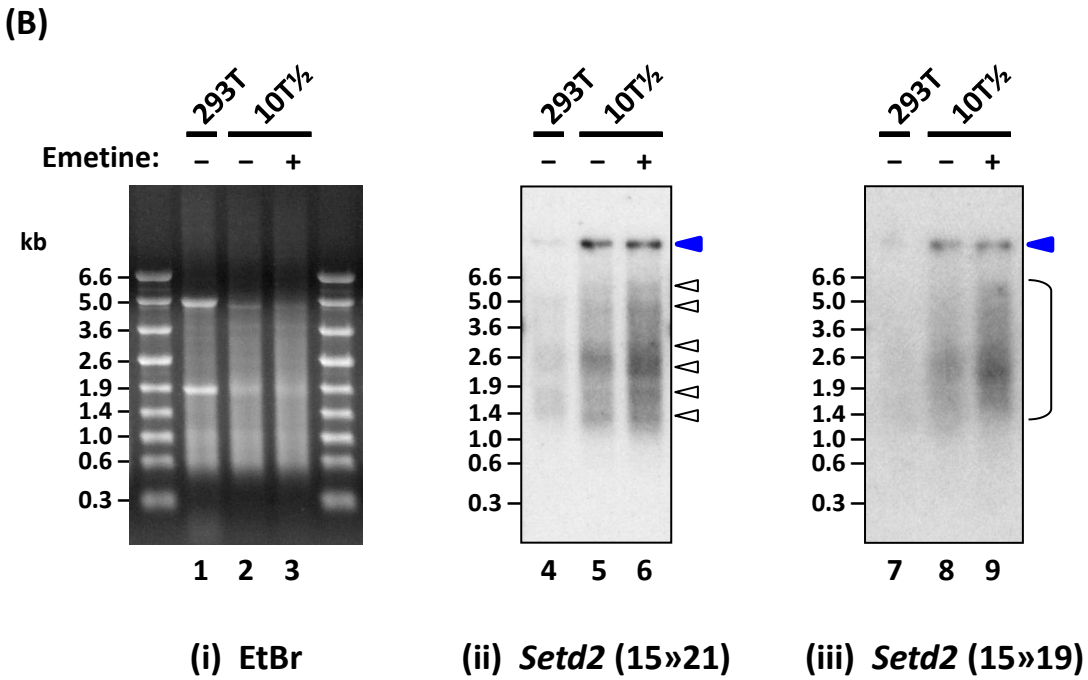
Poly(A)+ purification was verified by depletion of ribosomal and transfer RNA (**Figure 5.2A-i**) and enrichment for *Gapdh* mRNA (**Figure 5.2A-iii**) in the poly(A)+ fraction compared to total, poly(A)-minus and wash fractions. Hybridisation of a 3' region *Setd2* probe, spanning exons 15»19, detected two components in poly(A)+ RNA: a transcript larger than 6.6 kb, and an indistinct smear towards lower molecular mass; both were absent from poly(A)-minus and wash fractions (**Figure 5.2-ii**). The >6.6 kb transcript likely represents the major $p290^{SETD2}$-encoding transcript, containing exons 1»21 and predicted to be ≈8.3 kb in mouse (≈8.1/8.4 kb in humans). With increased RNA loading, several additional bands between ≈1.4 and 6.6 kb were more readily detected in 10T½ cells (**Figure 5.2B-ii/iii**)—most noticeably using a 15»21 probe (**Figure 5.2B-ii,** lanes 5 and 6). Similarly, the >6.6 kb transcript and smaller bands (≈1.4–5 kb) were also detected in 293T cells, but at lower intensity despite equal amounts of RNA loaded (**Figure 5.2B,** lanes 4 and 7). This

Figure 5.2 Multiple transcripts in poly(A)+ RNA are detected with probes against the 3' region of *Setd2*

(A) Validation of poly(A)-selection: Equal amounts (3 μg) of total RNA and poly(A)-purification fractions from quiesced 10T½ cells were separated by denaturing agarose gel electrophoresis. Ethidium bromide–staining verifies depletion of ribosomal RNA (arrowheads) from the poly(A)+ fraction **(i)**. Northern blotting using probes to a 3′ region of *Setd2* (exons 15»19) **(ii)** or to *Gapdh* **(iii)** demonstrates enrichment of mRNA transcripts in poly(A)+ fraction.

(B) Probes against the 3′ region of *Setd2* detect multiple transcripts: Poly(A)+ RNA was prepared from HEK 293T cells or quiesced 10T½ fibroblasts that had been treated with or without 10 μg/ml emetine for 2 hours. Five micrograms of poly(A)+ RNA was separated by denaturing agarose gel electrophoresis and stained with ethidium bromide **(i)** to visualise loading and integrity. After transfer to nylon membrane, RNA was hybridised with radiolabelled, double-stranded DNA probes from 3′ regions of murine *Setd2*, corresponding to exons 15»21 **(ii)** or exons 15»19 **(iii)**. Blue arrowheads indicate a common transcript with a size consistent with previous reports (≈7.5/9 kb) and the canonical gene model (≈8.4 kb).


(A)
Total
Poly(A)⁻
Wash 1
Wash 2
Poly(A)⁺
kb
6.6
5.0
3.6
2.6
1.9
1.4
1.0
0.6
0.3
28S
18S
tRNA
5S, 5.8S
1 2 3 4 5
(i) EtBr
(ii) Setd2 (15»19)
(iii) Gapdh


(B)
293T
10T½
Emetine:
– – +
kb
6.6
5.0
3.6
2.6
1.9
1.4
1.0
0.6
0.3
1 2 3
4 5 6
7 8 9
(i) EtBr
(ii) Setd2 (15»21)
(iii) Setd2 (15»19)

could reflect lower *SETD2* expression in 293T cells compared to 10T½; alternatively, mouse-derived probes may hybridise less effectively to homologous human transcripts.

Emetine appears to have little effect on the >6.6 kb transcript, whereas smaller transcripts all appear to increase in intensity: most noticeably for the ≈2.5 kb signal (**Figure 5.2B,** lanes 6 and 9), suggesting they may be labile and stabilised by translation inhibition.

5.4.2. Identification of a highly conserved genomic sequence between exons 12 and 13

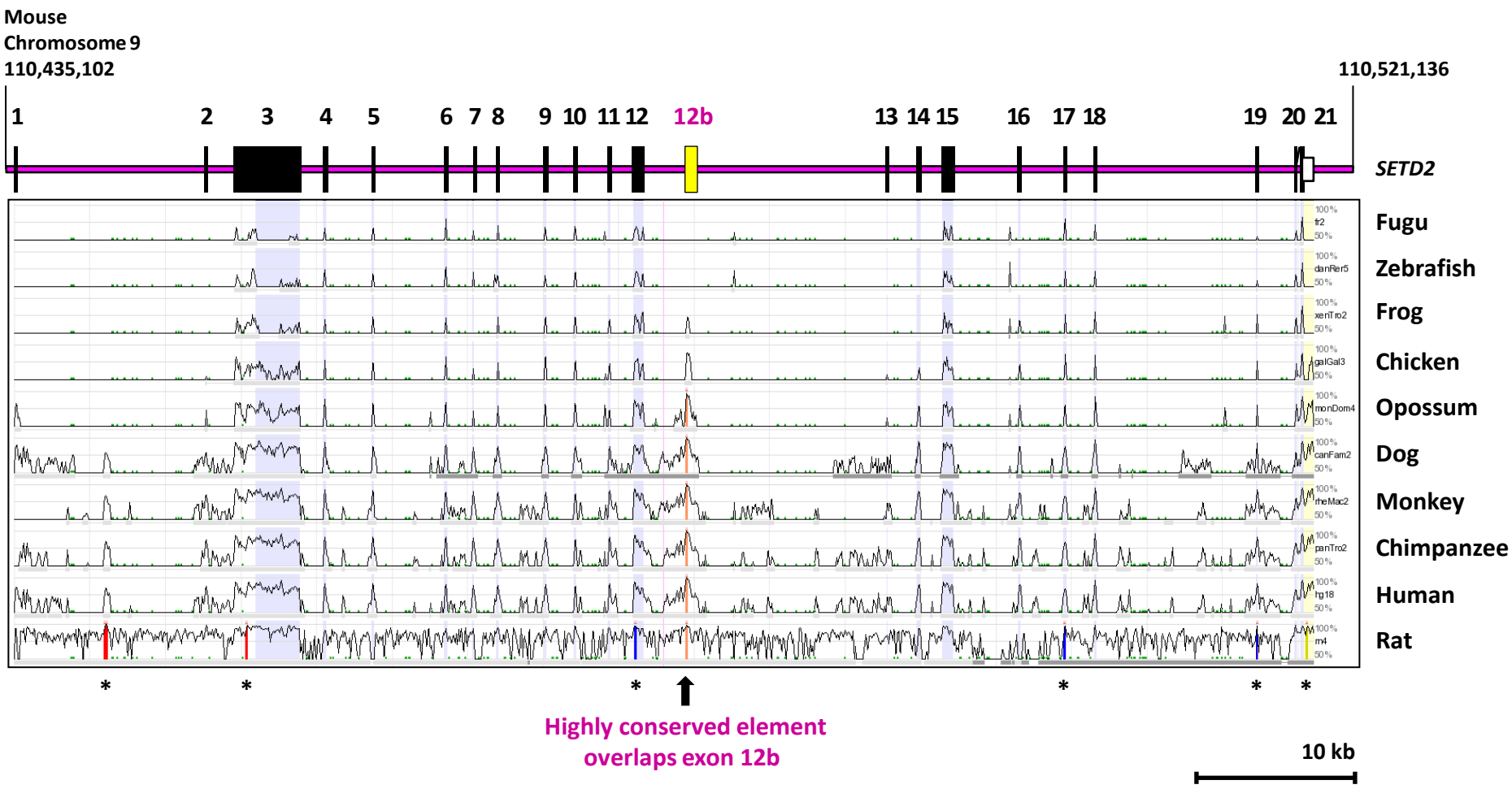
Pair-wise comparison of genomic sequence between species can highlight regions that are conserved throughout evolution. Such regions may correspond to exons, promoters and other regulatory elements. Sequence comparison of the *Setd2* locus between mouse and ten other vertebrates showed relatively moderate conservation across the 21 'constitutive' exons. A single, highly conserved element was identified between exons 12 and 13 (**Figure 5.3**), which satisfies the stringent criteria of >100 bp of 100% identity shared between mouse, rat and human sequences (Bejerano *et al.*, 2004). This region shares 153 bp identity with humans and 125 bp with rat; the region increases to ≈310 bp in both comparisons when considering 99% identity. Inspection of EST databases revealed that this region overlaps exon 12b (**Figure 5.1C**), an alternative exon present in three murine ESTs and single EST instances in human and rat databases.

Strikingly, this region is more strongly conserved than any of the constitutive exons defined by conventional models for mammalian *SETD2* genes, which is consistent with the global characteristics of alternative exons (Keren *et al.*, 2010).

Given the properties of this region described above, this region may provide an alternative exon to *Setd2* transcripts present in 10T½ cells. Its inclusion would incorporate termination codons in all three reading frames, and would thus predict premature termination of upstream open reading frames and promote the potential use of downstream initiation codons, depending on overall transcript structure and capacity for translation. For this

Figure 5.3 Identification of a highly conserved element between exons 12 and 13

Analysis of conserved genomic regions of *SETD2* using ECR Browser (Ovcharenko *et al.*, 2004). Murine *Setd2* genomic region (chromosome 9:110,435,102–110,521,136) was aligned to homologous regions of the following genomes: human, rat (*Rattus rattus*), chimpanzee (*Pan troglodytes*), Rhesus monkey (*Macaca mulatta*), dog (*Canis familiaris*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*), frog (*Xenopus laevis*), zebrafish (*Danio rerio*) and fugu (*Takifugu rubripes*). Percentage sequence identity within 100 bp–sliding windows of the genomic alignments (y-axis) is plotted against genome coordinate (x-axis). Highly conserved elements (pink) were identified as regions larger than 100 bp with 100% sequence identity between mouse, rat and humans. Asterisks indicate highly conserved elements shared between mouse and rat only.


Mouse
Chromosome 9
110,435,102
110,521,136
1 2 3 4 5 6 7 8 9 10 11 12 12b 13 14 15 16 17 18 19 20 21
SETD2
Fugu
Zebrafish
Frog
Chicken
Opossum
Dog
Monkey
Chimpanzee
Human
Rat
Highly conserved element
overlaps exon 12b
10 kb

reason, and given the evidence provided in Chapters 3 and 4 for the existence of smaller proteins containing exclusively N- or C-terminal epitopes, the prevalence of exon 12b inclusion was investigated by RT-PCR in 10T½ cells.

5.4.3. Identification of alternative *Setd2* transcripts that contain exon 12b spliced downstream to exon 13

The existence of transcripts containing exon 12b was investigated by reverse transcription–PCR (RT-PCR) of total RNA from 10T½ cells. Quiesced cells were employed to ensure a population as homogeneous as possible whilst total RNA was used to avoid bias in transcript selection through poly(A)-purification or subcellular fractionation. Although protein-encoding RNA is conventionally poly(A)+ and cytoplasmic, analysis of such a restricted RNA fraction would not permit analysis of alternative, poly(A)-minus or nuclear transcripts, which may have non-coding roles as equally important as conventional cytoplasmic poly(A)+ transcripts that may code for proteins.

For each RT-PCR reaction, RT-negative (lacking reverse transcriptase) and non-template controls (lacking cDNA) were prepared in parallel to verify that products derived solely from RNA and not from residual undigested genomic DNA or contaminating DNA. All PCR products of sufficient abundance were purified by gel extraction and verified by DNA sequencing using both forward and reverse primers. Sequences for the alternative RT-PCR products are provided in Appendix C.

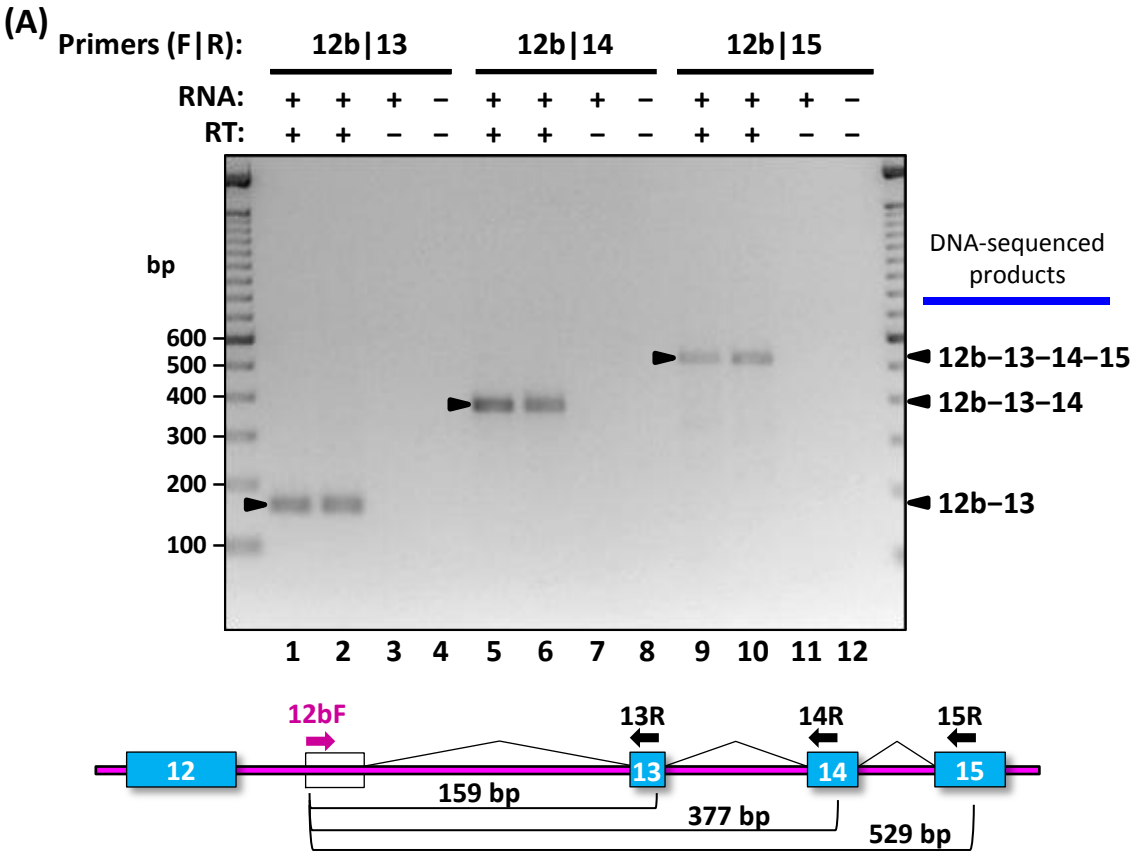
PCR primers were designed to amplify across exon 12b and downstream regions in exons 13, 14 and 15 (**Figure 5.4A**). Single PCR products were reproducibly obtained in the presence of cDNA only and migrated as the sizes predicted for splicing of exon 12b to 13 (**Figure 5.4A**, lanes 1 and 2), splicing of exon 12b to 13 and 14 (**Figure 5.4A**, lanes 5 and 6), and exon 12b to 13, 14 and 15 (**Figure 5.4A**, lanes 9 and 10). DNA sequencing confirmed product identity, demonstrating that exon 12b is including within the following contexts: (i) 12b–13 (**Appendix C-2.1**); (ii) 12b–13–14 (**Appendix C-2.2**); (iii) 12b–13–14–15 (**Appendix C-2.3**). The splicing of exon 12b to downstream exon 13 occurs via canonical GT/AG sites for the major spliceosome (**Appendix C-1.2**). No reaction produced products

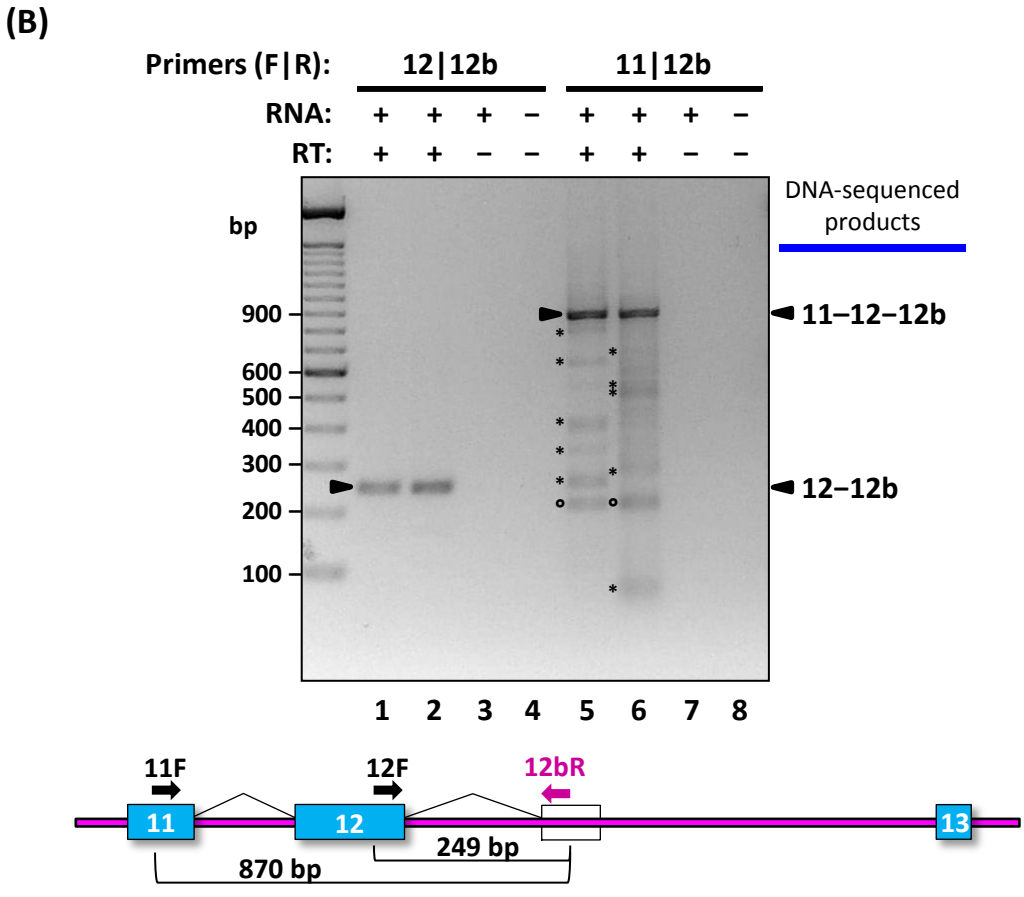
Figure 5.4 Exon 12b splices to upstream and downstream exons

(A) Exon 12b splices downstream to exon 13 within 12b–13, 12b–13–14 and 12b–13–14–15 transcripts: Total RNA prepared from quiesced 10T½ fibroblasts was digested with DNase and used to generate first-strand cDNA and RT-negative samples. PCR was performed with a forward primer to exon 12b and reverse primers within exon 13 (lanes 1–4), exon 14 (lanes 5–8) or exon 15 (lanes 9–12), using three different templates: (i) two cDNA replicates (reactions in lanes 1, 5 and 9 are repeated with independent cDNA in lanes 2, 6 and 10); (ii) RT-negative sample (lanes 3, 7 and 11); and (iii) water as a non-template control (lanes 4, 8 and 12). Arrowheads indicate products confirmed by DNA sequencing. The size and structure of confirmed RT-PCR products is indicated below the gel (not to scale).

(B) Exon 12b splices upstream to exon 12 within 12–12b and 11–12–12b transcripts: PCR was performed with a reverse primer to exon 12b and forward primers within exon 12 (lanes 1–4) or exon 11 (lanes 5–8), using templates prepared according to **(A)**: two cDNA replicates (lanes 1, 2, 5 and 6), RT-negative controls (lanes 3 and 7), and non-template controls (lanes 4 and 8). Arrowheads indicate products confirmed by DNA sequencing; asterisks indicate unreproducible PCR products; open circles indicate reproducible products of unknown identity. The size and structure of confirmed RT-PCR products is indicated below the gel (not to scale).

RT-PCR and electrophoresis in this experiment were performed by Jonathan Neve (Part II student) under my direct supervision.


(A)
Primers (F|R):
12b|13
12b|14
12b|15
RNA:
RT:
bp
600
500
400
300
200
100
DNA-sequenced products
12b–13–14–15
12b–13–14
12b–13
1 2 3 4 5 6 7 8 9 10 11 12
12bF
13R
14R
15R
12
13
14
15
159 bp
377 bp
529 bp


(B)
Primers (F|R):
12|12b
11|12b
RNA:
RT:
bp
900
600
500
400
300
200
100
DNA-sequenced products
11–12–12b
12–12b
1 2 3 4 5 6 7 8
11F
12F
12bR
11
12
13
249 bp
870 bp

that excluded exons 13 or/and 14, though the existence of such transcripts cannot be definitively excluded.

5.4.4. Identification of alternative *Setd2* transcripts that contain exon 12b spliced upstream to exon 12

RT-PCR was similarly employed to ascertain whether exon 12b splices upstream to exons 11 and 12 (**Figure 5.4B**). DNA sequencing confirmed that exon 12b splices upstream to exon 12 within two isolated cDNAs: (i) 12−12b (**Figure 5.4B**, lanes 1 and 2; **Appendix C-2.4**) and (ii) 11−12−12b (**Figure 5.4B**, lanes 5 and 6; **Appendix C-2.5**). Splicing of exon 12 to 12b occurs via canonical GT/AC sites for the major spliceosome (**Appendix C-1.1**).

The 11F and 12bR primers also amplified multiple smaller products less efficiently and reproducibly than the major 11−12−12b product (**Figure 5.4B**, lanes 5 and 6)**.** Their identity is unclear: they may represent additional alternatively spliced *Setd2* transcripts or they may result from non-specific amplification. The identity of a reproducible ≈200−250 bp product (**Figure 5.4B**, lanes 5 and 6, open circles) is unclear, but may represent 11−12b splicing with exon 12 exclusion (209 bp predicted).

5.4.5. Major transcripts across the 12−13 region exclude exon 12b

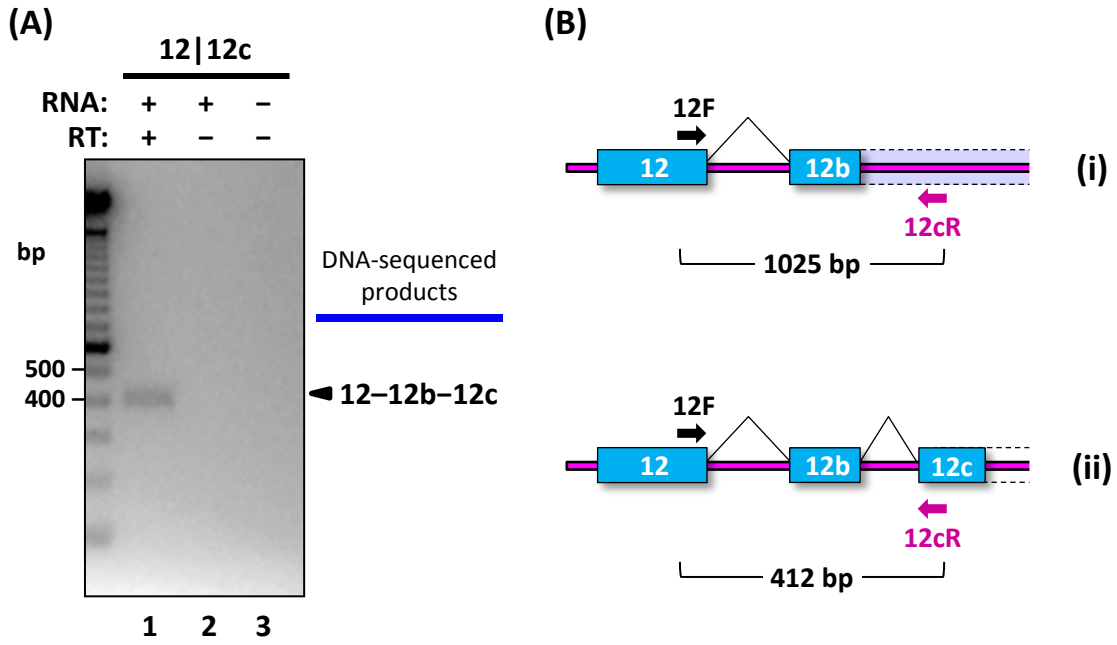
The identification of cDNAs that either harbour exon 12b spliced upstream to exon 12 or contain exon 12b spliced downstream to exon 13 raised the question of whether these partial cDNAs derive from a 12−12b−13 transcript or represent distinct transcripts. To address whether exon 12b is included within a 12−12b−13 splicing pattern, RT-PCR was performed using a forward primer in exon 12, a reverse primer in exon 13, and a PCR elongation time (120 s) that is theoretically sufficient to amplify the predicted size of a 12−12b−13 transcript (396 bp) (**Figure 5.5**).

Exon 12b is excluded from the most abundant products detected across the region spanning exons 11 to 14. Exon 12b is excluded from cDNAs: (i) 12−13 (**Figure 5.5**, lanes 1 and 2); (ii) 12−13−14 (**Figure 5.5**, lanes 5 and 6); and (iii) 11−12−13 (**Figure 5.5**, lanes 9 and 10). Additional less abundant products were observed for 12F|14R and 11F|13R

Figure 5.5 Major splice products across exons 12 and 14 exclude exon 12b

Total RNA prepared from quiesced 10T½ fibroblasts was digested with DNase and used to generate first-strand cDNA and RT-negative samples. PCR was performed with the indicated primers and three different templates: (i) two cDNA replicates (reactions in lanes 1, 5 and 9 are repeated with independent cDNA in lanes 2, 6 and 10); (ii) RT-negative sample (lanes 3, 7 and 11); and (iii) water as a non-template control (lanes 4, 8 and 12). Arrowheads indicate products verified by DNA sequencing; asterisks indicate additional unidentified products. The size and structure of RT-PCR products derived from transcripts with inclusion or exclusion of exon 12b is indicated below the gel (not to scale).

RT-PCR and electrophoresis in this experiment were performed by Jonathan Neve (Part II student) under my direct supervision.


(A)
12|12c
RNA: + + –
RT: + – –
bp
500
400
DNA-sequenced products
12–12b–12c
1 2 3
(B)
12F
12
12b
12cR
1025 bp
(i)
12F
12
12b
12c
12cR
412 bp
(ii)

reactions at higher molecular weight (**Figure 5.5**, lanes 5, 6, 9 and 10, asterisks), but were not obtained in sufficient amount to identify by direct sequencing. The lower abundance of these higher molecular weight products may also be attributed to the inherent bias of PCR to amplify smaller templates more efficiently. The identification of these additional bands will require further cloning and DNA sequencing in future.

Altogether, two mutually exclusive transcripts must minimally occur, defined by inclusion of the alternative exon 12b: (i) 12–13, and (ii) 12–12b and 12b–13, which may or may not derive from the same alternative transcript.

5.4.6. Discovery of exon 12c confirms existence of mutually-exclusive splicing events involving exon 12b

During the course of investigating the inclusion of exon 12b in a 12–12b-containing transcript by 3′ RACE, an additional novel exon, 12c, was identified by RT-PCR (**Figure 5.6A**, lane 1; **Appendix C-2.6**) using a reverse primer lying ≈600 bp downstream of the currently defined 3′ boundary of exon 12b, within the 12b–13 intron (**Figure 5.6B**). The splicing of exon 12b to exon 12c within a 12–12b–12c cDNA occurs via canonical GT/AC splices sites used by the major spliceosome (**Appendix C-1.3**).

No additional larger PCR products were observed, despite the use of an elongation time (120 s) theoretically sufficient to amplify a product of ≈1 kb, which would arise if the 12–12b–containing transcript continued contiguously downstream of exon 12b (**Figure 5.6B-i**).

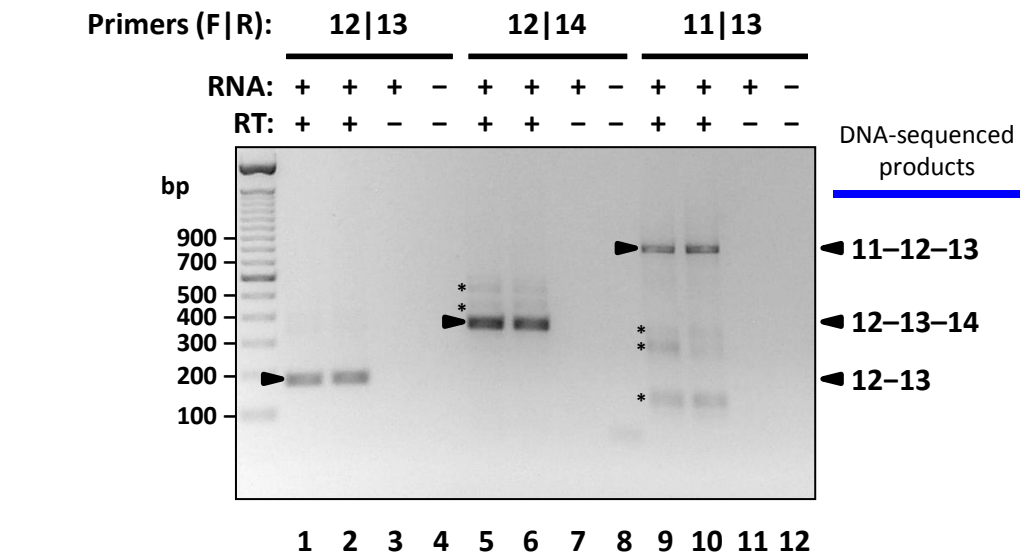
Nevertheless, since 12–12b–12c splicing is incompatible with 12–12b–13, amplification of 12–12b cannot simply derive exclusively from a transcript containing 12–12b–13. Consequently, splicing of 12b–12c demonstrates that the 12–12b and 12b–13 cDNAs cannot solely derive from a single transcript. Thus, alternative splicing within the 12–13 region defines at least three, mutually exclusive transcripts: (i) 12–13; (ii) 12–12b–12c; and (iii) 12b–13.

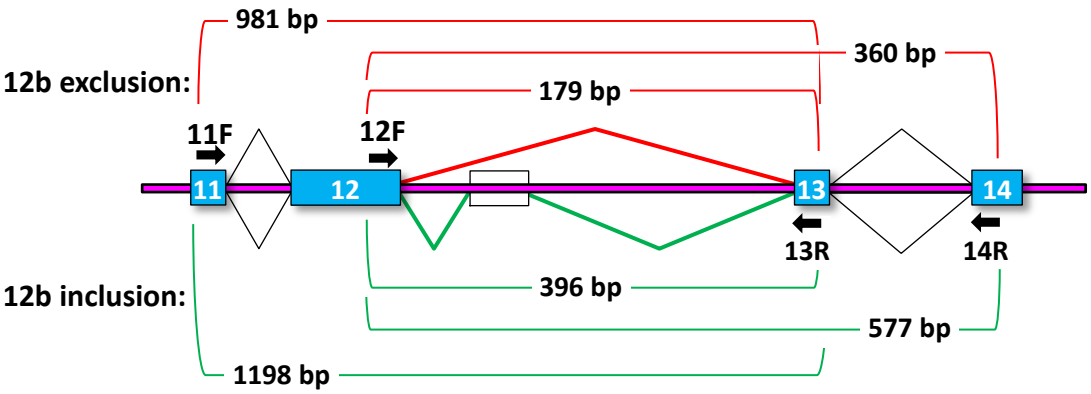
Figure 5.6 Exon 12b splices to a novel downstream exon (12c)

(A) RT-PCR analysis of alternative splicing downstream of the 12–12b transcript: Total RNA prepared from quiesced 10T½ fibroblasts was digested with DNase and used to generate first-strand cDNA and RT-negative samples. PCR was performed with a forward primer to exon 12 and reverse primer 12cR, which lies ≈600 bp downstream of the 12bR reverse primer used to define the 3′ region of exon 12b (highlighted in panel **B**). PCR reactions contained cDNA (lane 1), RT-negative sample (lane 2), or water as a non-template control (lane 3). The arrowhead indicates the 12–12b–12c product confirmed by DNA sequencing.

(B) Size and structure of predicted versus confirmed RT-PCR products: The product of the 12F|12cR PCR reaction was originally predicted to amplify contiguous sequence downstream of exon 12b **(i)**. However, DNA sequencing confirmed exclusion of ≈600 bp sequence between exons 12b and 12c **(ii)**.

RT-PCR and electrophoresis in this experiment were performed by Jonathan Neve (Part II student) under my direct supervision.


Primers (F|R):
12|13
12|14
11|13
RNA: + + + − + + + − + + + −
RT: + + − − + + − − + + − −
DNA-sequenced products
bp
900
700
500
400
300
200
100
11–12–13
12–13–14
12–13
1 2 3 4 5 6 7 8 9 10 11 12


12b exclusion:
981 bp
360 bp
179 bp
11F
12F
11
12
13
14
13R
14R
12b inclusion:
396 bp
577 bp
1198 bp

5.4.7. Alternative 3' splice sites define two 5' boundaries of exon 14

During the analysis of the DNA sequence of whole PCR products containing exon 14, mixed sequence reads with (n + 3)-displacement were obtained following the 13–14 boundary (**Figure 5.7A**). Such mixed sequence reads cannot be explained by slippage of primers since they occur internal to the cDNA. Further analysis showed that the population likely derives from alternative 3' splice sites within a NAGNAG tandem that define the 5' boundary of exon 14 (**Figure 5.7A** and **B**). Since both mixed sequence reads of (n + 3)-displacement were observed for both 12b–13–14 and 12–13–14 cDNAs, four varieties of alternatively spliced transcript likely exist in mouse cells: (i) 12–13–14; (ii) 12–13–*14; (iii) 12b–13–14; and (iv) 12b–13–*14 (**Figure 5.7C**). Direct confirmation of the existence of this variation, however, will require cloning of individual cDNAs.

Although the NAGNAG tandem splice site in Rhesus monkey and the second 3' splice site in opossum appear to be conserved, there is a lack of conservation of such a tandem sequence in the remaining 21 vertebrates analysed (**Figure 5.7B**). Although the utilisation of different 3' splice sites may be functionally significant in an organism-specific manner, it may simply reflect splicing noise.

5.4.8. Isolation of 'retained-intron' transcripts

In addition to the alternative splicing at exon 12b, several mRNAs and ESTs have been previously described with retained introns following exons 9, 10 and 18 in diverse mouse tissues (**Figure 5.1**). In order to minimise amplification of splicing intermediates or residual genomic DNA, RT-PCR was performed using forward primers located in upstream exons. RT-PCR and DNA sequencing confirmed the existence of three retained-intron transcripts: (i) 8–9.5 (**Figure 5.8A**, lane 1; **Appendix C-2.8**); (ii) 17–18.5 (**Figure 5.8A,** lane 4; **Appendix C-2.9**); and (iii) 6–7–8–9–10.5 (**Figure 5.8A**, lane 7; **Appendix C-2.7**). Although these cDNAs contain exons that are spliced upstream of the retained intron, they could still equally represent either intermediates of pre-mRNA splicing or *bona fide* alternative transcripts with either downstream termination or additional splicing events.

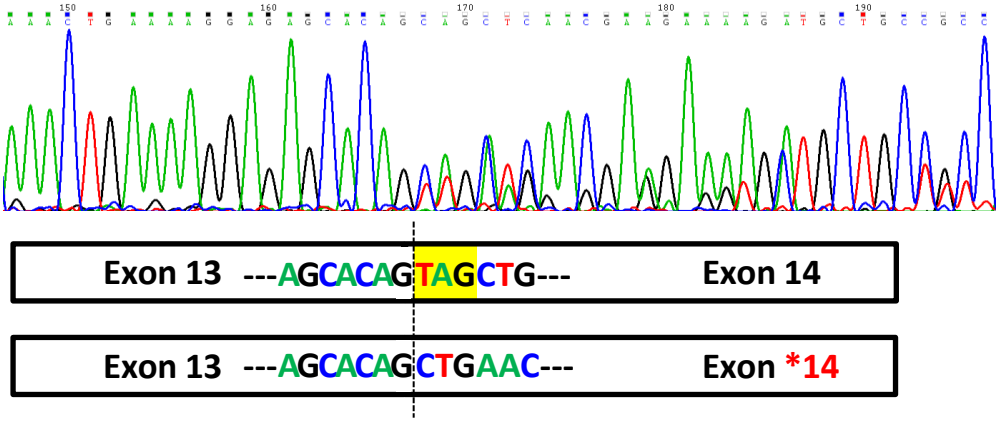
Figure 5.7 Alternative 3' splice sites define two 5’ boundaries of exon 14

(A) DNA chromatogram of sequenced 12b–13–14–15 product: mixed template reads after exon 13 correlate with the utilisation of alternative 3’ splice site in the 13–14 intron. Similar mixed reads were observed for 12b–13–14 product (data not shown).

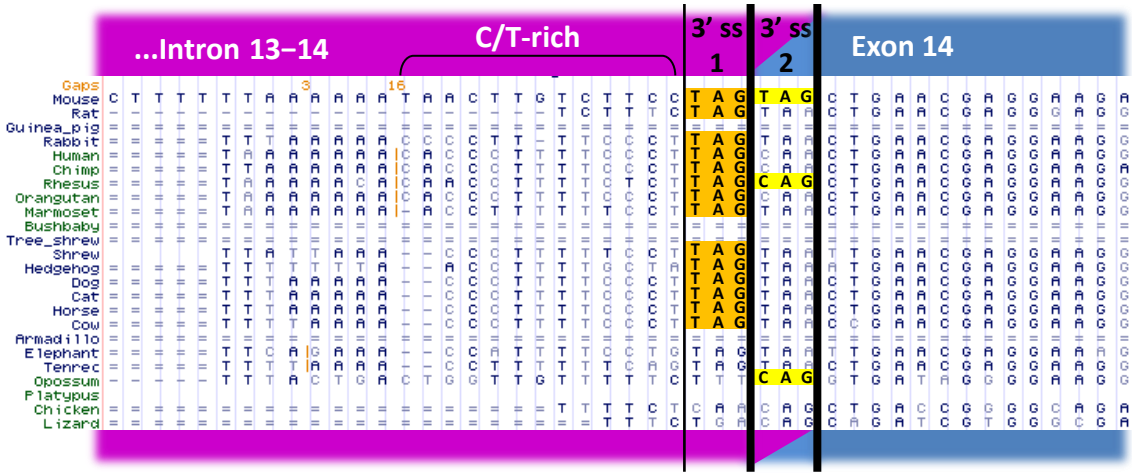
(B) Conservation of alternative 3’ splice sites upstream of exon 14: MultiZ alignment of *SETD2* genomic sequences from 24 vertebrates that are homologous to the region surrounding the 3’ splice sites upstream of exon 14 in murine *Setd2*. The two different 3’ splice site sequences (1 and 2) are indicated in orange and yellow, respectively.

(C) Structure of alternative splicing patterns across the region spanning exons 12 to 14

(A)


1
Exon 13 ---AGCACAGTAGCTG--- Exon 14
2
Exon 13 ---AGCACAGCTGAAC--- Exon *14

(B)


...Intron 13–14
C/T-rich
3' ss 1
3' ss 2
Exon 14
Gaps
Mouse
Rat
Guinea_pig
Rabbit
Human
Chimp
Rhesus
Orangutan
Marmoset
Bushbaby
Tree_shrew
Shrew
Hedgehog
Dog
Cat
Horse
Cow
Armadillo
Elephant
Tenrec
Opossum
Platypus
Chicken
Lizard

(C)

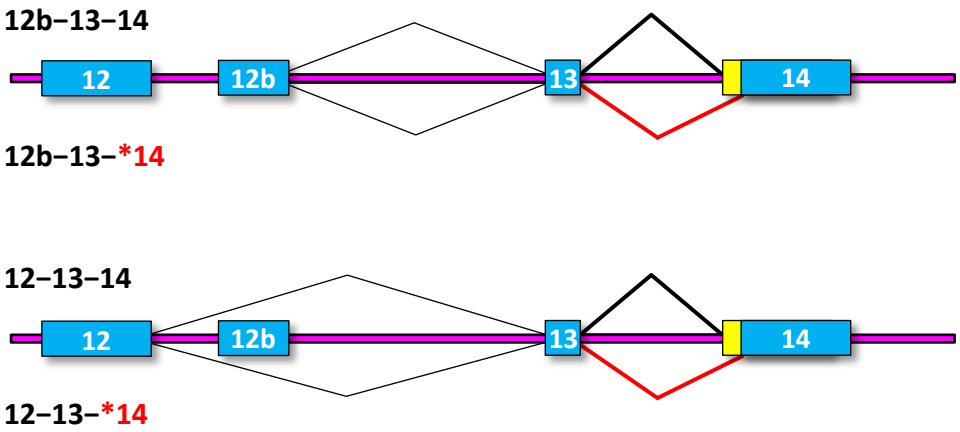

12b–13–14
12
12b
13
14
12b–13–*14
12–13–14
12
12b
13
14
12–13–*14

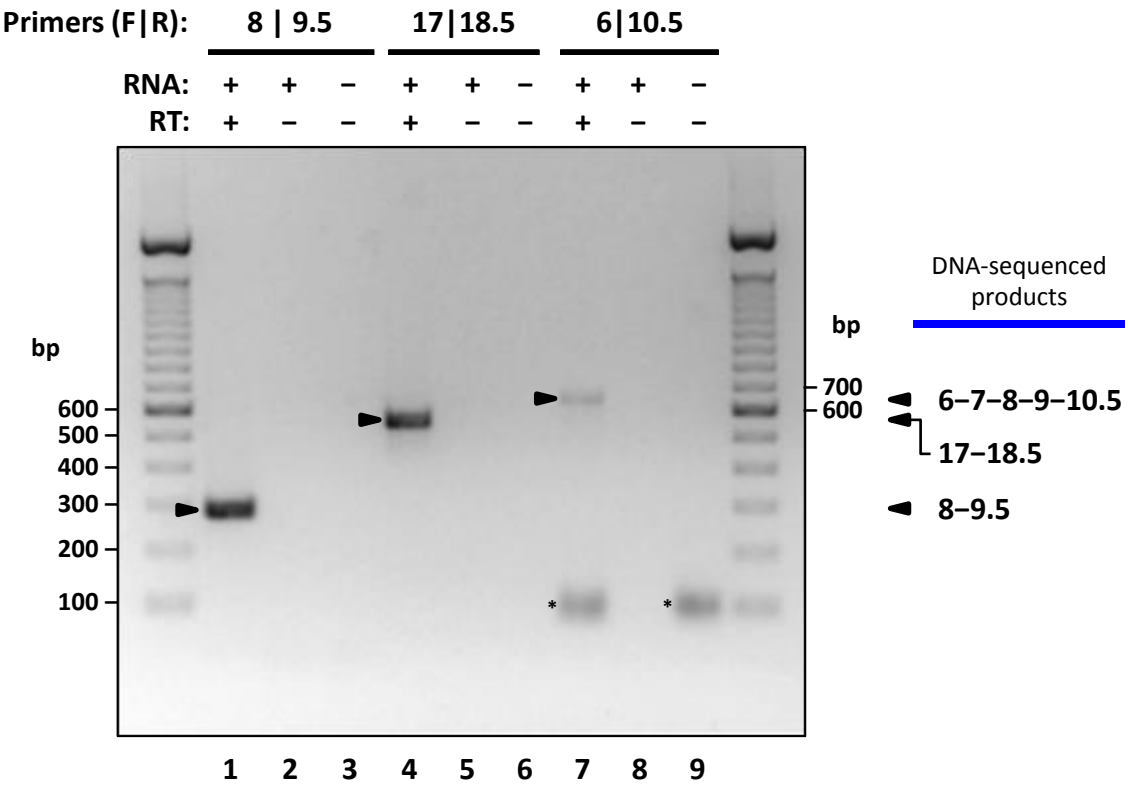
Figure 5.8 Identification of *Setd2* transcripts with retained introns

(A) Total RNA prepared from quiesced 10T½ fibroblasts was digested with DNase and used to generate first-strand cDNA and RT-negative samples. PCR was performed with the indicated primers and three different templates: (i) cDNA sample (lanes 1, 4 and 7); (ii) RT-negative sample (lanes 2, 5 and 8); and (iii) water as a non-template control (lanes 3, 6 and 9). Arrowheads indicate products confirmed by DNA sequencing; the asterisk indicates a probable primer-dimer product.

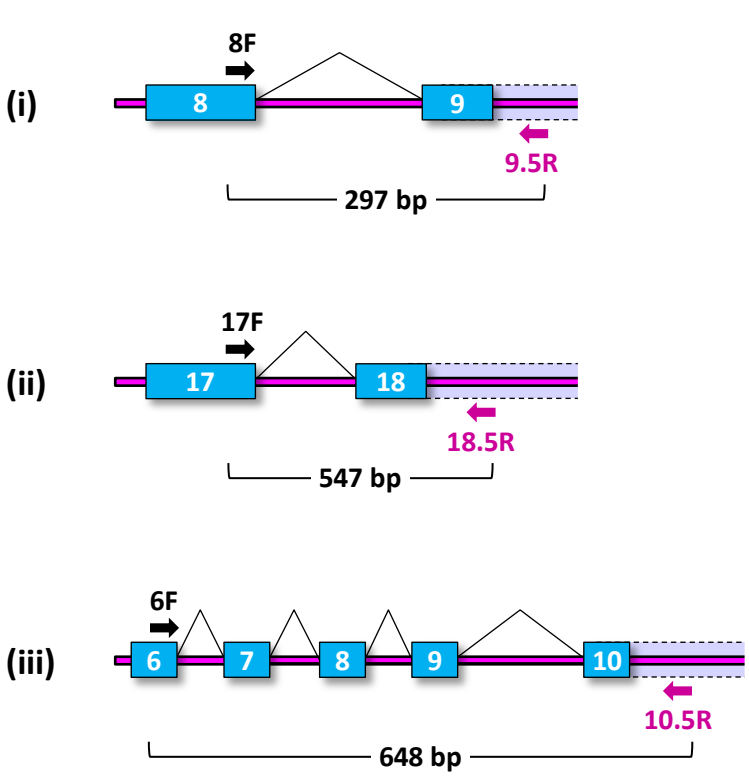
(B) Size and structure of retained intron transcripts confirmed by DNA sequencing (not to scale).

RT-PCR and electrophoresis in this experiment were performed by Jonathan Neve (Part II student) under my direct supervision.

(A)


Primers (F|R):
8 | 9.5
17|18.5
6|10.5
RNA: + + − + + − + + −
RT: + − − + − − + − −
bp
600
500
400
300
200
100
1 2 3 4 5 6 7 8 9
bp
700
600
DNA-sequenced products
6–7–8–9–10.5
17–18.5
8–9.5

(B)


(i)
8F
8
9
9.5R
297 bp
(ii)
17F
17
18
18.5R
547 bp
(iii)
6F
6
7
8
9
10
10.5R
648 bp

5.4.9. Identification of a mouse brain cDNA containing a complete novel ORF

Multiple attempts to amplify cDNA corresponding to 12b»21 and 12»21 from both total and poly(A)+ RNA by RT-PCR were unsuccessful, likely due to technical reasons. During the course of cloning experiments (presented in Chapter 6), a partially sequenced, commercially available cDNA clone (CA319728) from mouse brain, which corresponded to 12−12b−13−14, was fully sequenced and revealed to be: 12-12b-13-14-15-16-17-18-19-20-21 (**Figure 5.9A**).

This demonstrates that, at least in brain, murine cells indeed produce an alternatively spliced transcript containing 12b»21. It is unclear, however, how this relates to the complete transcript population in 10T½ cells—especially since the major species of exon 12b–containing transcripts do not appear to contain the 12−12b−13 pattern (**Figure 5.5**). The 12b»15 cDNA isolated (**Figure 5.4A**) may correspond to a fragment of the 12b»21 transcript.

Nevertheless, this cDNA clone represents the first identification of a novel, complete ORF (*i.e.* termination codons present upstream of predicted translation initiation sites) in mice, which is distinct from the ORF that encodes $p290^{SETD2}$. The predicted ORFs may produce a protein of ≈61 kDa, containing both C-terminal and CT21 epitopes, similar to the characteristics of $p66^{C/CT21}$ (**Figure 5.9B**).

Figure 5.9 The 12−12b−13»21 cDNA contains a novel ORF encoding a ≈61 kDa protein containing C-terminal and CT21 epitopes

(A) **ORF prediction within 12−12b−13»21 cDNA:** ORFs in-frame with p290^{SETD2} protein sequence were predicted within the 12−12b−13»21 cDNA by the Virtual Ribosome program (Wernersson, 2006). The sequence presented represents nucleotides 180–2520 of CA319728 clone (the omitted first 180 nt correspond to upstream exon 12 sequence) and the deletion of adenine-1412 is highlighted against a black background. Exon 12b is highlighted in blue text. Two different reading frames are presented for nucleotides 180–540: the first row represents protein sequence in frame with p290^{SETD2} whilst the second row represents sequence in frame with the ≈61 kDa ORF downstream. Translation initiation sequences are highlighted: [›››] represents ATG codons within Kozak consensus sequences; [)))] represents alternative Kozak-compatible initiation sequences. The first ATG initiation codon within the predicted p66 ORF is highlighted in cyan, whilst alternative upstream initiation sites are coloured green. Termination codons for the upstream ORF (in-frame with p290^{SETD2}) and the ≈61 kDa ORF are highlighted in red. The predicted ≈61 kDa ORF (highlighted in yellow) initiates from a Kozak-compatible initiation sequence (TTG at 596–599 bp). The C-terminal and CT21 epitopes are highlighted in blue and magenta, respectively.

(B) *Continued on page 205.*

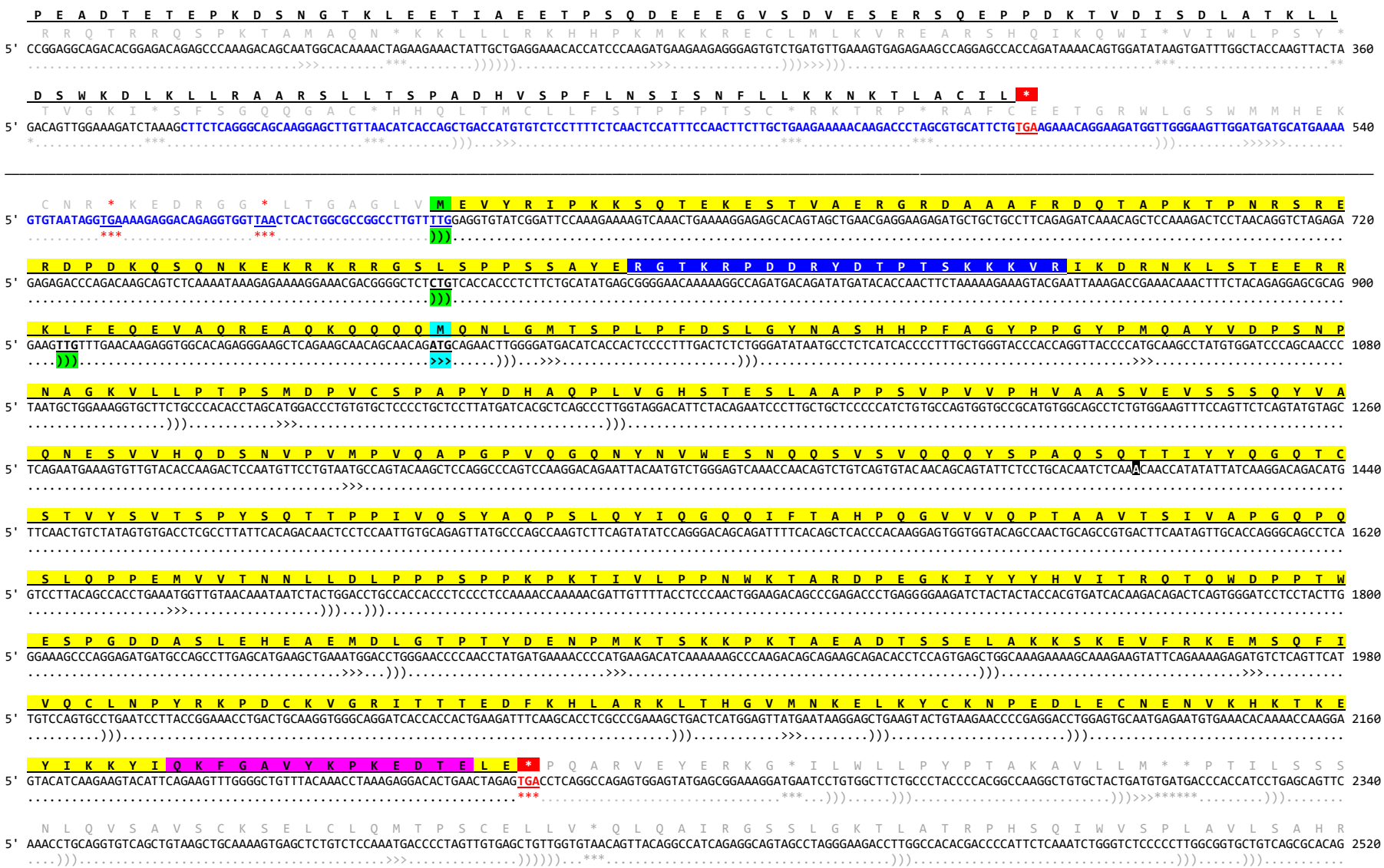
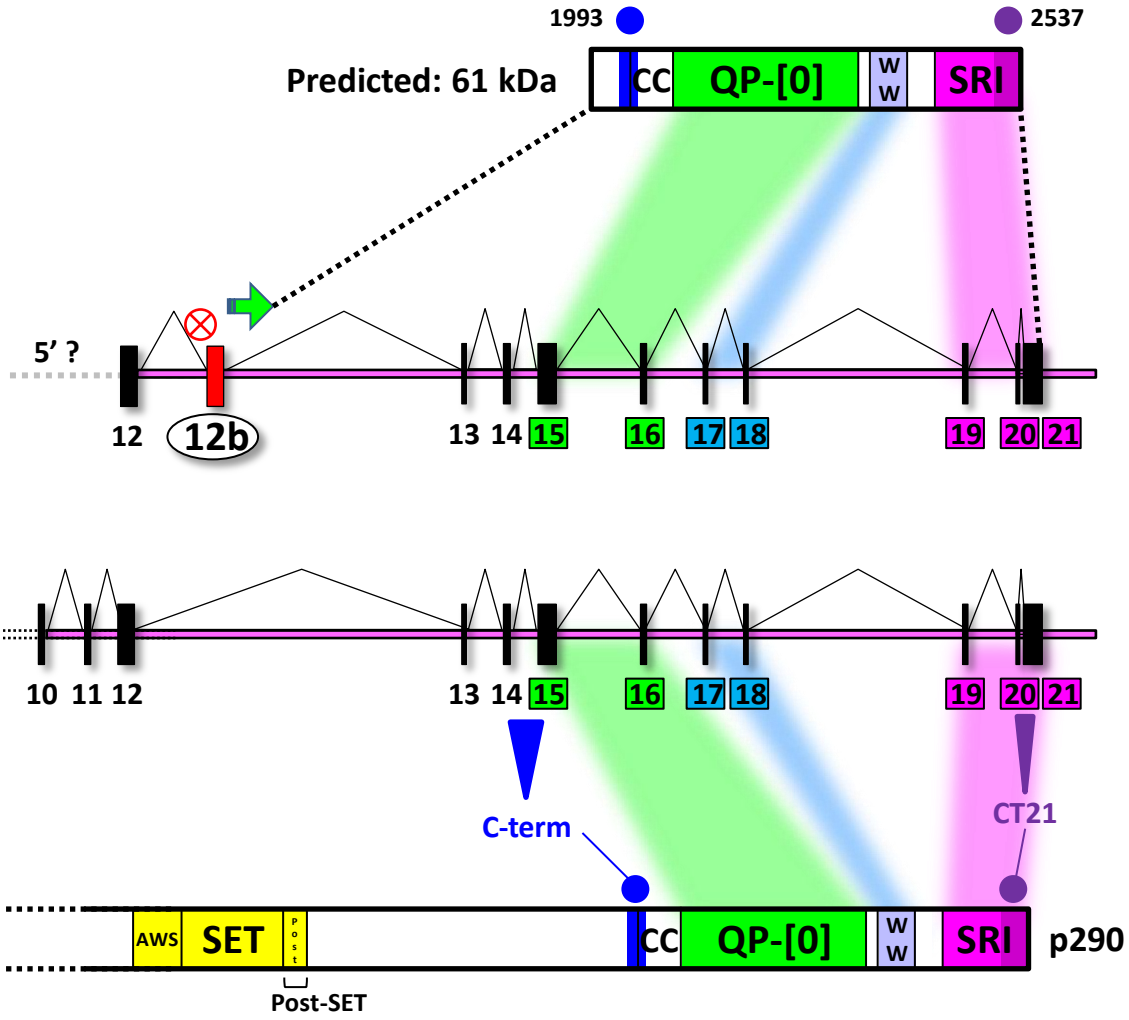

P E A D T E T E P K D S N G T K L E E T I A E E T P S Q D E E E G V S D V E S E R S Q E P P D K T V D I S D L A T K L L
R R Q T R R Q S P K T A M A Q N * K K L L L R K H H P K M K K R E C L M L K V R E A R S H Q I K Q W I * V I W L P S Y *
5' CCGGAGGCAGACACGGAGACAGAGCCCAAAGACAGCAATGGCACAAAACTAGAAGAAACTATTGCTGAGGAAACACCATCCCAAGATGAAGAAGAGGGAGTGTCTGATGTTGAAAGTGAGAGAAGCCAGGAGCCACCAGATAAAACAGTGGATATAAGTGATTTGGCTACCAAGTTACTA 360
D S W K D L K L L R A A R S L L T S P A D H V S P F L N S I S N F L L K K N K T L A C I L *
T V G K I * S F S G Q Q G A C * H H Q L T M C L L F S T P F P T S C * R K T R P * R A F C E E T G R W L G S W M M H E K
5' GACAGTTGGAAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTTGTTAACATCACCAGCTGACCATGTGTCTCCTTTTCTCAACTCCATTTCCAACTTCTTGCTGAAGAAAAACAAGACCCTAGCGTGCATTCTGTGAAGAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAA 540
C N R * K E D R G G * L T G A G L V M E V Y R I P K K S Q T E K E S T V A E R G R D A A A F R D Q T A P K T P N R S R E
5' GTGTAATAGGTGAAAAGAGGACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTGGAGGTGTATCGGATTCCAAAGAAAAGTCAAACTGAAAAGGAGAGCACAGTAGCTGAACGAGGAAGAGATGCTGCTGCCTTCAGAGATCAAACAGCTCCAAAGACTCCTAACAGGTCTAGAGA 720
R D P D K Q S Q N K E K R K R R G S L S P P S S A Y E R G T K R P D D R Y D T P T S K K K V R I K D R N K L S T E E R R
5' GAGAGACCCAGACAAGCAGTCTCAAAATAAAGAGAAAAGGAAACGACGGGGCTCTCTGTCACCACCCTCTTCTGCATATGAGCGGGGAACAAAAAGGCCAGATGACAGATATGATACACCAACTTCTAAAAAGAAAGTACGAATTAAAGACCGAAACAAACTTTCTACAGAGGAGCGCAG 900
K L F E Q E V A Q R E A Q K Q Q Q Q M Q N L G M T S P L P F D S L G Y N A S H H P F A G Y P P G Y P M Q A Y V D P S N P
5' GAAGTTGTTTGAACAAGAGGTGGCACAGAGGGAAGCTCAGAAGCAACAGCAACAGATGCAGAACTTGGGGATGACATCACCACTCCCCTTTGACTCTCTGGGATATAATGCCTCTCATCACCCCTTTGCTGGGTACCCACCAGGTTACCCCATGCAAGCCTATGTGGATCCCAGCAACCC 1080
N A G K V L L P T P S M D P V C S P A P Y D H A Q P L V G H S T E S L A A P P S V P V V P H V A A S V E V S S S Q Y V A
5' TAATGCTGGAAAGGTGCTTCTGCCCACACCTAGCATGGACCCTGTGTGCTCCCCTGCTCCTTATGATCACGCTCAGCCCTTGGTAGGACATTCTACAGAATCCCTTGCTGCTCCCCCATCTGTGCCAGTGGTGCCGCATGTGGCAGCCTCTGTGGAAGTTTCCAGTTCTCAGTATGTAGC 1260
Q N E S V V H Q D S N V P V M P V Q A P G P V Q G Q N Y N V W E S N Q Q S V S V Q Q Q Y S P A Q S Q T T I Y Y Q G Q T C
5' TCAGAATGAAAGTGTTGTACACCAAGACTCCAATGTTCCTGTAATGCCAGTACAAGCTCCAGGCCCAGTCCAAGGACAGAATTACAATGTCTGGGAGTCAAACCAACAGTCTGTCAGTGTACAACAGCAGTATTCTCCTGCACAATCTCAAACAACCATATATTATCAAGGACAGACATG 1440
S T V Y S V T S P Y S Q T T P P I V Q S Y A Q P S L Q Y I Q G Q Q I F T A H P Q G V V V Q P T A A V T S I V A P G Q P Q
5' TTCAACTGTCTATAGTGTGACCTCGCCTTATTCACAGACAACTCCTCCAATTGTGCAGAGTTATGCCCAGCCAAGTCTTCAGTATATCCAGGGACAGCAGATTTTCACAGCTCACCCACAAGGAGTGGTGGTACAGCCAACTGCAGCCGTGACTTCAATAGTTGCACCAGGGCAGCCTCA 1620
S L Q P P E M V V T N N L L D L P P P S P P K P K T I V L P P N W K T A R D P E G K I Y Y Y H V I T R Q T Q W D P P T W
5' GTCCTTACAGCCACCTGAAATGGTTGTAACAAATAATCTACTGGACCTGCCACCACCCTCCCCTCCAAAACCAAAAACGATTGTTTTACCTCCCAACTGGAAGACAGCCCGAGACCCTGAGGGGAAGATCTACTACTACCACGTGATCACAAGACAGACTCAGTGGGATCCTCCTACTTG 1800
E S P G D D A S L E H E A E M D L G T P T Y D E N P M K T S K K P K T A E A D T S S E L A K K S K E V F R K E M S Q F I
5' GGAAAGCCCAGGAGATGATGCCAGCCTTGAGCATGAAGCTGAAATGGACCTGGGAACCCCAACCTATGATGAAAACCCCATGAAGACATCAAAAAAGCCCAAGACAGCAGAAGCAGACACCTCCAGTGAGCTGGCAAAGAAAAGCAAAGAAGTATTCAGAAAAGAGATGTCTCAGTTCAT 1980
V Q C L N P Y R K P D C K V G R I T T T E D F K H L A R K L T H G V M N K E L K Y C K N P E D L E C N E N V K H K T K E
5' TGTCCAGTGCCTGAATCCTTACCGGAAACCTGACTGCAAGGTGGGCAGGATCACCACCACTGAAGATTTCAAGCACCTCGCCCGAAAGCTGACTCATGGAGTTATGAATAAGGAGCTGAAGTACTGTAAGAACCCCGAGGACCTGGAGTGCAATGAGAATGTGAAACACAAAACCAAGGA 2160
Y I K K Y I Q K F G A V Y K P K E D T E L E * P Q A R V E Y E R K G * I L W L L P Y P T A K A V L L M * * P T I L S S S
5' GTACATCAAGAAGTACATTCAGAAGTTTGGGGCTGTTTACAAACCTAAAGAGGACACTGAACTAGAGTGACCTCAGGCCAGAGTGGAGTATGAGCGGAAAGGATGAATCCTGTGGCTTCTGCCCTACCCCACGGCCAAGGCTGTGCTACTGATGTGATGACCCACCATCCTGAGCAGTTC 2340
N L Q V S A V S C K S E L C L Q M T P S C E L L V * Q L Q A I R G S S L G K T L A T R P H S Q I W V S P L A V L S A H R
5' AAACCTGCAGGTGTCAGCTGTAAGCTGCAAAAGTGAGCTCTGTCTCCAAATGACCCCTAGTTGTGAGCTGTTGGTGTAACAGTTACAGGCCATCAGAGGCAGTAGCCTAGGGAAGACCTTGGCCACACGACCCCATTCTCAAATCTGGGTCTCCCCCTTGGCGGTGCTGTCAGCGCACAG 2520

Figure 5.9 The 12–12b–13»21 cDNA contains a novel ORF encoding a ≈61 kDa protein containing C-terminal and CT21 epitopes *(continued)*

(B) Predicted ≈61 kDa protein contains C-terminal SETD2 domains: inclusion of exon 12b (red) terminates upstream translation and, with splicing to exon 13, permits predicted translation initiation (green arrow) of a ≈61 kDa protein containing**:** coiled-coil (CC), glutamine/proline-rich domain (QP), WW domain, and Set2–Rpb1 interaction (SRI) domain. Location of C-terminal and CT21 epitopes are indicated as blue and purple circles, respectively.


1993
2537
Predicted: 61 kDa
CC
QP-[0]
WW
SRI
5' ?
12
12b
13 14 15
16
17 18
19 20 21
10 11 12
C-term
CT21
AWS
SET
Post
Post-SET
p290

5.5. Discussion

5.5.1. Summary of results

- Multiple alternative *Setd2* transcripts were identified in C3H 10T½ cells by RT-PCR and DNA sequencing. Northern blotting with 3′ region probes detected a >6.6 kb transcript in addition to multiple smaller transcripts (**Figure 5.2**).

- Five mutually exclusive transcript variants were defined by the combination of: (i) alternative inclusion of exons 12b and 12c across the region spanning exons 12–13 (**Figures 5.4–5.6**); and (ii) splicing of exon 13 to alternative 3′ splice sites upstream of exon 14 (**Figure 5.7**):
    - (i) 12–13–14
    - (ii) 12–13–*14
    - (iii) 12–12b–12c
    - (iv) 12b–13–14
    - (v) 12b–13–*14

- Exon 12b was also found in 11–12–12b and 12b–13–(*)14–15 cDNAs (**Figure 5.4**).

- Partial intron-retention was observed after exons 8, 9 and 18 (**Figure 5.8**):
    - (i) 8–9.5
    - (ii) 6–7–8–9–10.5
    - (iii) 17–18.5

- Identification of cDNA spanning exons 12–12b–13»21 (**Figure 5.9**). This transcript contains a novel, full-length ORF in-frame with SETD2 protein sequence. The predicted ≈61 kDa protein contains C-terminal and CT21 epitopes and may be related to $p66^{C/CT21}$.

5.5.2. How many alternative *Setd2* transcripts exist?

5.5.2.1. *Existence of the canonical 1»21 transcript*

Northern blotting detected a single transcript greater than 6.6 kb. The size is similar to previous reports and is consistent with the canonical gene model for production of the p290^{SETD2} protein in 10T½ (validated in Chapter 3). Two previous studies have identified human *SETD2* transcripts from a variety of sources. Although both of these studies report a consistent size of transcript across multiple cell types, the reports conflict with respect to the reported size: ≈7.5 versus ≈9 kb (Faber *et al.*, 1998; Rega *et al.*, 2001). The >6.6 kb transcript is assumed to represent the 1»21 prediction, which would encode p290^{SETD2} in 10T½ cells. The discussion of RT-PCR products below assumes this, but the validity of this interpretation is called into question in the next chapter.

5.5.2.2. *Evidence for additional transcripts in previously published data*

The previous reports use poly(A)+ RNA from whole tissues, which is more susceptible to ribonuclease-mediated degradation, than RNA prepared from cultured cells (Faber *et al.*, 1998; Rega *et al.*, 2001). Furthermore, whole-tissue RNA reflects a population average over the transcriptome of different cell types; thus rare transcripts expressed at low levels or in specific sub-populations may escape detection. The use of a homogeneous population of quiesced 10T½ cells has an advantage in this respect—and the transcript profile is directly comparable to the protein analysis under identical culture conditions.

In certain tissues, smaller transcripts are faintly visible (Faber *et al.*, 1998; Rega *et al.*, 2001), and approximate the size of the more prominent smaller transcripts observed in 10T½ cells (**Figure 5.2B**).

The identity of the transcripts below 6.6 kb is unclear and the signals themselves are poorly discernible at present. They are unlikely to derive from RNA degradation during sample preparation, since the *Gapdh* probe detects a single, discrete transcript with minimal smearing, and total poly(A)+ RNA appears intact by ethidium bromide–staining (**Figure 5.2Ai** and **Bi**). Non-specific cross-reactivity with poly(A)+ RNA is also unlikely, because

signal intensities do not correlate with the extent of total RNA migration, and would be expected to be comparable between 10T½ and 293T cells.

If these smaller transcripts truly represent alternative *Setd2* variants, they must contain sequences within the 3′ region (15 to 19 or 21) and differ with respect to splicing, initiation or termination, compared to the major >6.6 kb transcript. In this respect, RT-PCR demonstrated that the 17−18.5 and 12−12b−13»21 cDNAs derive from alternative transcripts, which should be detected in the northern blots above (**Figure 5.2-Aii, Bii** and **iii**). Since these transcripts are alternatively spliced, it is likely (though not conclusive) that their size is different to that of the >6.6 kb transcript. Furthermore, preliminary analysis by semi-quantitative and real-time RT-PCR have suggested 12b−13−14−15 cDNAs are stabilised by emetine (data not shown), similar to the effect observed by northern blotting. However, without full-length cDNA information, it is impossible to assign these partial cDNAs to specific transcripts.

There is little conclusive evidence to support the existence of alternative promoters: the major peaks of promoter-associated H3K4me3 overlap the CpG island at exon 1 in all human and mouse cells analysed by the ENCODE project (Ernst *et al.*, 2011). However, minor peaks of H3K4me3 are observed over exon 12 in a subset of specific CD^+ cell types. Furthermore, in human *SETD2*, several different regions have been identified with capped 5′ termini, suggesting the possible existence of alternative promoters.

5.5.2.3. *The extent of transcript variation from RT-PCR data*

Several independent regions of mutually exclusive splicing patterns can be inferred from the RT-PCR data. Firstly, analysis of the region spanning exons 12 to 14 identified five independent splicing combinations—these definitively exist as mutually exclusive patterns (*i.e.* full-length transcripts containing 12−13−14 could not also physically contain the pattern 12b−13−*14). Secondly, the region spanning exons 8 to 11 defines an independent region of mutually exclusive splicing, assuming that exon 10 splices downstream to exon 11 (as predicted by the 1»21 transcript). Thirdly, intron retention after exon 18 defines

an additional, independent splicing region (assuming that exon 18 splices downstream to exon 19 in the 1»21 transcript).

Although different splicing patterns *within* these regions cannot physically occur in a single transcript, different permutations *between* the regions are feasible. For example, the existence of a transcript containing 12–13–14–15–16–17–18.5 is as equally consistent with the available data as 12b–13–*14–15–16–17–18–19. Since the 3′ boundaries of the retained introns are unknown, even these transcripts may contain further downstream splicing events (*e.g.* 6–7–8–9–10–intron–11–12...)

A crucial limitation in RT-PCR analysis of splicing is the inherent uncertainty of the 5′ and 3′ regions that flank the amplicon under analysis. Although the above permutations between mutually exclusive spliced regions are physically possible, cautious interpretation is necessary in the absence of direct evidence for a transcript's existence. For example, whilst RT-PCR demonstrated the existence of cDNAs corresponding to 11–12–12b and 12–12b–12c, it cannot be inferred that they derive solely from amplification from within a single 11–12–12b–12c transcript.

However, even conservative interpretation of the RT-PCR data supports five distinct transcript variants that definitively exist at a minimum (defined by the mutually exclusive splicing region between exons 12 and 14).

In conclusion, this is the first demonstration, as far as we are aware, of concurrent alternative splicing of the murine *SETD2* gene within a homogenous cell system.

The experiments here likely characterise, however, only a small proportion of the total alternative transcripts that probably occur *in vivo*. Several experiments reveal weakly amplified products that were of too low abundance to directly sequence and will require cloning for further analysis (**Figures 5.4B** and **5.5**). Recently released RNA-sequencing data confirm low-level splicing of exon 12b both upstream and downstream in human cell lines (Ernst *et al.*, 2011). In fact, RNA-sequencing of human *SETD2* reveals even more complexity of alternative splicing than has been determined from EST/cDNA evidence alone. For

example, exon 1 can splice downstream to exon 4 or 9; these events could potentially produce smaller proteins lacking the N-terminal region (which contributes ≈150 kDa to $p290^{SETD2}$) but containing C-terminal or CT21 epitopes, putatively similar to $p140^{C}$ or $p130^{CT21}$ proteins. Given the degree of conservation, it would not be surprising for the murine homolog to exhibit even more complex alternative splicing than has currently been determined.

5.5.3. The biological significance of alternative *Setd2* transcripts

The biological significance of the majority of alternative splicing in mammalian transcriptomes is poorly understood. All biological processes are afforded a certain degree of noise, and a proportion of alternative splicing events may arguably reflect inherent stochasticity of spliceosome action (Melamud & Moult, 2009a).

Considering non-random biological significance, the alternative transcripts may function at one or more levels. Firstly, the transcript itself may not be important, but rather the process of transcription (Hainer *et al.*, 2010) or targeting of co-transcriptional chromatin modifications, which may be influenced by splicing (de Almeida *et al.*, 2011; Kim *et al.*, 2011), may differentially affect chromatin regions. Secondly, alternative transcripts may not encode protein, but still perform regulatory roles; they may interact with other proteins or RNAs (such as miRNAs) influencing their concentration, activity and subcellular localisation. Thirdly, transcript variants may encode identical proteins, but variation in untranslated regions may confer different mechanisms of translation regulation (*e.g.* upstream ORFs), or transcript stabilities through AU-rich elements, miRNA binding sites or NMD induction (Moore & Proudfoot, 2009). Finally, alternative transcripts may produce distinct protein isoforms, differing subtly or substantially in structure and function.

5.5.3.1. *The fate of transcripts: translation and turnover*

Transcript structure critically influences transcript fate in the cytoplasm: the occurrence of splice junctions more than 50 nt downstream of a termination codon generally induces nonsense-mediated decay (NMD) (Moore & Proudfoot, 2009). The protein coding potential

of alternative transcripts identified here is therefore influenced by both upstream translation and downstream splicing events, both of which are currently unclear. Without full-length cDNA, it is impossible to predict exact protein-coding potential. However, extrapolation of flanking RNA sequences allows prediction of potentially encoded proteins.

Although alternative transcripts have the potential to produce proteins unrelated to $p290^{SETD2}$ sequence by translation in different reading frames, the following analysis focuses exclusively on ORFs in-frame with $p290^{SETD2}$, which would produce alternative SETD2 isoforms.

5.5.3.2. *Retained introns predict premature termination*

Retained intron transcripts should be viewed with caution, since they could represent splicing intermediates; however, intron retention is a genome-wide phenomenon, and has reported functional consequences for specific genes—e.g. *fosB* versus *ΔfosB* (Mumberg *et al.*, 1991); *NSD1* transcripts (Lucio-Eterovic *et al.*, 2010) and *GCN5* (Xu et al., 1998). Without 3′ information, it is unclear whether these *Setd2* transcripts retain partial or complete introns, which undergo further splicing downstream, or whether the retained intron represents an alternative terminal exon. Nevertheless, inclusion of intronic sequence after exons 9 and 10 incorporates termination codons into the $p290^{SETD2}$-reading frame. Premature termination would predict isoforms containing the KMT-related domains, uncoupled from the C-terminal regions, containing CC, QP-rich, WW and SRI domains (**Figure 5.10-A/B**), ascribed important roles in protein–protein and protein–DNA interactions (as discussed in the introduction). Such isoforms would lack C-terminal and CT21 epitopes but could potentially contain the N-terminal epitope, if the upstream region includes exon 3 in frame. Such proteins may be related to the N-terminal-specific $p120^{N}$ immunoform described in Chapter 3.

Likewise, intron retention after exon 18 predicts premature termination of translation, which would produce a protein with an intact WW domain but lack the C-terminal SRI

domain and CT21 epitope (**Figure 5.10-C**). The relevance of this observation will be revisited in Chapter 6 in relation to the identity of $p290^{SETD2}$ and the major *Setd2* transcript.

5.5.3.3. *Exon 12b and 12c*

The inclusion of exon 12b and 12c demonstrates that exon 12b is not a simple cassette that is differentially spliced between exons 12 and 13.

Both exon 12b and 12c insert termination codons into all six reading frames. However, distinct predictions can be made for the 11–12–12b and 12–12b–12c transcripts, if they are translated upstream. If the 11–12–12b–containing transcript terminates downstream without additional splicing events, the variant would avoid NMD and produce a truncated, N-terminal-specific protein (potentially consistent with $p120^{N}$). The 12–12b–12c–containing transcript, however, contains a splice junction >50 bp downstream of the premature termination codon in exon 12b, and thus would be expected to trigger NMD if translation initiation occurs upstream.

Because exon 12b inserts termination codons into all six reading frames, its inclusion within the $p290^{SETD2}$ ORF will always result in termination of upstream translation (irrespective of the actual identity of upstream exons). Exon 12b could therefore be considered a 'poison' exon. Interestingly, this type of alternative exon strongly correlates with highly conserved and ultra-conserved non-coding elements within genes encoding SR proteins and SR-related factors (Lareau *et al.*, 2007; Ni *et al.*, 2007; Saltzman *et al.*, 2008). Involved in both constitutive and alternative splicing, SR proteins can regulate their own levels of expression in a negative feedback loop: they promote inclusion of a highly conserved, poison exon within their own transcript, which triggers NMD and a decrease in SR protein levels (Lareau *et al.*, 2007; Ni *et al.*, 2007; Saltzman *et al.*, 2008). Given that exon 12b represents a highly conserved, poison exon (**Figure 5.3**), that SETD2 contains an RS-like domain and has roles in regulation of alternative splicing (discussed in section 1.4.3), and that $p290^{SETD2}$ is a relatively labile protein (**Figure 3.9**), it is tempting to speculate that a similar regulatory system may operate here.

Inhibition of translation and NMD by emetine treatment appears to increase the levels of transcripts detected by *Setd2* 3′ region probe, but not the >6.6 kb transcript. Furthermore, preliminary RT-PCR analysis indicates progressive increases in 12b»15 transcript with increasing duration of emetine treatment (up to two hours) and real-time RT-PCR analysis indicates an approximate ten-fold increase in splicing across the 12b−13 boundary after exposure to emetine for one hour (data not shown). The increase in transcripts could result from several mechanisms: (i) loss of labile nucleases to which *Setd2* transcripts are susceptible; (ii) emetine may act as an agonist, stimulating increased *Setd2* transcription; (iii) inhibition of translation-coupled RNA degradation—possibly via nonsense-mediated decay (NMD) or other regulatory processes, such as Staufen 1–mediated decay (Kim *et al.*, 2005).

5.5.3.4. *Novel ORF within 12−12b−13»21 cDNA*

Whilst the above transcripts contain only partially defined ORFs in frame with p290^{SETD2}, the 12−12b−13»21 cDNA contains a novel, complete ORF that would produce a protein related to SETD2 sequence.

Despite the complete sequence of the CA319728 clone demonstrating that *Setd2* definitively produces an alternative transcript in mouse brain, its existence in 10T½ cells is currently unclear, though evidence for the existence of 12b−13−14−15 was confirmed (**Figure 5.4A**).

Multiple rounds of DNA sequencing confirmed a nucleotide deletion (loss of adenine-1412) within exon 15 of this specific clone. This represents a deletion compared to the genome and all other exon 15–containing ESTs/cDNAs. Although it could represent RNA editing, a single observation suggests it more likely represents a single nucleotide polymorphism or a cloning error. For this reason, the sequence analysed in **Figure 5.9** retains adenine-1412.

The 12−12b−13»21 cDNA contains multiple ORFs in-frame with p290^{SETD2} sequence, which differ with respect to their initiation site (detailed in Table 5.1). The first ATG codon in-frame with p290^{SETD2} sequence occurs at +956 but three alternative, Kozak-compatible initiation codons exist upstream of this site (reviewed in (Wegrzyn *et al.*, 2008) and (Touriol

et al., 2003)). The most upstream initiation codon (TTG at +596) occurs within a Kozak-consensus and translation from this site would predict the generation of a protein of 61.0 kDa, containing both C-terminal and CT21 epitopes, similar to $p66^{C/CT21}$.

Table 5.1 – Predicted proteins within 12b–13»21 cDNA

Predicted mass (kDa)	Position* / Exon†	Canonical‡ consensus −+ 7654321123456789	Alternative‡ consensus −+ 7654321123456789	Epitopes C-term	Epitopes CT21	Human conservation
61.0	596/12b	cttGtttTGGAggtG	CCTtGttTTGGAGGTG	√	√	√
54.1	776/14	gGGCtCtcTGtCacca	gGgCtCtCTGtcaCca	√	√	×
49.1	904/14	cAGGAaGtTGtttgaa	CagGAAGTTGtttGaa	×	√	√
47.0	956/15	gcaacAGATGcAGaaC	gCAacAGATGcAGaaC	×	√	√

* position of predicted translation site within CA319728 clone.
† exon numbers correspond to murine *Setd2* (NCBI RefSeq: NM_001081340.2).
‡ canonical ATG versus alternative initiation consensus sequences are derived from (Wegrzyn *et al.*, 2008). Boldface indicates a match to the top two scoring nucleotides.

The potential for translation of the alternative ORFs within the 12–12b»21 cDNA *in vivo*, however, is dependent on how this cloned fragment was derived from original full-length transcripts.

The 12–12b»21 cDNA may derive from a variant transcript that lacks the upstream (1»11) exons—for instance, if it were to arise from an alternative promoter proximal to exon 12. Although exon 12 contains a short upstream ORF (<50 nt), the largest ORF initiates within exon 12b, starting at an alternative TTG initiation codon at +596—a site which is formed across the splice junction of exons 12b and 13. In this situation, the 12b»21 ORF at +596 could be translated by one of two mechanisms. Cap-dependent, leaky scanning may occur where the 40S ribosomal subunit may scan through the upstream ORF in exon 12 without translation initiation, subsequently forming an initiation complex at the downstream +596 site (Wang & Rothnagel, 2004). Alternatively, the upstream ORF may be translated, but the 40S subunit of the ribosome may remain associated with the 3′ UTR following termination of the upstream ORF; this post-termination complex could then undergo re-initiation to translate downstream ORFs (Kochetov *et al.*, 2008).

If the 12b»21 cDNA derives from a larger transcript (for example, 1»12−12b−13»21), exon 12b would prematurely terminate translation from the canonical $p290^{SETD2}$-encoding ORF within the upstream 1»12 region. Translation through this upstream region and premature termination would be expected to trigger NMD due to the splicing of exon 12b to exon 13, which occurs more than 50 nucleotides downstream of the premature termination site. For translation of the downstream 12b»21 ORF within this context, such a transcript must not only escape degradation by NMD but also permit initiation at the alternative translation initiation sites that occur downstream of the termination codon in exon 12b. This could putatively occur by either the re-initiation mechanism discussed above, or by Cap-independent initiation through the use of internal ribosome entry sites (Hellen & Sarnow, 2001).

5.6. Conclusion

The objective of this chapter was to characterise *Setd2* transcripts in 10T½ cells. Northern blotting detected multiple transcripts with 3′ region *Setd2* probes; RT-PCR and DNA-sequencing verified the existence of multiple, mutually exclusive transcript variants, differing in two alternative exons and three regions of intron retention, in 10T½ cells (**Figure 5.10**). Such transcript diversity may be relevant to observation of putative protein isoforms detected exclusively by either N-terminal or C-terminal antibodies (Chapters 3 and 4). Additionally, a cDNA variant from mouse brain was identified that contains a distinct, novel ORF with potential to encode a protein with similar characteristics to $p66^{C/CT21}$. The relationship between this transcript variant and $p66^{C/CT21}$ will be investigated in the next chapter.

Figure 5.10 Summary of alternatively spliced transcripts of murine *Setd2* in 10T½ fibroblasts

Schematic diagram of alternatively spliced *Setd2* transcripts in murine 10T½ fibroblasts (confirmed by DNA sequencing). Pink represents intronic sequence. The green arrows and red crossed-circles represent predicted sites of initiation and termination of translation, respectively.

(A) 8–9.5 transcript

(B) 6»10.5 transcript

(C) 17–18.5 transcript

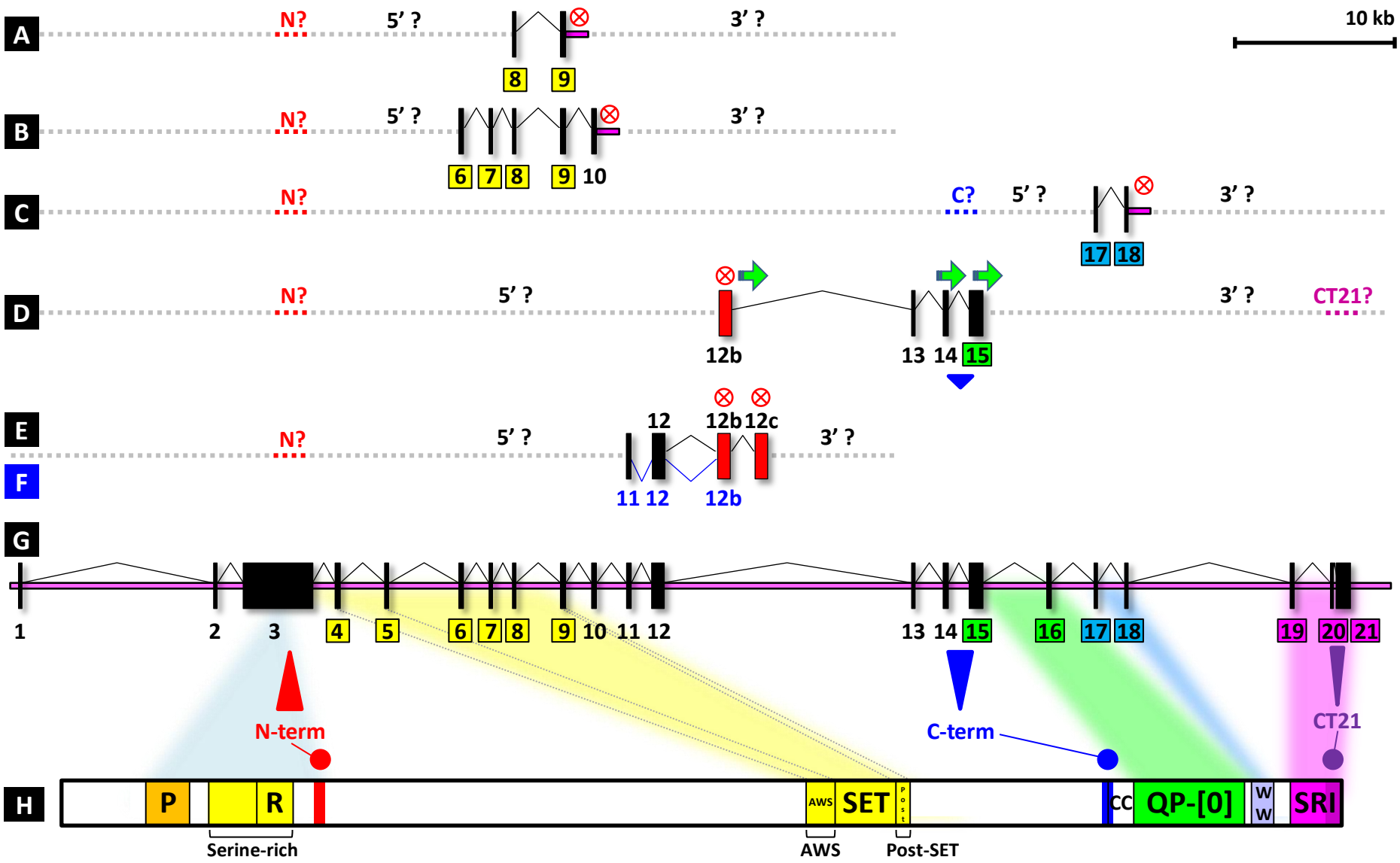
(D) 12b–13–14–15 transcript

(E) 12–12b–12c transcript: splicing indicated in black.

(F) 11–12–12b transcript: splicing indicated in blue.

(G) 1»21 transcript: existence of the canonical transcript (RefSeq NM_001081340.2) is inferred based on immunodetection of $p290^{SETD2}$ protein and detection of a transcript > 6.6 kb by northern blotting.

(H) SETD2 protein structure: approximate correlation between protein domains, antibody epitopes and exons in the 1»21 transcript is indicated.


A
B
C
D
E
F
G
H
N?
5' ?
3' ?
10 kb
8
9
6
7
8
9
10
C?
17
18
12b
13
14
15
CT21?
12
12b
12c
11
12
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
N-term
C-term
CT21
P
R
AWS
SET
Post
CC
QP-[0]
WW
SRI
Serine-rich
AWS
Post-SET

Chapter 6 Ectopic expression of a SETD2 isoform encoded by a variant transcript and comparison with p66$^{C/CT21}$

6.1. Introduction

The novel alternatively spliced *Setd2* cDNA (12−12b−13»21) harbours a predicted ORF for a 61 kDa protein containing both C-terminal and CT21 epitopes (Chapter 5). These characteristics are reminiscent of the putative SETD2 isoform, p66$^{C/CT21}$, which is detected with antibodies raised against both CT21 and C-terminal epitopes (Chapters 3 and 4). The major objective of this chapter is to ectopically express this variant SETD2 isoform, validate its existence immunologically with these antibodies, and compare its physical and extraction properties with endogenous p66$^{C/CT21}$.

Chapters 3 and 4 demonstrated specific properties of p66$^{C/CT21}$, distinct from other α-SETD2 immunoforms:

(i) Enrichment in nuclease–high-salt–resistant fractions (**Figure 4.2**)

(ii) Enrichment in specific subcellular fractions prepared by differential detergent extraction and centrifugation (DDC) (**Figure 4.7**)

(iii) Characteristic migration by 2D−AU/SDS−PAGE (**Figure 4.9**)

(iv) Requirement of thermo-urea denaturation for detection with C-terminal antibody after SDS-PAGE (**Figure 3.6**)

These characteristics provide multiple diagnostic criteria to assess whether ectopically expressed proteins derived from *Setd2*(13»21) cDNA correspond to endogenous p66$^{C/CT21}$ protein. To distinguish ectopic SETD2(13»21) from endogenous p66$^{C/CT21}$, epitope-tagged *Setd2*(13»21) expression constructs were generated as N-terminal 3×FLAG or C-terminal

HA forms, expressed from plasmids containing a cytomegalovirus (CMV) promoter. These constructs were transfected into 10T½ fibroblasts to compare their subcellular fractionation and immunodetection characteristics compared to p66$^{C/CT21}$.

In summary, these experiments demonstrated that the predicted SETD2(13»21) protein does not fully recapitulate endogenous p66$^{C/CT21}$—the identity of p66$^{C/CT21}$ remains unclear. However, as a representative C-terminal-specific isoform derived from the 12−12b−13»21 transcript variant, SETD2(13»21) exhibits unique fractionation properties, which are not shared with p290^{SETD2}. The fractionation and immunodetection of SETD2(13»21) also influences interpretation of the exact nature of the verified p290^{SETD2} methyltransferase itself.

6.2. Results

6.2.1. Design of constructs to express epitope-tagged SETD2(13»21) proteins

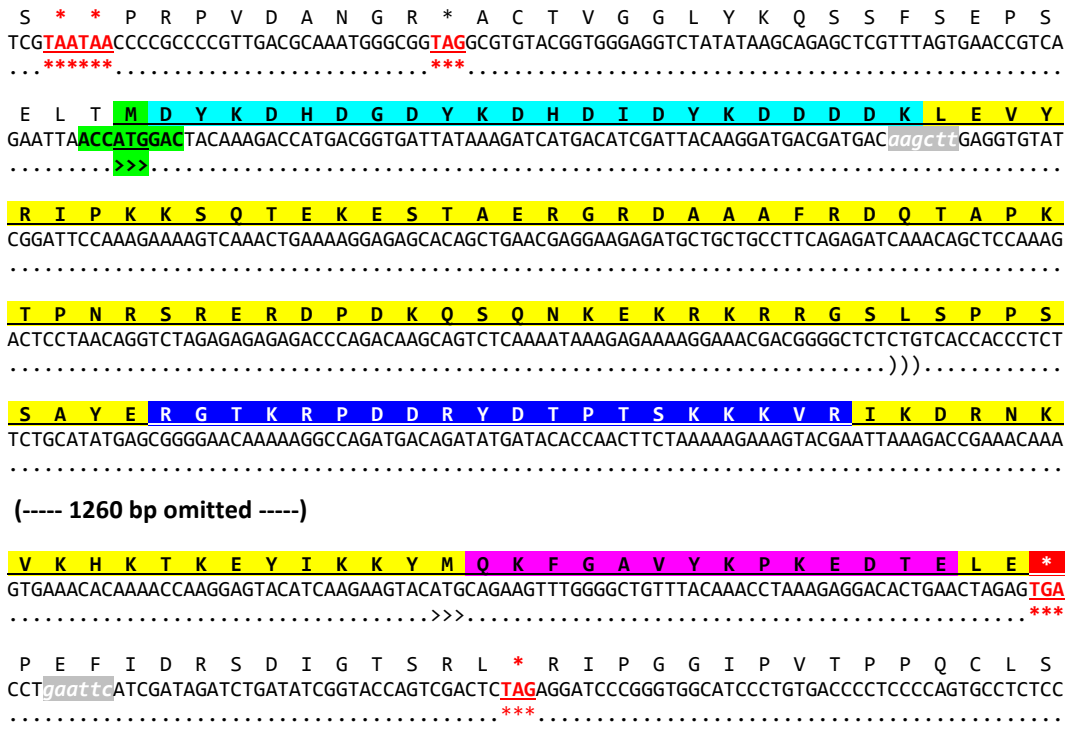
Distinct protein isoforms, defined by N-terminal truncation, can derive from alternative translation initiation ((Ivanov *et al.*, 2011); reviewed in (Kochetov, 2008)): precedent exists for initiation from UUG, CUG, GUG, ACG and AUA codons (Scharnhorst *et al.*, 1999; Wegrzyn *et al.*, 2008; Gerashchenko *et al.*, 2010) in addition to use of multiple internal sites (Scharnhorst *et al.*, 1999; Kobayashi *et al.*, 2009). Although alternative translation initiation has only been experimentally confirmed in a limited number of cases, bioinformatic analyses predict a more widespread occurrence in mammals than is currently appreciated (Wegrzyn *et al.*, 2008; Ivanov *et al.*, 2011).

Within the alternative *Setd2*(12−12b−13»21) cDNA, multiple ORFs in-frame with p290^{SETD2} protein sequence are predicted (**Table 5.1**). The first AUG codon in a Kozak consensus sequence resides at +956, predicting a ≈47 kDa protein, containing the CT21 epitope but lacking the C-terminal epitope. Upstream of this AUG, two alternative Kozak-compatible initiation sequences exist at +596 and +776, which would both produce proteins predicted to contain both C-terminal and CT21 epitopes, consistent with p66$^{C/CT21}$. The +596 sequence

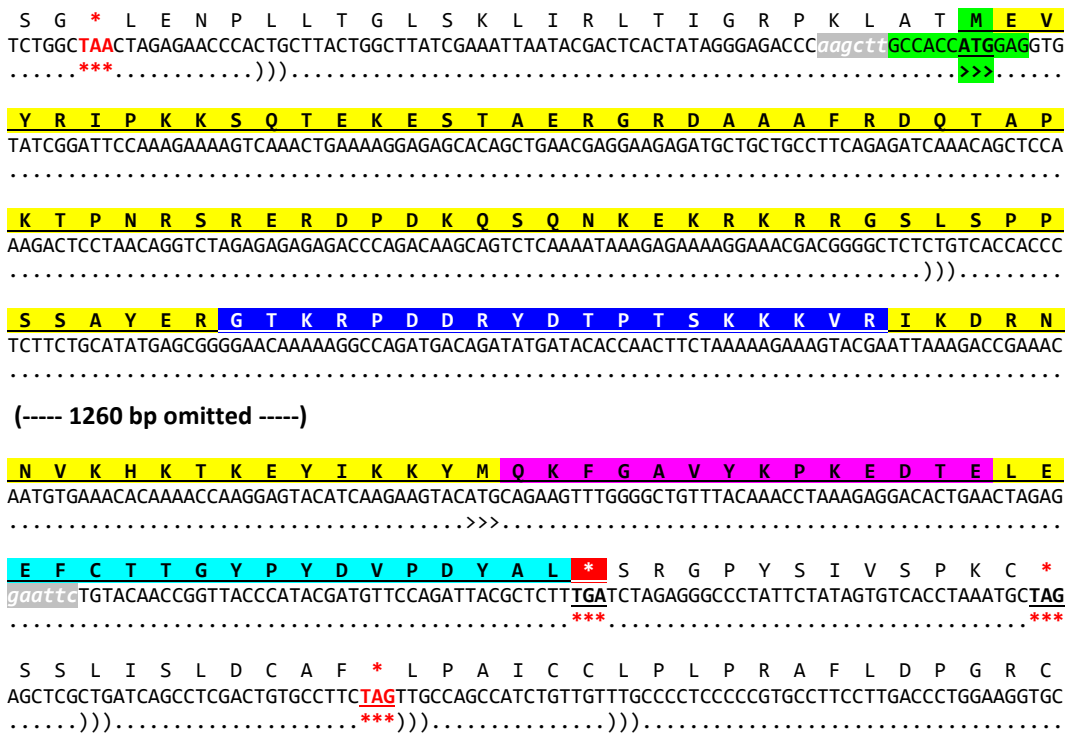
Figure 6.1 Epitope-tagged SETD2(13»21) constructs

Nucleotide and corresponding amino acids sequences of *Setd2*(13»21) cloned into p3×FLAG−CMV-10 **(A)** and pcDNA3−HA **(B)** epitope-tagging vectors. SETD2-related sequence is highlighted in yellow with C-terminal and CT21 epitopes emphasised in blue and magenta, respectively. Translation initiation and termination sequences are highlighted in green and red, respectively. The FLAG and HA epitopes are highlighted in cyan. Flanking non-highlighted text represents vector sequence, which contains upstream and downstream termination codons (red text). EcoRI and HindIII sites used for cloning are marked as white italicised text on grey background.

(A) N-terminal FLAG-tagged SETD2(13»21) construct



(B) C-terminal HA-tagged SETD2(13»21) construct



was deemed more important than +776 for potential p66$^{C/CT21}$ generation for several reasons: firstly, the former predicts a protein of 61.0 kDa, which is more consistent with the apparent mass of p66$^{C/CT21}$ judged by SDS-PAGE, than the 54.0 kDa protein initiated from +776. Secondly, the +596 sequence (**CCTtGttTTGGAGGTG**) more closely matches the consensus sequence for alternative initiation than +776 (Wegrzyn *et al.*, 2008). Thirdly, +596 is conserved between human and mouse and is specifically contributed by exon 12b–inclusion. The importance of this latter factor was justified by high conservation of exon 12b in human and mouse (**Figure 5.3**) and detection of p66$^{C/CT21}$ in both human and murine cells (**Figure 3.7**).

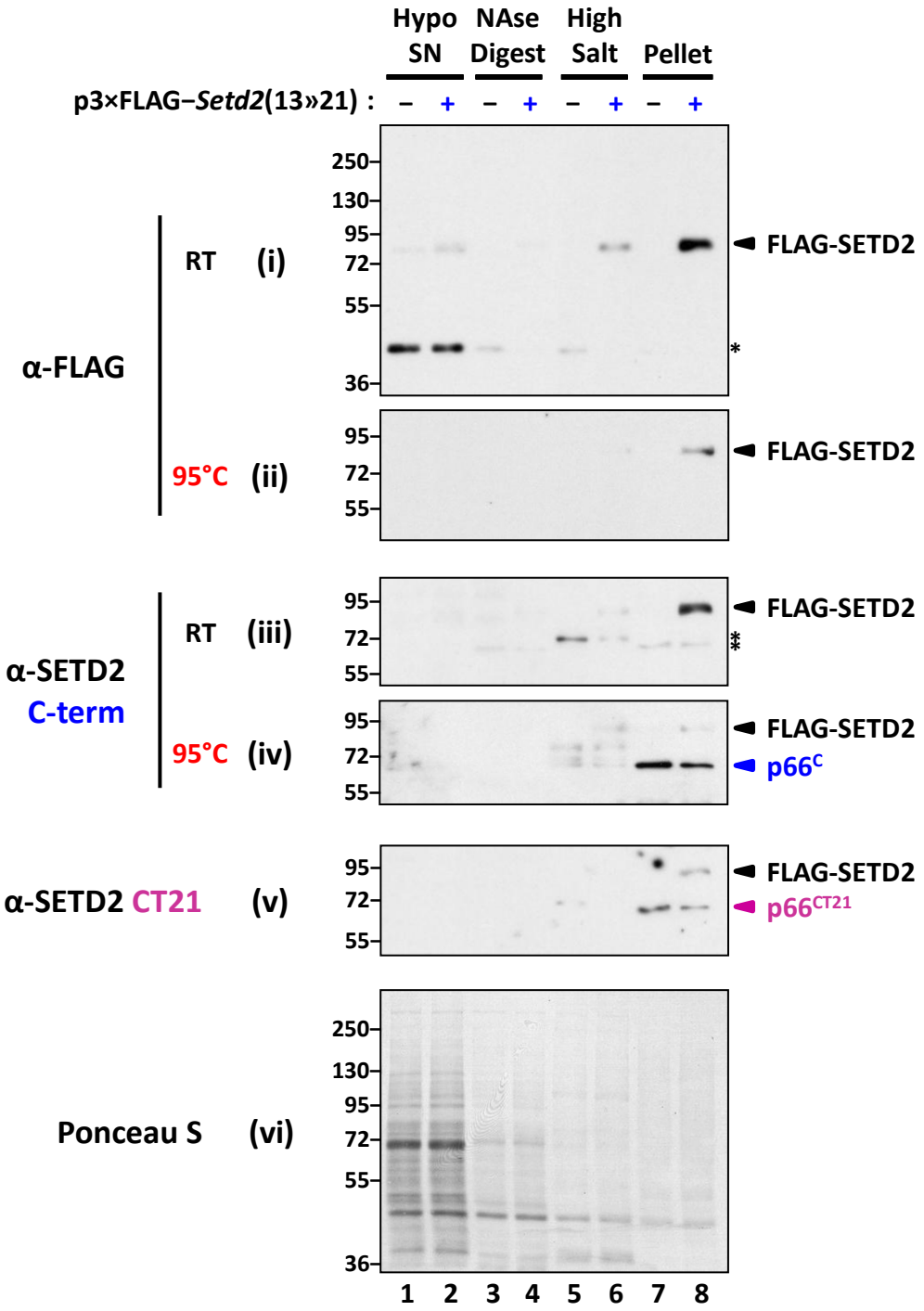
I therefore generated N-terminal FLAG-tagged and C-terminal HA-tagged SETD2(13»21) constructs in mammalian over-expression vectors, which were both verified by multiple rounds of DNA sequencing (**Figure 6.1**). Both constructs were engineered to express SETD2(13»21) from strong AUG-containing Kozak-sequences, to promote maximal expression, but maintain SETD2 protein sequence comparable to the +596-initiated ORF.

6.2.2. SETD2(13»21) and p66$^{C/CT21}$ co-fractionate by nuclease–high-salt extraction but exhibit different electrophoretic migration by SDS-PAGE

10T½ fibroblasts were transfected with vector-only or N-terminal 3×FLAG-tagged SETD2(13»21) constructs and fractionated by nuclease–high-salt extraction (**Figure 4.1**). Immunoblotting with α-FLAG antibody detected FLAG–SETD2(13»21) in extracts derived from cells transfected with p3×FLAG–*Setd2*(13»21)–CMV-10 but not empty vector (**Figure 6.2-i**). Surprisingly, FLAG–SETD2(13»21) migrated by SDS-PAGE with an apparent molecular mass of 88 ± 2 kDa (n = 3) (**Figure 6.2-i/ii**), which is not only distinct from that of p66$^{C/CT21}$ (**Figure 6.2-iii**) but is also anomalous compared to its predicted molecular mass of the epitope-tagged sequence (64.0 kDa). Immunodetection of additional 88 kDa signals, exclusively in FLAG–SETD2(13»21)-transfected extracts, by both α-SETD2 C-terminal and CT21 antibodies confirmed these bands as FLAG-tagged SETD2(13»21) (**Figure 6.2**:**iii–v**).

Figure 6.2 FLAG-tagged SETD2(13»21) co-fractionates with p66$^{C/CT21}$ after nuclease–salt extraction but migrates anomalously at ≈88 kDa

Expression of FLAG–SETD2(13»21) protein in 10T½ fibroblasts: cells were transfected with either p3×FLAG-CMV-10 vector alone (odd lanes) or p3×FLAG–*Setd2*(13»21)–CMV-10 construct (even lanes) by electroporation. Quiesced cells were harvested 48 hours after transfection, by hypotonic lysis and sequential extraction by nuclease digestion and high-salt extraction. Soluble extracts were precipitated with TCA and all samples were resolubilised in urea/SDS-PAGE buffer at room temperature **(i, iii, v, vi)** or at 95°C for 10 minutes **(ii, iv)**. Equivalent proportions of each sample were separated by 8% SDS-PAGE. Gels were transferred to PVDF in Towbin buffer; membranes were stained with Ponceau S **(vi)** before immunoblotting with α-FLAG **(i, ii)**, α-SETD2 C-terminal **(iii, iv)** or α-SETD2 CT21 **(v)** antibodies. Asterisks indicate background immunoreactive bands, which in **(iii)** may correspond to weak detection of p66^{C} in the absence of thermo-urea denaturation due to the long exposure needed for detection of FLAG–SETD2.


Hypo SN
NAse Digest
High Salt
Pellet
p3×FLAG–Setd2(13»21) : – + – + – + – +
α-FLAG
RT (i)
95°C (ii)
α-SETD2 C-term
RT (iii)
95°C (iv)
α-SETD2 CT21 (v)
Ponceau S (vi)
250
130
95
72
55
36
FLAG-SETD2
p66C
p66CT21
*
**
1 2 3 4 5 6 7 8

Since p66^{C} absolutely required thermo-urea denaturation for optimal detection (**Figure 3.6A**), effects of this condition on epitope-tagged proteins were also investigated. Thermo-urea denaturation, however, impaired detection of FLAG–SETD2(13»21) and importantly, no additional α-FLAG-reactive bands appeared (**Figure 6.2-ii** versus **i**), in contrast to the appearance of p66^{C} by α-SETD2 C-terminal antibody (**Figure 6.2-iv** versus **iii**).

Despite the above differences in electrophoretic migration and effects of denaturation, SETD2(13»21) and p66^{CT21} exhibited similar resistance to extraction by nuclease digestion and high-salt extraction (**Figure 6.2-i/iii**, lane 8). This is distinct from previously observed fractionation of p290^{SETD2} and p140^{C} predominantly in nuclease-digest and high-salt extracts (**Figure 4.2**).

Similar immunodetection and fractionation of p66$^{C/CT21}$ here compared to previous experiments (**Figures 3.4** and **4.2**) indicates that transfection did not noticeably affect the properties of endogenous p66$^{C/CT21}$. However, both p140^{C} and p290^{SETD2} were barely detectable in all transfection experiments presented in this chapter (data not shown): although this is likely a technical issue (since problems with their immunodetection were concurrently encountered in independent, unrelated experiments), lack of their detection even in vector-only transfections means that no rigorous, conclusive analysis can be made about these proteins. Their fractionation alongside SETD2(13»21) will be important to re-assess concurrently, because SETD2(13»21) did not appear to recapitulate the fractionation of p290^{SETD2} inferred from previous experiments (**Figures 4.2** and **4.7**), suggesting that N-terminal regions of p290^{SETD2} may play more crucial roles in subcellular interactions than previously appreciated. Whether SETD2(13»21) transfection substantially influences expression of p290^{SETD2} and p140^{C} (and indeed H3K36 methylation) remains inconclusive and deserves resolution in future.

Despite CMV-mediated expression, SETD2(13»21) protein did not appear to be substantially over-expressed, since no additional Ponceau S–stained band was visible in subcellular fractions derived from FLAG–SETD2(13»21) transfection compared to those

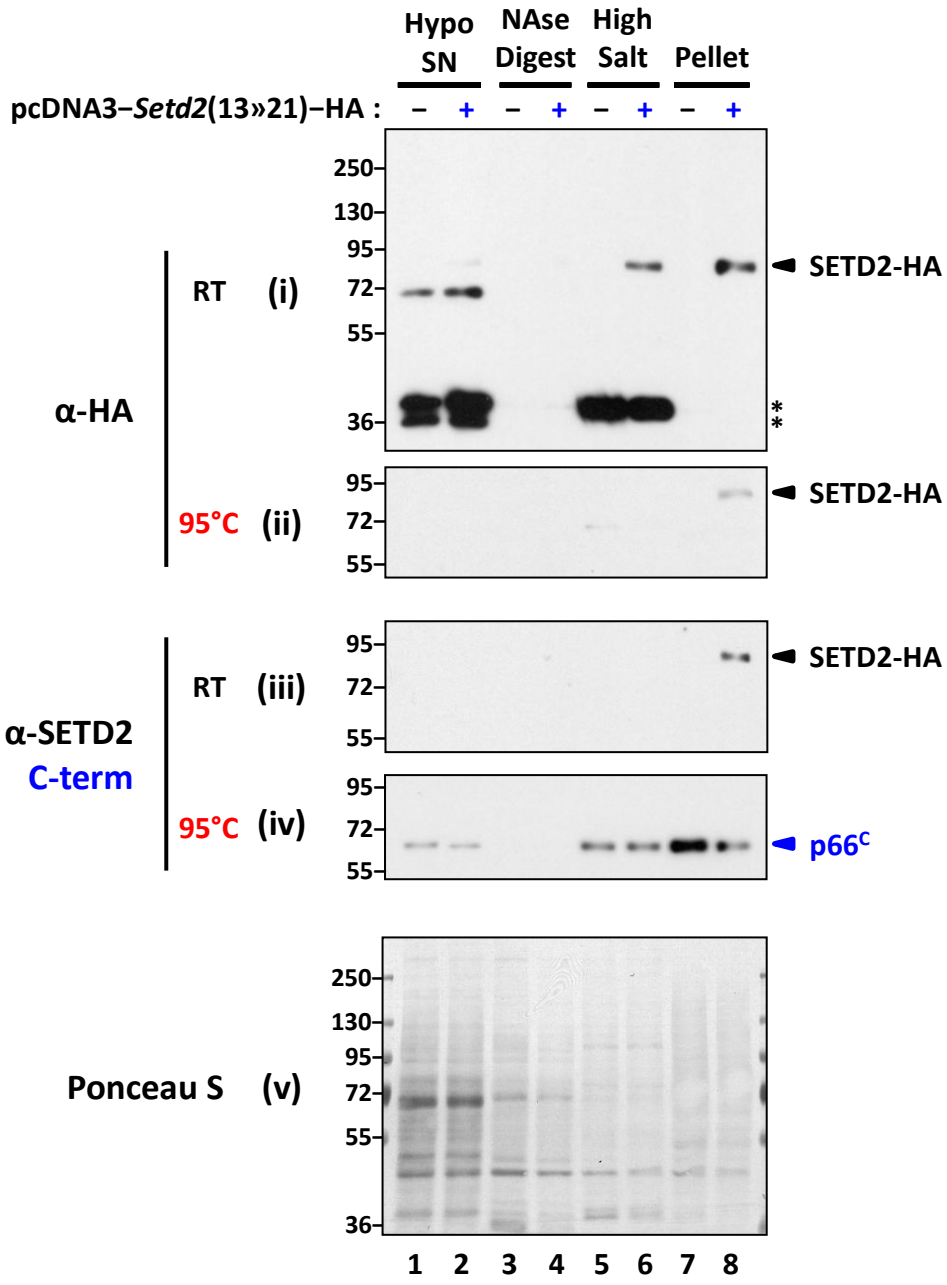
derived from the vector-only control (**Figure 6.2-vi**). Furthermore, immunodetection of similar levels of SETD2(13»21) and p66^{CT21} by α-SETD2 CT21 antibody (**Figure 6.2-v**) suggest that epitope-tagged protein expression is comparable to endogenous p66^{CT21} (assuming equivalent capacities for epitope recognition). A quantitative comparison of SETD2(13»21) against p66^{C} is complicated, however, by the different effects of thermo-urea denaturation on the intensities of immunoreactive bands. Detection of SETD2(13»21) by α-CT21 antibody (**Figure 6.2-v**) is noteworthy, because this crucially verifies detection of cognate SETD2 epitope—despite depletion of no α-CT21 signals (including 'p290^{CT21}') by RNAi (**Figure B.2** in Appendix B—implications of this for p66$^{C/CT21}$ and p290^{SETD2} will be discussed below).

In summary, although both epitope-tagged SETD2(13»21) proteins and p66$^{C/CT21}$ fractionated comparably by nuclease–high-salt extraction, their differences in electrophoretic mobility and denaturation requirements demonstrated that SETD2(13»21) does not fully recapitulate p66$^{C/CT21}$ and rather reflects properties of a potential C-terminal isoform with its own distinct properties *in vivo*.

To assess the generality of anomalous migration of SETD2(13»21), nuclease–high-salt fractionation was repeated for fibroblasts transfected with N-terminal HA-tagged SETD2(13»21) construct (**Figure 6.3**). SETD2(13»21) was detected by α-HA antibody only in extracts transfected with SETD2(13»21)–HA construct (**Figure 6.3-i/ii**). Comparable to FLAG-tagged protein, HA-tagged SETD2(13»21) exhibited anomalous migration by SDS-PAGE with an apparent molecular mass of ≈85 kDa (compared to 63.0 kDa predicted for tagged protein)

Figure 6.3 HA-tagged SETD2(13»21) co-fractionates with p66$^{C/CT21}$ after nuclease–salt extraction but migrates anomalously at ≈85 kDa

Expression of SETD2(13»21)–HA in 10T½ fibroblasts: cells were transfected in the absence (odd lanes) or presence of pcDNA3–*Setd2*(13»21)–HA construct (even lanes). Quiesced cells were harvested 48 hours after transfection, by hypotonic lysis and sequential fractionation by nuclease digestion and high-salt extraction. Soluble extracts were precipitated with TCA and all samples resolubilised in urea/SDS-PAGE buffer at room temperature **(i, iii, v)** or at 95°C for 10 minutes **(ii, iv)**. Equivalent proportions of each fraction were separated by 8% SDS-PAGE and then transferred to PVDF in Towbin buffer. Membranes were stained with Ponceau S **(v)** before immunoblotting with α-HA **(i, ii)** or α-SETD2 C-terminal **(iii, iv)** antibodies. Asterisks indicate non-specific bands unrelated to SETD2 constructs.


Hypo SN
NAse Digest
High Salt
Pellet
pcDNA3–Setd2(13»21)–HA :
– + – + – + – +
250
130
95
72
55
36
SETD2-HA
α-HA
RT (i)
95°C (ii)
α-SETD2 C-term
RT (iii)
95°C (iv)
p66C
Ponceau S (v)
1 2 3 4 5 6 7 8

(**Figure 6.3-i**) and was detectable by α-SETD2 C-terminal antibody in the absence of thermo-urea denaturation (**Figure 6.3-iii**). Lack of detection of SETD2(13»21)−HA by α-SETD2 C-terminal antibody after heating at 95°C (**Figure 6.3-iv**, lane 8) is likely due to insufficient exposure (inferred from **Figure 6.5** below: panel **iii** versus **iv**, lane 10). A small proportion of SETD2(13»21)−HA was extracted by high-salt buffer (**Figure 6.3-i**, lane 6) but most remained in the pellet resistant to extraction by nuclease digestion and high-salt buffer (**Figure 6.3-i**, lanes 7 and 8). This parallels the extractability of p66^{C} (**Figure 6.3-iv**) and recapitulates the fractionation of FLAG-SETD2(13»21) (**Figure 6.2**).

Although both FLAG- and HA-tagged SETD2(13»21) proteins fractionated similarly to p66$^{C/CT21}$, their substantial differences in migration by SDS-PAGE compared to p66$^{C/CT21}$ demonstrated that epitope-tagged SETD2(13»21) does not represent endogenous p66$^{C/CT21}$. Based on these differences, further analysis by 2D−AU/SDS−PAGE was not deemed useful.

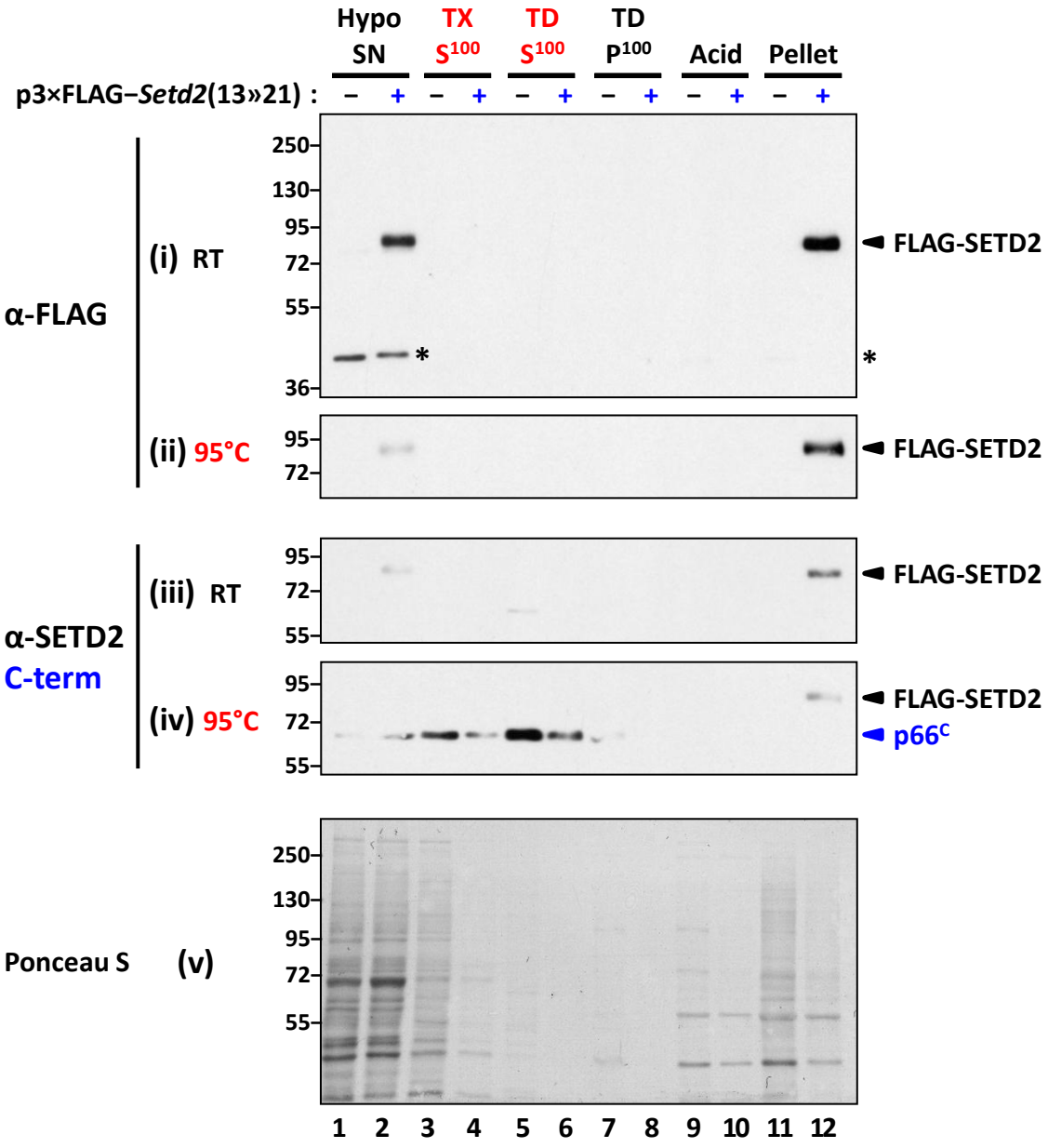
6.2.3. SETD2(13»21) does not recapitulate detergent-extractability of p66$^{C/CT21}$

Despite differences in electrophoretic migration, epitope-tagged SETD2(13»21) and p66$^{C/CT21}$ exhibited similar fractionation by nuclease/high-salt extraction. I therefore investigated whether SETD2(13»21) would also co-fractionate with p66$^{C/CT21}$ after differential detergent extraction and centrifugation.

DDC-fractionation of 10T½ cells transfected with FLAG−SETD2(13»21) constructs recapitulated specific detection of FLAG−SETD2(13»21) and its anomalous ≈88 kDa migration, by both α-FLAG and α-SETD2 C-terminal antibodies (**Figure 6.4**). However, whilst endogenous p66^{C} fractionated predominantly in TX-S^{100} and TD-S^{100} fractions (**Figure 6.4-iv**, lanes 3−6), FLAG-tagged SETD2(13»21) was instead enriched in the final pellet (**Figure 6.4:i–iv**, lane 12). The SETD2(13»21) extracted in the initial hypotonic supernatant (**Figure 6.4:i–iii**, lane 2) is not reproducibly observed (*cf.* Hypo-SN in **Figures 6.2, 6.3 and 6.5**) and may be contributed by minor contamination from the insoluble pellet during removal of the initial hypotonic supernatant, since a small amount of p66^{C} is also detected in this fraction (**Figure 6.4-iv**, lanes 1 and 2). Again, whilst thermo-urea denaturation revealed p66^{C} (**Figure 6.4-iv**), no additional α-FLAG signals in

Figure 6.4 FLAG-tagged SETD2(13»21) does not recapitulate the DDC fractionation properties of p66$^{C/CT21}$

DDC fractionation of 10T½ fibroblasts expressing FLAG–SETD2(13»21): cells were transfected with either p3×FLAG-CMV-10 vector alone (odd lanes) or p3×FLAG–*Setd2*(13»21)–CMV-10 construct (even lanes) by electroporation. Quiesced cells were harvested 48 hours after transfection, by differential detergent extraction and centrifugation (DDC), with an additional acid extraction of the Triton X-100-resistant (TX-P) fraction to yield acid-soluble extracts (lanes 7 and 8) and final insoluble pellets (lanes 9 and 10). Soluble samples were precipitated with TCA, resolubilised in urea/SDS-PAGE buffer, and incubated at room temperature **(i, iii, v)** or at 95°C for 10 minutes **(ii, iv)**. Equivalent proportions of each fraction were separated by 8% SDS-PAGE. Gels were transferred to PVDF in Towbin buffer; membranes were stained with Ponceau S **(v)** before immunoblotting with α-FLAG **(i, ii)** or α-SETD2 C-terminal **(iii, iv)** antibodies. Asterisks indicate non-specific bands unrelated to SETD2 constructs.

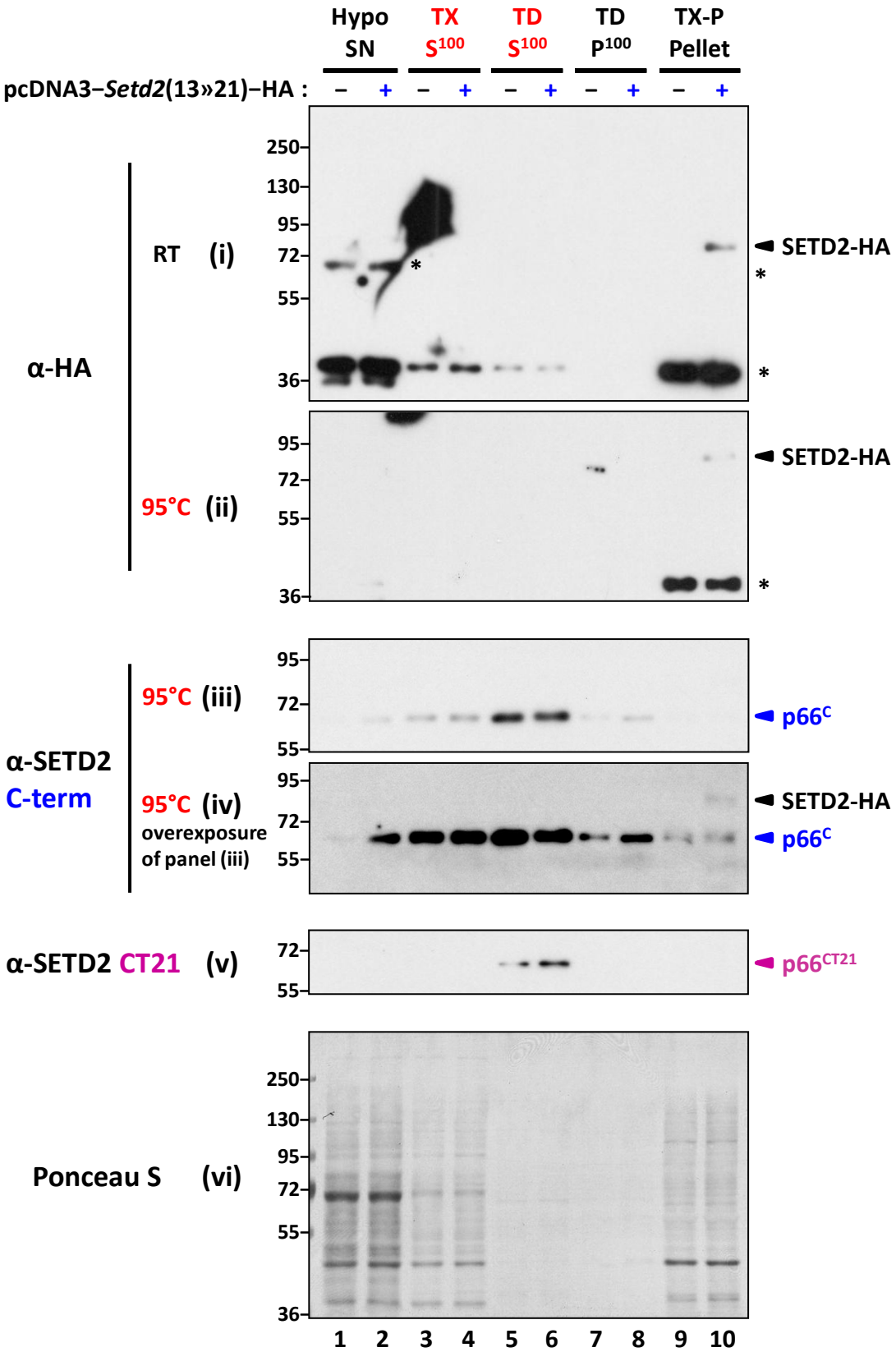

Hypo SN
TX S100
TD S100
TD P100
Acid
Pellet
p3×FLAG–Setd2(13»21) :
– + – + – + – + – + – +
α-FLAG
(i) RT
(ii) 95°C
α-SETD2 C-term
(iii) RT
(iv) 95°C
Ponceau S
(v)
250
130
95
72
55
36
FLAG-SETD2
p66C
*
1 2 3 4 5 6 7 8 9 10 11 12

any subcellular fraction were detected (**Figure 6.4-ii**) compared to samples solubilised at room temperature (**Figure 6.4-i**). Transfectants with HA-tagged construct exhibited a similar protein distribution after DDC fractionation: SETD2(13»21)−HA enrichment in the Triton X-100−resistant pellet (**Figure 6.5-i/ii/iv,** lane 10); p66^{C} and p66^{CT21} fractionated predominantly in TD-S^{100} and were virtually absent from TX-P (**Figure 6.5-iii/v**).

Thermo-urea denaturation also did not reveal any additional α-FLAG-reactive signal and led to a decrease in detection of α-FLAG signal and also α-C-terminal signal. This is in contrast to p66^{C}, which was optimally detected by α-SETD2 C-terminal antibody only after thermal denaturation (**Figure 6.4-iv**). Although levels are hard to rigorously compare due to differential effects of thermo-urea denaturation, over-exposure of p66^{C} signal shows that even after overexpression, the detection of FLAG–SETD(13»21) is much weaker than p66^{C}. This is probably due to denaturation-mediated proteolysis or epitope-modification (as observed for other SETD2-reactive bands, such as p120^{N} or p290^{SETD2}). Clearly, the SETD2(13»21) construct does not recapitulate the striking increase in p66^{C} immunodetection after thermo-urea denaturation.

Figure 6.5 HA-tagged SETD2(13»21) does not recapitulate the DDC fractionation properties of p66$^{C/CT21}$

DDC fractionation of 10T½ fibroblasts expressing SETD2(13»21)–HA: cells were electroporated in the absence (odd lanes) or presence of pcDNA3–*Setd2*(13»21)–HA construct (even lanes); quiesced cells were harvested 48 hours after transfection, by the DDC method. Samples were precipitated with TCA, resolubilised in urea/SDS-PAGE buffer, and incubated at room temperature **(i, v, vi)** or at 95°C for 10 minutes **(ii–iv)**. Equivalent proportions of each fraction were separated by 8% SDS-PAGE and then transferred to PVDF in Towbin buffer. Membranes were stained with Ponceau S **(vi)** before immunoblotting with α-HA **(i, ii)**, α-SETD2 C-terminal **(iii, iv)** or α-SETD2 CT21 **(v)** antibodies. Panel **(iv)** represents an over-exposure of panel **(iii)**. Asterisks indicate non-specific bands unrelated to SETD2 constructs.


Hypo SN
TX S100
TD S100
TD P100
TX-P Pellet
pcDNA3–Setd2(13»21)–HA :
– + – + – + – + – +
α-HA
RT (i)
95°C (ii)
250
130
95
72
55
36
SETD2-HA
*
α-SETD2 C-term
95°C (iii)
95°C (iv) overexposure of panel (iii)
p66C
α-SETD2 CT21 (v)
p66CT21
Ponceau S (vi)
1 2 3 4 5 6 7 8 9 10

6.3. Discussion

6.3.1. Summary of results

- HA- and FLAG-tagged SETD2(13»21) proteins expressed in 10T½ cells migrated anomalously by SDS-PAGE with apparent molecular mass of ≈85–88 kDa (**Figures 6.2–6.5**).

- SETD2(13»21) co-fractionated with p66$^{C/CT21}$ by nuclease–high-salt extraction (**Figures 6.2** and **6.3**) but not by differential detergent extraction (**Figures 6.4** and **6.5**).

- SETD2(13»21) did not fully recapitulate fractionation characteristics of either endogenous p66$^{C/CT21}$ or p290^{SETD2} in 10T½ cells (*cf.* **Figures 4.2** and **4.7**).

6.3.2. Anomalous migration of SETD2(13»21)

Apparent molecular masses of FLAG- and HA-tagged proteins estimated by SDS-PAGE are 88 ± 2 kDa and 85 ± 2 kDa, respectively. Fusion of 3×FLAG- or HA-tag adds 3.0 or 2.0 kDa to the predicted 61 kDa mass of SETD2(13»21) to produce epitope-tagged constructs of 64 and 63 kDa, respectively. After subtracting the contributions from the two different epitope tags, the average mass of untagged SETD2(13»21) protein was estimated as 84 ± 2 kDa (n = 4). Therefore, the mass of untagged SETD2(13»21) estimated by SDS-PAGE is ≈23 kDa greater than that predicted from sequence (≈40% deviation), and ≈18 kDa greater than the apparent mass of p66$^{C/CT21}$ (≈30% deviation).

Since both sequence constructs were completely verified by DNA sequencing and restriction digestion, the anomalous migration is likely due to inherent properties of the expressed protein: either the inherent structure of the SETD2(13»21) protein, or differential modification *in vivo*. Although epitope tags—small and terminally fused—are unlikely to significantly affect protein migration directly, their influence cannot formally be ruled out, without analysis of expression of untagged protein in parallel; they may influence protein folding, interactions or post-translational modifications *in vivo*, and thus indirectly

contribute towards production of a protein defined by more than just its predicted primary structure.

Anomalous migration may be due to differential SDS-binding (Rath *et al.*, 2009) or residual protein structure. Both of these may result from inherent protein characteristics (hydrophobicity; QP-rich region; predicted coiled-coils, for example, in the C-terminal region); or they may result from differential post-translational modification of SETD2(13»21) *in vivo*, compared to proteins containing identical SETD2 sequence (*i.e.* p290^{SETD2}; potentially p66$^{C/CT21}$). Anomalous migration by SDS-PAGE is not uncommon; for example, p53 is a 43.7 kDa protein, which runs as 53 kDa, due to high proline content (Hansen *et al.*, 1998b). Considering SETD2(13»21) contains the QP-rich region followed by a region of extreme low-charge, such composition likely contributes to anomalous migration of SETD2(13»21). As discussed earlier, the C-terminal sequence encoded within exons 13−21 is significantly enriched in proline, glutamine and depleted for charged amino acids. This is consistent with anomalous SDS-PAGE migration of *in vitro* translated c-*myc*, c-*fos*, p53 and adenovirus E1A, which all contain proline-rich regions (MacConnell & Verma, 1983; Watt *et al.*, 1985; Yee & Branton, 1985; Hansen *et al.*, 1998b). Histones also migrate slower than would be expected for their mass, likely due to high content of lysine and arginine, reducing expected net negative charge of SDS–histone complexes. Histone H4 has a predicted molecular mass of 11.2 kDa but migrates anomalously by SDS-PAGE (**Figure 4.2A**). Post-translational modification *in vivo* may also influence migration; for example, single serine-phosphorylation retards migration of adenovirus E1A by 2−4 kDa (Smith *et al.*, 1989). Addition a single phosphoryl group cannot explain an increase of 2−4 kDa by mass-charge addition alone; modifications are thus likely to affect SDS interaction or influence residual secondary structures. *In vitro* translation would be useful for assessing this possibility.

Such anomalous migration confounds estimation of isoform mass—having implications for predicting relationships between putative SETD2 isoforms and ORFs. This has implications for prediction of other SETD2 isoforms—if similar regions that confer anomalous migration are present in p66$^{C/CT21}$ or p140^{C}, then would these proteins also migrate

anomalously? For example, p66$^{C/CT21}$ may in reality be a <66 kDa protein (if assuming a similar % deviation from the predicted mass of SETD2(13»21), p66$^{C/CT21}$ would be ≈45 kDa. Such anomalous behaviour cannot be extrapolated to other forms conclusively, but its demonstration for a defined SETD2 C-terminal segment highlights an additional factor that needs to be considered, which further complicates prediction in relationships between *Setd2* cDNAs and putative protein isoforms.

6.3.3. Subcellular fractionation of SETD2(13»21)

After nuclease–high-salt fractionation, p290^{SETD2} and p140^{C} distribute predominantly in nuclease-digest and high-salt fractions, whereas p66$^{C/CT21}$ remains largely resistant to extraction (**Figure 4.2**). Conversely, by DDC fractionation, p66$^{C/CT21}$ is virtually all extracted by Triton buffer into TX-S^{100} and TD-S^{100} fractions while p290^{SETD2} and p140^{C} are resistant to Triton extraction (**Figure 4.7**).

Interestingly, SETD2(13»21) does not fully recapitulate fractionation behaviour of either p290^{SETD2} or p66$^{C/CT21}$: it is consistently detected in the pellet fraction, resistant to extraction by nuclease/high-salt buffer and by Triton-extraction buffer. Enrichment of SETD2(13»21) in insoluble material is unlikely to reflect incorporation into 'inclusion bodies' from overloading the translation system due to overexpression, since no additional Ponceau S–stainable bands are discernible—even in subcellular fractions highly depleted of protein. Furthermore, α-CT21 antibody detects comparable levels of p66^{CT21} and SETD2(13»21) (**Figure 6.2**).

Different nuclease/salt fractionation for SETD2(13»21) and p290^{SETD2} suggests that N-terminal domains present in p290^{SETD2} but lacking in SETD2(13»21) are required to confer such extractability. This highlights a largely neglected importance of N-terminal regions for specifying subcellular interactions; SET domains have been implicated solely in methyltransferase activity, rather than influencing subcellular localisation of protein–protein interactions, for example. The SETD2(13»21) construct lacks the tripartite SET domain, proline-rich region and N-terminal RS-like domains, which may modulate subcellular localisation through specific protein–protein interactions. The RS domains are

required for appropriate subcellular localisation of SR proteins (Cazalla *et al.*, 2002). Alternatively domains present in SETD2(13»21) are sufficient, but simply do not fold properly in the absence of N-terminal sequence.

The differences in electrophoretic mobility and fractionation between p66^{C} and SETD2(13»21) could also be caused by the addition of an epitope-tag, which could potentially interfere with correct protein folding, protein–protein interactions, post-translational modifications, and subcellular localisation.

6.3.4. Implications for the identity of p66$^{C/CT21}$

Although these transfection experiments demonstrate that SETD2(13»21) constructs do not fully recapitulate properties of endogenous p66$^{C/CT21}$, the results do not refute the possibility that p66$^{C/CT21}$ is a SETD2 isoform. The co-fractionation and two-dimensional co-migration of two α-SETD2 signals still supports p66$^{C/CT21}$ as a putative SETD2 isoform. Given the anomalous migration of SETD2(13»21), sequence prediction of a C-terminal p66$^{C/CT21}$ isoform from alternative ORFs containing both C-terminal and CT21 epitopes becomes harder, since it is unclear what aspect of sequence or acquired endogenous behaviour confers such electrophoretic migration to transfected proteins.

One plausible hypothesis is that alternative transcripts exist that encode p66$^{C/CT21}$. The characterisation of alternative transcripts in Chapter 5 is non-exhaustive; indeed, recent RNA-sequencing data suggest a multitude of rare, alternatively spliced transcripts over the 3′ region of human *SETD2* (and thus potential to encode smaller C-terminal-specific proteins).

Although these experiments fairly conclusively indicate that the tagged proteins expressed from the transfected vectors do not behave comparably in most respects to endogenous p66$^{C/CT21}$, the experiments cannot absolutely refute the possibility that the 12–12b–13»21 cDNA (and indeed transcript(s) from which it was cloned) may be related to p66$^{C/CT21}$ generation. Indeed, ectopic, plasmid-based, epitope-tagged constructs do not fully recapitulate the original transcript, from which the (likely) partial cDNA derives—let alone

the process of transcription and RNA processing from the native gene. Eukaryotic protein expression is more complex than attempts to recapitulate an *in silico* translation of a cDNA. For example, 5′ and 3′ UTRs are critical in correct regulation of expression, subcellular localisation and co-translational modification of proteins: different mass isoforms of CCCTC-binding factor (CTCF) are produced from cDNAs that differ solely in 3′ UTRs (Klenova *et al.*, 1997); 3′ UTRs may even critically regulate post-translational modification of C/EBPβ (Basu *et al.*, 2011). Additionally, the nature of the endogenous promoter of a gene and the process of splicing (which are not recapitulated in the expression of the ectopic cDNA constructs) may influence gene expression through deposition of exon-junction complexes, which may regulate translation and NMD, or through the interaction of specific RNA-binding proteins that may influence post-transcriptional events, such as subcellular localisation, which would affect the biological function and extractability of a protein (Moore & Proudfoot, 2009; Shen *et al.*, 2010).

These transfection experiments do not diminish the assertion that p66$^{C/CT21}$ is a plausible, novel SETD2 isoform, based upon co-immunodetection, co-fractionation and two-dimensional co-migration under multiple conditions. However, critical independent evidence—most likely from further mass spectrometry—is required to more rigorously conclude on the true identity of p66$^{C/CT21}$.

These experiments also suggest that the way SETD2(13»21) protein folds *in vivo* is not a hindrance to detection with the C-terminal epitope. If p66$^{C/CT21}$ does represent a SETD2 isoform, then this indicates that the thermo-urea denaturation requirement of p66$^{C/CT21}$ must require additional criteria than the peptide sequence present in the transfected construct—perhaps post-translational modification or strong interactions with other proteins.

6.3.5. Implications for p290^{SETD2}

The detection and fractionation properties of SETD2(13»21) have implications for the nature of p290^{SETD2} and the H3K36 methyltransferase.

Firstly, the C-terminal protein-protein interaction motifs present in SETD2(13»21) are not sufficient to recapitulate the fractionation properties of p290^{SETD2}, suggesting that sequences in the N-terminus are likely required for this. This is rather significant, since all previously reported SETD2-protein interactions map to the C-terminal region contained in the 13»21 construct (with the exception of p53 interaction, which maps with the tripartite SET domain in addition).

Secondly, p290^{SETD2} is defined by recognition as a ≈290 kDa protein detected by N-terminal and C-terminal SETD2 antibodies, which is depleted by RNAi (**Figure 3.8**). However, the 287 kDa protein predicted from the conventional *Setd2* transcript model of 21 exons predicts that p290^{SETD2} should also contain the CT21 epitope. However, under conditions where p290^{C} and p290^{N} were eliminated by RNAi, no α-CT21-reactive bands were reduced—not even the p290^{CT21} protein (**Figure B.2**). This demonstrated that α-CT21 does not detect the p290^{SETD2} protein detected by α-N-/C-terminal antibodies. Note that, technically, these experiments do not refute that p290^{CT21} may represent a p290$^{N/C}$-distinct form of SETD2 (which is too stable to knockdown by RNAi; with N-/C-terminal epitopes being occluded or absent).

Nevertheless, α-CT21 antibody recognises CT21 epitope present in SETD2(13»21) (**Figure 6.2-v**). Why then does α-CT21 detect SETD2(13»21) but not p290$^{N/C}$? Several explanations are possible:

(i) p290$^{N/C}$ may contain the CT21 epitope, but endogenous PTMs present on p290$^{N/C}$, but not SETD2(13»21), prevent antibody binding. This is plausible given SETD2(13»21) did not fully recapitulate p290^{SETD2} fractionation (**Figures 6.2** and **6.3**)—thus differential modification could result from differential subcellular localisations and interactions.

(ii) p290$^{N/C}$ may not contain the CT21 peptide itself—the C-terminus of p290$^{N/C}$ but not SETD2(13»21) may be removed by proteolysis.

(iii) p290^{SETD2} derives from an alternative transcript where inclusion of exon 21 does not necessitate incorporation of peptide sequence; for example, a theoretical *Setd2* transcript that retains the intron between exons 20 and 21 (*e.g.* 1»19–20–intron–21) would escape NMD, and produce a ≈284 kDa protein including both N- and C-terminal epitopes, but prematurely truncated before the CT21 epitope, and yet still be susceptible to exon 21–specific siRNA. Partial cDNAs with intron retention events between exons 19–21 (covering the entire SRI domain) exist in mouse, rat and human. Indeed, identification of 17–18.5 cDNA in 10T½ cells (**Figure 5.8**) further highlights potential for encoding proteins lacking CT21 epitope.

Regardless of the exact reason, the data suggest conflict in understanding the final C-terminal residues of SETD2 (at least the final 16 residues), which comprise the C-terminal section of the SRI domain. NMR structures of the human SRI domain suggest that although CT21 sequence contributes towards hydrophobic packing of the core structure, it does not interact directly with phospho-CTD (Li *et al.*, 2006). Post-translational modification to this region, such as phosphorylation, however, may plausibly abrogate binding to the phospho-CTD through electrostatic repulsion. The possibility of broader disruption of the SRI domain from alternative transcripts (e.g. retention of the intron between exons 20 and 21) has implications for the mechanism of SETD2 function and K36me3 generation.

The assertions that SETD2 interacts with hyperphosphorylated RNAPII (whether this be directly or indirectly) and that the isolated SRI interacts with pCTD are still valid. However, what is questionable is whether mammalian SETD2 as the p290^{SETD2} methyltransferase interacts with RNAPII via a SRI domain, which may or may not even be a component of the protein. This is important because it has implications for the mechanism of SETD2 recruitment and activity (discussed in section 1.3.3). Recent reports conflict on the significance of the pCTD–SRI interaction for H3K36me3 targeting: one report suggests that pCTD interaction is crucial for splicing-coupled K36me3 generation (de Almeida *et al.*, 2011); whilst another report suggests that K36me3 can be deposited at

transcribed genes in a splicing-independent manner that does not correlate with the distribution of RNAPII (Kim *et al.*, 2011). Indeed, high resolution ChIP-seq demonstrates that K36me3 does not correlate with the distribution of phosphoserine-2 (de Almeida *et al.*, 2011). Indeed, both yeast and human SETD2 lacking an SRI domain can still interact (albeit less strongly) with hyperphosphorylated RNAPII (Phatnani *et al.*, 2004), and is targeted *in S. cerevisiae* like full-length protein (Youdell *et al.*, 2008). Furthermore, SRI-pCTD-independent recruitment to chromatin in mammals could be alternatively achieved through interactions demonstrated with elongation factor complex IWS1/SPT6 (Yoh *et al.*, 2008), and other chromatin-binding proteins (like p53 and Huntingtin) (Faber *et al.*, 1998; Xie *et al.*, 2008).

Because no cDNAs have actually been isolated that splice exons encoding KMT-related domains to the terminal exons encoding the SRI domain, it is plausible that p290^{SETD2} does not exist as a methyltransferase with an SRI domain. The only evidence available for the existence is the concatenation of ESTs to generate a gene model, which is consistent with—but not definitively validated against—an extrapolated model from yeast and other organisms. A crucial test of the role of the SRI domain *in vivo* in mammalian SETD2, which has never been formally investigated, would be to see whether expression of SETD2 lacking the SRI domain could recapitulate normal dynamics and distribution of H3K36me3 in *Setd2*$^{-/-}$ cells.

6.3.6. The nature of the endogenous 12−12b−13»21-containing transcript variant and the role of exon 12b in *Setd2* expression

The 12−12b−13»21 transcript definitively exists, but the actual protein that it encodes for *in vivo* is still unclear, since several ORFs exist within it. The SETD2(13»21) protein clearly does not represent p66$^{C/CT21}$ but may be related to the p90^{CT21} protein since they share similar fractionation in the final pellet after nuclease/salt extraction. However, endogenous p90^{CT21} protein is not detected by the C-terminal antibody. It is unlikely that SETD2(13»21) is related to the p85^{C} protein detected after two-dimensional analysis in **Figure 4.10**, since SETD2(13»21) did not distribute in the TD-S^{100} fraction.

Three other translation initiation sites are found in the 12−12b−13»21 transcript, which may produce proteins that reflect endogenous putative isoforms. If this particular SETD2 sequence exhibits 38% difference between predicted and observed mass, a 64 kDa protein is predicted from the ORF initiating from the first ATG codon in-frame with p290^{SETD2}; although this would more predict a more compatible molecular weight with p66^{CT21}, it is unknown which regions of the 13»21 region actually contribute towards anomalous migration. Furthermore, the predicted protein would lack the C-terminal epitope.

Perhaps exon 12b acts simply as a poison exon to trigger NMD. Exon 12b has characteristics similar to the exons found in AS−NMD regulation of splicing factors, such as SR proteins. Given SETD2 is also implicated in regulation of alternative splicing, perhaps a similar mechanism exists to regulate *Setd2* gene expression via AS−NMD coupling? Exon 12b is more likely to play an important, conserved in the biology of *Setd2*. Chapter 5 has already discussed the possibility that exon 12b may function as a non-coding element—unrelated to generation of protein isoforms—by NMD induction.

6.4. Conclusion

Although the SETD2(13»21) protein construct expressed from an ORF within the 12−12b»13−21 transcript variant does not represent p66$^{C/CT21}$, it displays specific subcellular fractionation and electrophoretic characteristics that may be representative of a unique C-terminal-specific SETD2 isoform derived from the variant transcript *in vivo*. Characterisation of SETD2(13»21) suggests N-terminal domains crucially influence p290^{SETD2} behaviour *in vivo* and also indicates that the CT21 epitope cannot be recognised in p290^{SETD2}.

Chapter 7 Final discussion and future directions

7.1. Identification of putative SETD2 isoforms

This thesis aimed to identify and characterise endogenous transcriptional and protein isoforms of mammalian SETD2 within murine C3H 10T½ fibroblasts.

During this study, affinity-purified antibodies were generated and specifically detected multiple putative isoforms of SETD2. The $p290^{SETD2}$ H3K36-trimethyltransferase was detected by α-N-terminal and α-C-terminal antibodies and was verified by RNAi (**Figure 3.8**). The N-terminal-specific $p120^{N}$ and major C-terminal-specific proteins, $p140^{C}$ and $p66^{C}$, were not depleted by transient siRNA transfection, which is likely due to their relatively slow turnover *in vivo* (**Figure 3.9**). Their generation from alternative transcripts not targeted by the siRNAs is also plausible. RNAi further indicates that these putative isoforms are unlikely to derive from biological or artefactual degradation of $p290^{SETD2}$.

Although multiple forms of SETD2 have been described in the literature, it is difficult to make rigorous comparisons because of limited information on antibody specificity, accurate size information, and comparisons in different cell types. In this work, only $p290^{SETD2}$ was observed migrating above 250 kDa, unlike reports of multiple proteins greater than 290 kDa (**Table 1.4**). A ≈130 kDa isoform identified by mass spectrometry may possibly be related to $p120^{N}$ or $p140^{C}$. In general, immunoforms detected in C3H 10T½ were also observed in human MRC5 and 293T cells, suggesting evolutionary conservation (**Figure 3.7**).

$p66^{C/CT21}$ is the most convincing isoform candidate, being the only protein to be detected by both α-C-terminal and α-CT21 antibodies, and which co-distributed in specific subcellular fractions derived from two different methods of sequential extraction (**Figures 4.7 and B.1**). Furthermore, both antibody signals co-migrated by two-dimensional gel electrophoresis (**Figure 4.9**). Further co-immunoprecipitation experiments will be required, however, to formally evaluate whether the two antibodies recognise the same polypeptide.

All of these smaller immunoforms require further verification by an independent technique—ideally direct identification by mass spectrometry or N-terminal sequencing. A fractionation procedure was successfully developed to enrich $p66^{C/CT21}$ for proteomic identification (**Figure 4.6**). Although MS identified CKAP4 as the predominant protein overlapping $p66^{C/CT21}$ (**Figure 4.11**), the intensity characteristics and identification of other bands within the fraction (along with inconsistencies from the literature) suggest $p66^{C/CT21}$ is not CKAP4. It is noted that the major staining spot at this location is not likely to represent SETD2, which would be present at much lower levels in the cell. The $p66^{C/CT21}$-enriched fraction provides an ideal starting point for future biochemical purification by chromatography or immunoprecipitation for downstream proteomic identification.

Other approaches for identification could include: (i) immunoblot analysis in *Setd2*$^{-/-}$ knock-out cells; (ii) longer duration of RNAi by stable transfection of shRNA-expressing plasmids; (iii) epitope-tagging a BAC-cloned or genomic copy of the *Setd2* gene; (iv) RNAi–*in vitro* translation: using antibodies to detect SETD2-related products derived from *in vitro* translation of RNA extracted from mouse fibroblasts that have been transfected with siRNAs to *Setd2*; and (v) for $p66^{C}$ specifically, screening cDNA expression libraries for proteins immunodetected by α-C-terminal antibody only after thermo-urea denaturation.

7.2. Verified alternative transcripts indicate potential protein diversity

The concurrent existence of multiple, mutually exclusive, alternatively spliced transcript variants was demonstrated by RT−PCR and confirmed by DNA-sequencing (**Figure 5.10**). Whilst their existence is definitive, their functional significance and contribution towards protein diversity can only be discussed speculatively at present, without knowledge of complete transcript structure. The identification of multiple, distinct events involving exon 12b predicts the potential existence of proteins defined by C-terminal truncation or novel translation initiation. Retained-intron transcripts may generate truncated forms, lacking C-terminal domains; for example, the 8»9.5 and 6»10.5 cDNAs may putatively produce KMTs without domains for interaction with RNAPII, transcription factors and Huntingtin. Transcript variants containing premature termination codons (6»10.5, 8–9.5

and 11−12−12b) may generate $p120^{N}$. Transcript variant 17−18.5 may contribute to the production of $p120^{N}$ or $p140^{C}$ (**Figure 5.10**).

In all cases, the partial cDNAs obtained here will need to be further characterised by mapping their 5' and 3' ends by RACE, primer extension or nuclease mapping. Ultimately, full-length cDNA cloning will be essential to understand their coding potential. Further characterisation of transcript composition will facilitate future design of ectopic expression constructs, and design of appropriate siRNAs to target specific variants or multiple transcripts in parallel. Retained-intron transcripts may reflect *bona fide* transcript variants or fortuitous amplification of pre-mRNA, which could be further assessed by analysing their presence in cytoplasmic or polysome-associated RNA, and by northern blotting using intronic probes.

Truncated proteins predicted from alternative splicing contain isoform-specific peptide sequence contributed from retained introns or exon 12b. Antibodies could therefore be generated against these peptides to assess whether transcripts encode functional proteins, and whether these proteins exhibit characteristics similar to the putative SETD2 isoforms detected by the three antibodies employed in this thesis. The development of fractionation procedures and characterisation of SETD2 immunoforms as established here provides further criteria to rigorously evaluate the interpretations of such experiments.

7.3. Functional significance of SETD2 diversity

7.3.1. General comments on putative isoforms

The $p140^{C}$, $p120^{N}$ and $p130^{CT21}$ isoforms remain unclear as *bona fide* SETD2 isoforms, and without knowledge of protein structure, little can be said of potential function. Once verified, isoforms could be further characterised through analysis of subcellular localisation by immunofluorescence, interactions by co-immunoprecipitation and pull-down assays, protein-specific overexpression, and chromatin interactions. Current evidence for SETD2 interaction with other cellular proteins does not allow rigorous assessment of whether Huntingtin, p53, hnRNP-L, RNAPII and Ataxin-1 interact with

p290^{SETD2} or other isoforms *in vivo*. This could be tested by co-immunoprecipitation and pull-down assays. It would also be interesting to assess where these SETD2-interactors distribute compared to p290^{SETD2} and p66$^{C/CT21}$ after subcellular fractionation.

Loss of virtually all H3K36me3 is associated with depletion of p290^{SETD2}. If these other putative SETD2 isoforms were methyltransferases, it is unlikely that they are thus involved in generation of H3K36me3—they are definitely not sufficient for global H3K36me3, but a requirement cannot be ruled out (Edmunds *et al.*, 2008; Yoh *et al.*, 2008). Since SETD2 *in vitro* appears to generate both mono- and di-methylated H3K36 (in addition to the major trimethylated product), there may still be a role for alternative isoforms in generating a proportion of H3K36me1/2. Furthermore, given many methyltransferases act on non-histone substrates and SETD2 exhibited auto-methylation activity (Sun *et al.*, 2005), the generation of a KMT isoform (lacking chromatin-targeting domains, for instance) may be significant for methylation of putative non-histone substrates. Once further isolated by biochemical purification, potential enzymatic activity could be tested *in vitro* with methylation assays.

7.3.2. The SETD2(13»21) variant: C-terminal interaction domains independent of KMT activity ?

The complete sequence of a novel, mouse brain cDNA was identified, and a derived construct was successfully expressed and characterised in mouse fibroblasts, revealing a novel isoform with distinct electrophoretic and fractionation properties. The recombinant construct, however, may not appropriately reflect endogenous translation initiation from this cDNA, so future work should aim to assess this by appending an epitope tag to the 3′ terminus of the original cDNA. Nevertheless, the existence of the cDNA suggests generation of a C-terminal-specific isoform that would lack KMT activity but would contain domains important for interaction with hyperphosphorylated RNAPII, DNA and Huntingtin (Faber *et al.*, 1998; Rega *et al.*, 2001; Sun *et al.*, 2005). The constructs already designed may be useful for further cloning to assess domain requirements for fractionation, interactions with RNAPII, Huntingtin, p53 and other interactors (**Figure 1.2B**).

7.3.3. Implications for p290^{SETD2} structure: the SRI domain and mechanism of H3K36me3 generation

The lack of p290^{SETD2} detection by α-CT21 antibody suggests that the final C-terminal residues are modified or absent *in vivo*. Since this epitope comprises part of the SRI domain, it would be interesting to further analyse whether p290^{SETD2} actually contains an intact SRI domain or whether it may be modified by post-translational modifications, which may confer α-CT21 occlusion. RT-PCR could be used to investigate potential transcriptional variation within exons 19 to 21, which encode the SRI domain. If indeed p290^{SETD2} lacks a functional SRI domain, this calls into question its true significance in SETD2 recruitment and activity. It would also be of further interest to analyse whether p290^{SETD2} can actually interact with RNAPII, in its various phosphorylated states, *in vivo* in the presence or absence of the SRI domain and the effect this would have on H3K36me3 generation.

7.3.4. p66$^{C/CT21}$ as a putative C-terminal SETD2 isoform

p66$^{C/CT21}$ has been identified as a putative, novel SETD2 isoform that appears to contain two C-terminal epitopes. It does not result from proteolytic processing of p290^{SETD2} nor from translation of an ORF spanning exons 13»21. It may represent an isoform with C-terminal interaction domains uncoupled from KMT activity. Its incorporation into complexed material is interesting and further characterisation of the proteins and complexes that co-extract with p66$^{C/CT21}$ may provide a clue to its function. Indeed, what is the molecular basis of the complexed form of p66$^{C/CT21}$? How large is it? Does p66$^{C/CT21}$ interact with other proteins or is the complex homomeric? Is the complex physiological or an artefactual result of extraction conditions? The DDC fractionation procedure developed here provides a foundation on which to attempt further biochemical purification of p66$^{C/CT21}$ and its complex, which could be further isolated and characterised by density gradient centrifugation, gel filtration or blue-native PAGE.

What is the basis for the resistance of p66$^{C/CT21}$ to extraction by nucleases and high salt? Does it interact with the cytoskeleton, or perhaps the nuclear matrix? Intriguingly, both

association with nuclear matrix and detergent-soluble vesicle complexes are characteristics shared with Huntingtin as a regulator of transcription (Kegel *et al.*, 2002; Zuccato *et al.*, 2003; Sayer *et al.*, 2005; Benn *et al.*, 2008) and vesicular trafficking (DiFiglia *et al.*, 1995; Hoffner *et al.*, 2002; Caviston *et al.*, 2007), respectively.

As far as we are aware, little precedent exists for immunodetection of a protein specifically after thermal-urea denaturation. Indeed, irrespective of the true identity of $p66^{C}$—SETD2 or otherwise—the biochemical properties of this protein highlight a gap in understanding the effect of denaturants on protein structure, and are sufficiently novel and interesting to warrant further investigation—especially given involvement of protein aggregates in neurodegenerative disease and the association of SETD2 with HD and SCA1 (Passani *et al.*, 2000; Lim *et al.*, 2006). Any functional significance and biological relevance, however, remains critically dependent on identifying this protein. Structural and biophysical techniques could be used to rationalize the immunodetection requirements for $p66^{C}$.

If $p66^{C/CT21}$ corresponds to SETD2, it would also contain the putative transcriptional activation domain/DNA-binding region (Rega *et al.*, 2001; Sun *et al.*, 2005). It would be interesting to see if $p66^{C/CT21}$ associates with particular regions of the genome by ChIP-seq, which may provide clues as to its function.

7.3.5. What regulates alternative splicing of *Setd2*? Is splicing altered during development or disease?

Given SETD2 regulates RNA processing through H3K36me3, does SETD2 regulate alternative splicing of its own gene? This could be investigated by RT-PCR analysis and H3K36me3 ChIP after siRNA-mediated knockdown of SETD2.

What is the significance of exon 12b inclusion? Does it contribute towards generation of non-coding transcripts or distinct N- or C-terminal-specific isoforms? Given its striking conservation between vertebrates and similarity to poison exons that mediate regulation of SR protein expression by AS–NMD coupling, could a similar expression mechanism act to regulate synthesis of the H3K36 trimethyltransferase? Preliminary RT-PCR experiments

suggest exon 12b–containing transcripts are stabilised by emetine (data not shown); a putative role for NMD here could be investigated by treatment with NMD inhibitors (such as caffeine) or by RNAi of *SMG1*, an essential NMD factor (McIlwain *et al.*, 2010). SR gene–like AS–NMD coupling appears plausible for SETD2 as a protein that is implicated in regulation of RNA splicing, and like many splicing regulators, contains an RS-like domain. Furthermore, this study has revealed moderate lability of $p290^{SETD2}$ (**Figure 3.9**), suggesting that protein levels may be effectively influenced by variation of transcript abundance and composition.

Because of its critical effect on $p290^{SETD2}$ generation, it will also be important to understand what controls exon 12b utilisation: is its inclusion a stochastic event or is it regulated by defined factors or/and in response to specific stimuli? Aberrant exon 12b inclusion would be expected to disrupt $p290^{SETD2}$ expression and H3K36me3. Given loss of *SETD2* function is associated with ccRCC (Dalgliesh *et al.*, 2010; Duns *et al.*, 2010), it would be interesting to analyse the frequency of exon 12b inclusion in cancer cell lines, to ascertain whether dysregulation of exon 12b splicing may also contribute towards disruption of SETD2 function in disease states.

7.3.6. Significance for SETD2 function in chromatin modification, tumourigenesis and Huntington's disease

The results presented here in conjunction with rudimentary precedents in the literature suggest generation of SETD2 isoforms. Previous characterisation of truncated SETD2 constructs indicated that specific regions have discrete activities that may be affected by the presence of adjacent domains (Sun *et al.*, 2005).

The majority of *SETD2* mutations identified in ccRCC are predicted to insert premature termination codons into the reading frame encoding $p290^{SETD2}$ (Dalgliesh *et al.*, 2010; Duns *et al.*, 2010). The existence of transcripts in 10T½ cells as well as ESTs in human, mouse and rat that have been isolated despite being predicted targets of NMD raises the question of whether a specific mechanism acts on *SETD2* transcripts to allow them to escape NMD.

Although ccRCC is associated with a loss of H3K36me3 (Duns *et al.*, 2010), the impact of *SETD2* mutations on transcript structure and protein synthesis is unclear. Are truncated SETD2 proteins still produced in cells with mutations predicted to cause NMD? Is the putative tumour suppressor function of *SETD2* due to loss of H3K36me3 or some other aspect? Given the involvement of SETD2 in the biology of p53, Huntingtin, splicing and transcriptional regulation, could expression of a truncated SETD2 protein contribute towards additional dysregulation that may occur independently of H3K36me3 generation? If the *SETD2* gene concomitantly produces $p290^{SETD2}$ along with additional C-terminal proteins, what would be the effect of these additional proteins on cellular processes in the absence of $p290^{SETD2}$?

There is an emerging nuclear role for Huntingtin in transcriptional regulation and it has recently been identified as a regulator of PRC2 and H3K27me3—a chromatin modification that is impeded by H3K36me2/3 (Benn *et al.*, 2008; Seong *et al.*, 2010). Given the interaction of Huntingtin with C-terminal fragments of SETD2, it will be important to understand whether Huntingtin interacts with the full-length methyltransferase $p290^{SETD2}$, or whether its endogenous interaction occurs only with a C-terminal-specific variant—such as SETD2(13»21) or $p66^{C/CT21}$. Since these proteins are expected to lack methyltransferase activity, differential interaction with isoforms would have implications for interpreting whether or not Huntington may influence H3K36 methylation—and thus potential cross-talk with H3K27me3 regulation—within the important context of transcriptional regulation.

7.4. Conclusion

This thesis highlights the complexity of *Setd2* expression, definitively in terms of transcript variants, and putatively in terms of potential protein isoforms. The existence of alternative forms would have significant implications for the proposed functions and interactions of SETD2 in the regulation of gene expression and its contribution to Huntington's disease and cancer development.

Appendix A Supplementary results for Chapter 3:

Immunodetection of p290^{SETD2} and putative SETD2 isoforms

A-1 Estimation of apparent molecular masses of immunoreactive proteins

Apparent molecular masses were estimated by linear regression of the relationship between logarithm of molecular mass and relative electrophoretic mobility (R_f) by SDS-PAGE (Weber & Osborn, 1969). Standard curves were determined using molecular weight markers. Independent measurements were made for at least five independent biological experiments, and the mean and standard deviation calculated (**Table 7.1**).

Table 7.1 – Estimation of molecular mass of α-SETD2 immunoforms

Protein	Mean molecular mass (kDa)	Standard deviation (kDa)	*n*
p140^{C}	140	5	12
p120^{N}	120	2	5
p66^{C}	66	3	15
p45^{C}	45	1	6
p66^{CT21}	67	2	5

Given that at higher molecular mass, the log(m)–R_f relationship becomes hyperbolic, and less reliable to estimate accurately (Weber & Osborn, 1969), the p120^{N} and p140^{C} proteins have been estimated to two significant figures. Furthermore, the molecular mass of p290^{SETD2} cannot be reliably extrapolated using this technique; '290' refers to the estimate based on gene predictions, but given that the exact identity of sequence present in the p290^{SETD2} protein is unknown, an exact molecular mass is not specified.

A-2 Characterisation of minimal conditions for $p66^{C}$ detection

A-2.1. Thermo-urea denaturation causes loss of $p120^{N}$ detection

Immunodetection of $p120^{N}$ is impaired by heating at 95°C in the presence of urea (**Figure A.1**) in a manner seemingly reciprocal to the appearance of $p66^{C}$ (*cf.* **Figure 3.6B**). The disappearance of $p120^{N}$ is not coupled to the appearance of $p66^{C}$, however, since the majority of each protein can be segregated into independent subcellular fractions (*cf.* **Figure 4.2B-ii/iii**).

A-2.2. Immunodetection of $p66^{C}$ requires concomitant heating with urea

Optimal immunodetection of $p66^{C}$ only occurs in extracts solubilised in the presence of 4 M urea that have been thermally denatured at 95°C (**Figure A.2-i**, lane 1 versus 2). Since this condition causes artefactual carbamylation which would hinder downstream proteomic analysis, it was of interest to ascertain whether $p66^{C}$ could be detected by heating extracts first at 95°C in SDS-PAGE loading buffer, and then cooling to room temperature, before addition of 4 M urea. Under such a condition, $p66^{C}$ was not observed (**Figure A.2-i**, lane 4)—the extract required subsequent heating at 95°C once the extract already contained urea (**Figure A.2-i**, lane 6). Thus $p66^{C}$ detection by the α-SETD2 C-terminal antibody appears to require concomitant thermal denaturation in the presence of urea.

Figure A.1 Loss of p120^{N} when extracts are heated in urea

TX–DOC extracts from quiesced 10T½ cells were solubilised in 1× SDS-PAGE loading buffer with the indicated concentration of urea (0–4 M) and either incubated on ice (lane 1) or 70°C (lanes 2–5) or 95°C for (lanes 6–9) for the indicated times. Samples were separated by 8% SDS-PAGE and transferred to PVDF in Towbin buffer. Membranes were probed with α-SETD2 N-terminal antibody.

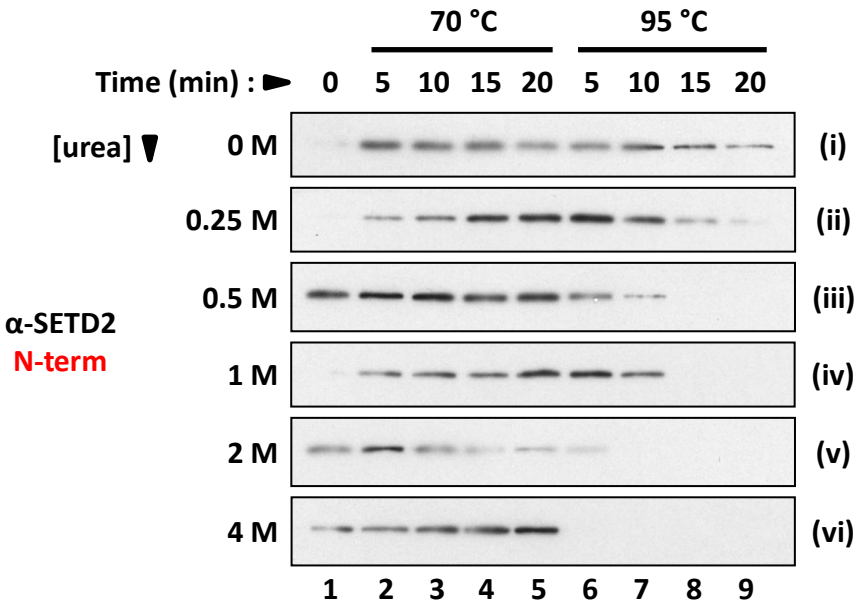
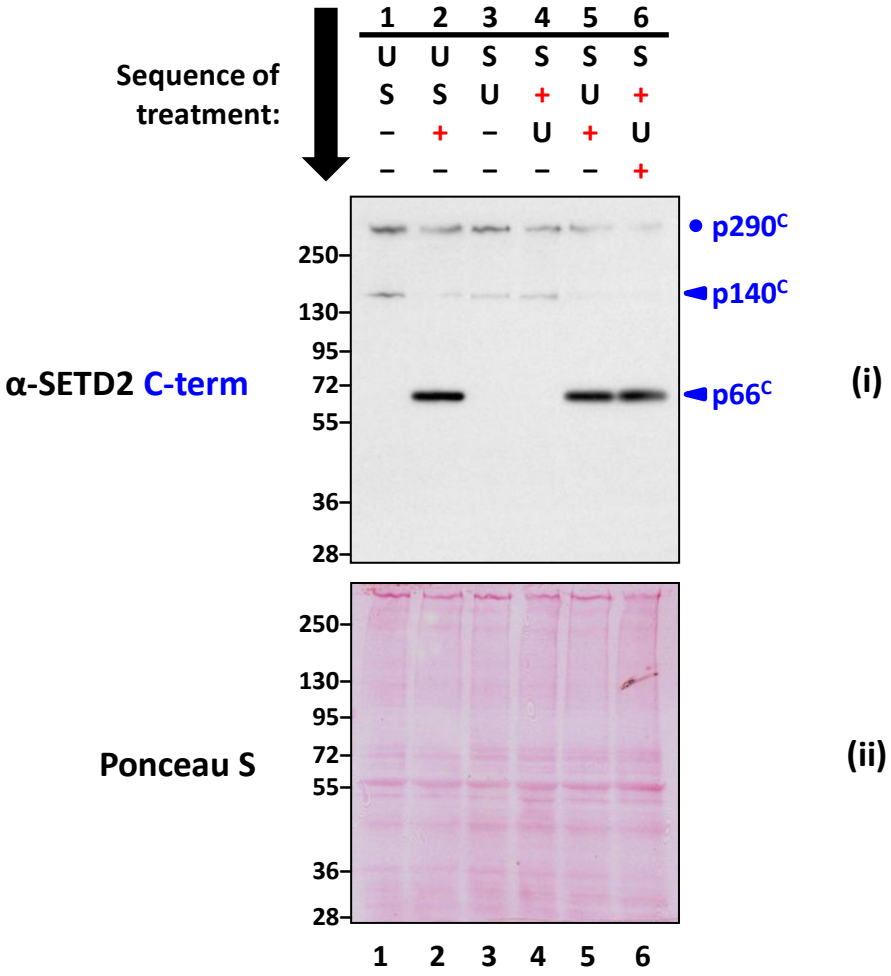

70 °C
95 °C
Time (min) :
0
5
10
15
20
5
10
15
20
[urea]
0 M
0.25 M
0.5 M
1 M
2 M
4 M
(i)
(ii)
(iii)
(iv)
(v)
(vi)
α-SETD2
N-term
1
2
3
4
5
6
7
8
9

Figure A.2 p66^{C} detection requires concomitant thermo-urea denaturation

Immunodetection of p66^{C} is influenced by varying the order of sample denaturation with respect to adjustment to 4 M urea (U), 1× SDS-PAGE loading buffer (S), and denaturation at 95°C for 15 minutes (+). Incubation at room temperature for 15 minutes is indicated by (−). Samples represent identical aliquots of high-salt extract (prepared according to **Figure 4.1**) from quiesced 10T½ cells. Samples were resolved by 8% SDS-PAGE and transferred to PVDF in SDS transfer buffer. After staining with Ponceau S to ensure equal loading **(ii)**, membranes were probed with α-SETD2 C-terminal antibody **(i)**.

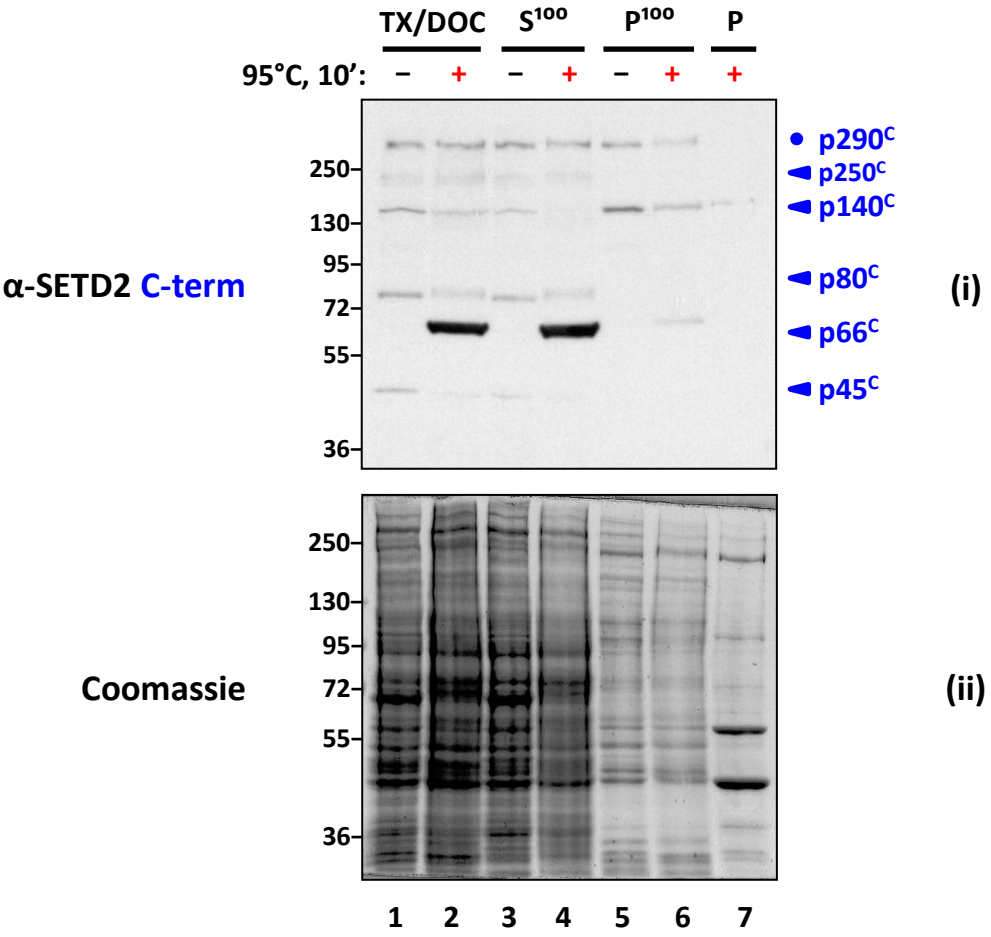

Sequence of treatment:
1 2 3 4 5 6
U U S S S S
S S U + U +
– + – U + U
– – – – – +
α-SETD2 C-term
p290C
p140C
p66C
250
130
95
72
55
36
28
(i)
Ponceau S
(ii)
1 2 3 4 5 6

A-3 Distribution of p66^{C} after ultracentrifugation of TX−DOC extracts

The inability to detect p66^{C} in the absence of thermo-urea denaturation may result from its incorporation into a protein complex that does not enter the resolving gel during SDS-PAGE (even after the relatively harsh conditions of 4 M urea, 2% (*w*/*v*) SDS, 5% (*v*/*v*) β-mercaptoethanol and denaturation at 70°C for 20 minutes – *cf.* **Figure 3.6B**). Large complexes, such as ribosomes, sediment in TX−DOC buffer after ultracentrifugation at 100,000*g* for 1 hour (Mahadevan *et al.*, 1989). Under these conditions, virtually all p66^{C} remains soluble in the ultracentrifugation supernatant S^{100} (**Figure A.3**, lane 4), indicating that under these buffer and centrifugation conditions, the majority of p66^{C} does not exist in a highly complexed form. Similarly, the p250^{C}, p80^{C} and p45^{C} immunoforms remain soluble (**Figure A.3**, lanes 3 and 4) whereas p290^{SETD2} and p140^{C} are divided between complexed and non-complexed material (**Figure A.3**, lane 3 versus 5). This result indicates that the thermo-urea denaturation requirement for p66^{C} immunodetection is independent of sedimentation properties – and thus unrelated to the behaviour of the [URE3] prion in budding yeast (*cf.* section 3.4.5.2). This result does not exclude the possibility that thermo-urea denaturation is required to remove a protein that strongly interacts with the α-C-terminal-reactive epitope within p66^{C}, which may together represent complexed material that is insufficiently large to sediment under the aforementioned centrifugation conditions. Furthermore, it should be noted that, in alternative extracts, virtually all p66^{C} does in fact sediment under the centriguation conditions employed above (*cf.* **Figures 4.3**, **4.4B** and **B.1**). Sodium deoxycholate appears to be a critical component that converts p66^{C} into a non-complexed form (*cf.* **Figure 4.4C**).

Figure A.3 p66^{C} is virtually absent from complexed material in TX–DOC extracts

Soluble TX–DOC extract (lanes 1 and 2) and TX-DOC-resistant pellet (lane 7) were prepared from quiesced 10T½ cells. A proportion of TX–DOC supernatant was centrifuged at 100,000g for 1 hour at 4°C to yield 'non-complexed' S^{100} supernatant (lanes 3 and 4) and 'complexed' P^{100} pellet fractions (lanes 5 and 6). Soluble fractions were precipitated with TCA. All samples in pellet form were solubilised in urea/SDS-PAGE loading buffer and then incubated at room temperature (odd lanes) or 95°C (even lanes) for 10 minutes. Equivalent proportions of each fraction were separated by 8% SDS-PAGE. Gels were either Coomassie-stained **(ii)** or transferred to PVDF in SDS transfer buffer for immunoblotting with α-SETD2 C-terminal antibody **(i)**.


TX/DOC
S100
P100
P
95°C, 10’:
– + – + – + +
250
130
95
72
55
36
p290C
p250C
p140C
p80C
p66C
p45C
α-SETD2 C-term
(i)
Coomassie
(ii)
1 2 3 4 5 6 7

A-4 Effect of emetine treatment on SETD2 immunoforms

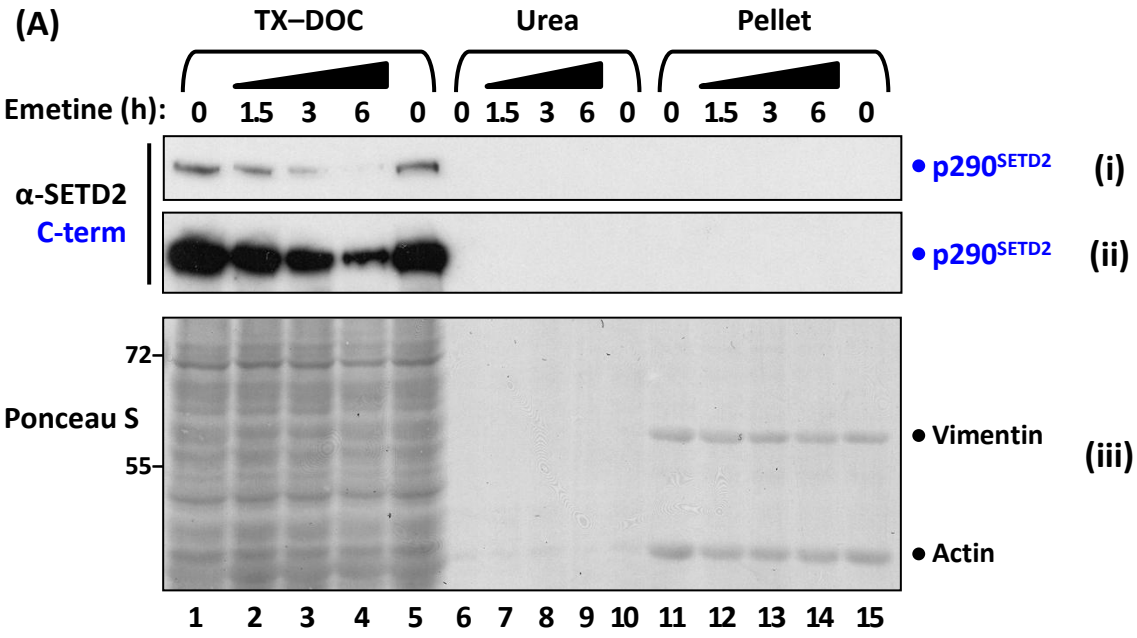
The decrease in $p290^{SETD2}$ in the TX−DOC extract following emetine treatment (**Figure A.4**, lanes 1−5) is not due to redistribution into TX-DOC-resistant fractions (**Figure A.4**, lanes 6−15). The figure also provides quantitative presentation of the effect of emetine on $p290^{SETD2}$, $p120^{N}$, $p150^{C}$ and $p66^{C}$ proteins, averaged from at least three independent experiments (**Figure A.4B**). The loss of EGF-inducible synthesis of c-JUN protein after emetine treatment is also demonstrated as a positive control for inhibition of translation by emetine (**Figure A.4C**).

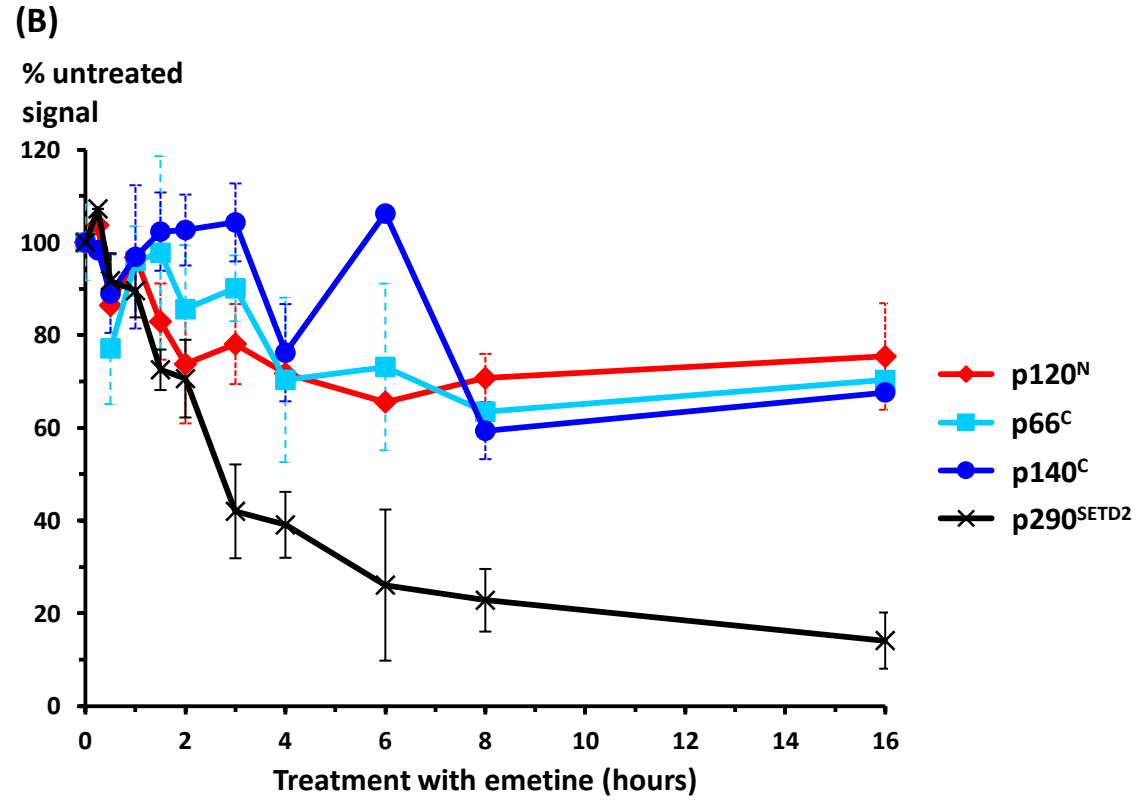
Figure A.4 Effect of emetine on p290^{SETD2}

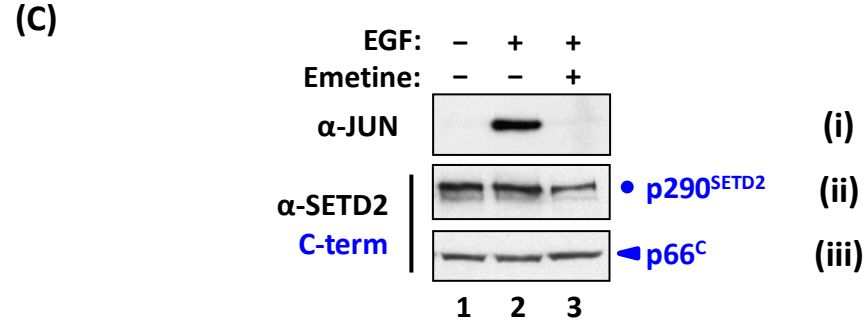
(A) Emetine does not cause redistribution of p290^{SETD2} into TX–DOC insoluble fractions: Quiesced 10T½ cells were treated with 10 μg/ml emetine for the indicated durations before harvesting in TX–DOC buffer to yield TX–DOC supernatant (lanes 1–5) and sequentially extracting TX-DOC-resistant material with urea-solubilisation buffer (lanes 6–10) and urea/SDS-PAGE loading buffer (lanes 11–15). TX–DOC extracts were precipitated with TCA and all samples adjusted to urea/SDS-PAGE loading buffer at room temperature. Equivalent proportions of each fraction were separated by 8% SDS-PAGE and then transferred to PVDF in SDS buffer. Membranes were stained with Ponceau S **(iii)** and probed with α-SETD2 C-terminal **(i, ii)**. Panel (ii) represents a longer exposure of panel (i).

(B) Quantified summary of the effect of emetine treatment on α-SETD2 immunoforms: immunoblot signals were quantified using ImageJ from a minimum of three independent experiments, presented as mean percentage signal compared to untreated control. Error bars represent standard error of the mean.

(C) Emetine treatment inhibits EGF-mediated JUN synthesis. Quiesced 10T½ cells were left untreated (lane 1), treated with 50 ng/ml EGF for 30 minutes (lane 2) or treated with 10 μg/ml emetine for 15 minutes followed by 50 ng/ml EGF for 30 minutes (lane 3). TX–DOC extracts were prepared, precipitated with TCA and resuspended in urea/SDS-PAGE loading buffer and incubated at room temperature **(i, ii)** or at 95°C for 10 minutes **(iii)**. Samples were separated by 8% SDS-PAGE and then immunoblotted using α-JUN **(i)** or α-SETD2 C-terminal **(ii, iii)** antibodies.


(A)
TX–DOC
Urea
Pellet
Emetine (h): 0 1.5 3 6 0 0 1.5 3 6 0 0 1.5 3 6 0
α-SETD2
C-term
p290SETD2
(i)
p290SETD2
(ii)
72
Ponceau S
55
Vimentin
Actin
(iii)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15


(B)
% untreated signal
120
100
80
60
40
20
0
0 2 4 6 8 10 12 14 16
Treatment with emetine (hours)
p120N
p66C
p140C
p290SETD2


(C)
EGF: – + +
Emetine: – – +
α-JUN
(i)
α-SETD2
C-term
p290SETD2
(ii)
p66C
(iii)
1 2 3

Appendix B Supplementary results for Chapter 4:

Generation of a p66-enriched fraction

B-1 $p66^{C}$ and $p66^{CT21}$ co-fractionate after extensive extraction of nuclease-resistant material

Since the majority of $p66^{C}$ sediments by ultracentrifugation of the high-salt extract after nuclease–salt fractionation (*cf.* **Figure 4.3B**) and virtually all $p66^{C}$ can be extracted from crude nuclei by Triton-extraction buffer (*cf.* **Figure 4.4A**), it was of interest to investigate whether these fractionation procedures could be combined to derive an extract further enriched with $p66^{C}$. In overview, this involved the DDC extraction of the material that sediments after centrifugation of the entire nuclease-resistant pellet solubilised in high-salt buffer. Centrifugation of this salt resuspension was performed at either 16,000*g* for 10 minutes or 100,000*g* for 60 minutes, and the resulting pellets extracted by the DDC method (**Figure B.1**). Consistent with previous results, centrifugation of the salt resuspension at 16,000*g* produced a supernatant Salt-S^{16} (**Figure B.1-i,** lane 3), which contained a smaller proportion of total $p66^{C}$ than the pellet (represented by the summation of immunoblot signal in lanes 4–7). After centrifugation of the salt resuspension at 100,000*g*, $p66^{C}$ was virtually absent from the supernatant Salt-S^{100} (**Figure B.1-i,** lane 8) and fractionated instead in the pellet (represented by the summation of immunoblot signal in lanes 9–12). Since the distribution of $p66^{CT21}$ mirrors that of $p66^{C}$ (**Figure B.1-ii**), $p66^{CT21}$ also shares the extraction and sedimentation properties of $p66^{C}$—namely, resistance to extraction by nuclease digestion and high-salt, and incorporation into complexed material within the high-salt extract.

The pellet remaining after centrifugation at 100,000*g* (summation of lanes 9–12) represents a single fraction that contains virtually all $p66^{C/CT21}$ (along with components of the cytoskeleton and lamins). Considering Triton-extraction is sufficient to extract virtually all $p66^{C/CT21}$ from the aforementioned contaminants in the standard DDC protocol, it was of interest to see whether this would also occur from this single $p66^{C/CT21}$-enriched fraction. Two interesting observations derive from this analysis. Firstly, the Triton-extraction buffer

is insufficient to extract $p66^{C/CT21}$ from material that resists extraction by nuclease digestion and high-salt (**Figure B.1,** lanes 7 and 12). Secondly, the proportion of $p66^{C/CT21}$ that is solubilised with Triton-extraction does not sediment after ultracentrifugation (**Figure B.1,** lanes 4 and 9), suggesting that the complexed forms of $p66^{C/CT21}$ are retained in the final pellet. Overall, the co-fractionation of $p66^{C}$ and $p66^{CT21}$ throughout these extraction procedures further supports their equivalence.

B-2 α-SETD2 CT21 antibody does not recognise $p290^{SETD2}$

RNAi analysis of TX–DOC extracts from murine fibroblasts using α-SETD2 CT21 antibody showed no loss of any α-CT21 immunoform (**Figure B.2-iii**, lanes 1 and 2), despite depletion of $p290^{SETD2}$ detected by N-terminal and C-terminal α-SETD2 antibodies (**Figure B.2-i/ii,** lanes 1 and 2).

Figure B.1 Co-fractionation of p66^{C} and p66^{CT21} in extracts derived from high-salt and detergent solubilisation of nuclease-resistant material

Samples from quiesced 10T½ cells were prepared according to a method combining nuclease–salt fractionation and DDC. The former protocol was followed to yield the hypotonic supernatant (lane 1) and nuclease-digest fraction (lane 2). The nuclease-resistant pellets from two replicates were individually resolubilised in high-salt buffer and then processed along parallel lines that differed solely in the initial centrifugation speed: (i) **16,000*g* route** (lanes 3–7)**:** the resuspension was centrifuged at 16,000*g* for 10 minutes to yield a high-salt supernatant Salt-S^{16} (lane 3) and a salt-resistant pellet; or (ii) **100,000*g* route** (lanes 8–12): the resuspension was centrifuged at 100,000*g* for 60 minutes to produce a high-salt supernatant Salt-S^{100} (lane 8) and a salt-resistant pellet. For both extraction routes, salt-resistant pellets were subsequently processed according to the DDC protocol from the point of resolubilisation with Triton-extraction buffer to generate TX-S^{100}, TD-S^{100} and TD-P^{100} fractions. The DDC fractions representing Triton-resistant pellets (TX-P in DDC) are denoted as P^{16} (lane 7) and P^{100} (lane 12); they represent material resistant to extraction with nucleases, high-salt and Triton-extraction buffer, from the 16,000*g* and 100,000*g* routes, respectively. Soluble extracts were precipitated with TCA and then all extracts solubilised in urea/SDS-PAGE loading buffer at room temperature **(ii)** or at 95°C for 10 minutes **(i)**. Equivalent proportions of each fraction were separated by 8% SDS-PAGE and then transferred to PVDF in SDS buffer. Membranes were stained with Ponceau S **(iii)** prior to immunoblotting with α-SETD2 C-terminal **(i)** or CT21 **(ii)** antibodies.(Carninci *et al.*, 2005)

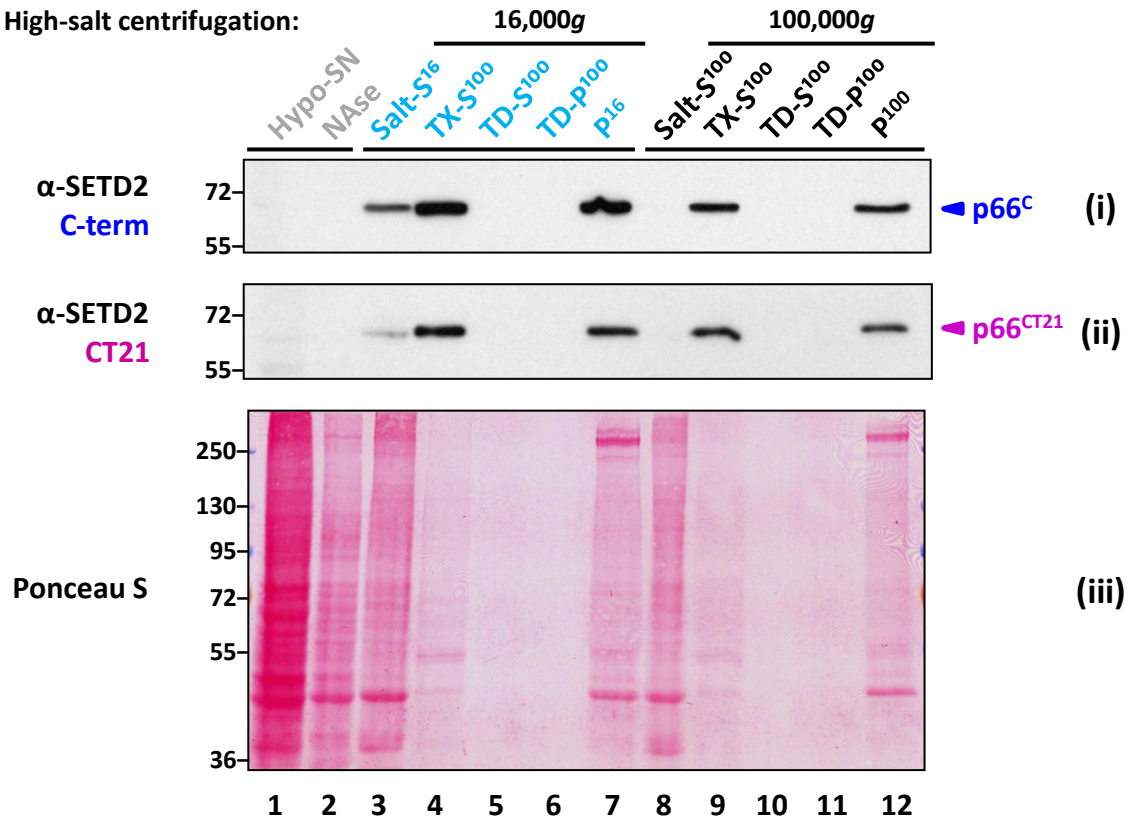
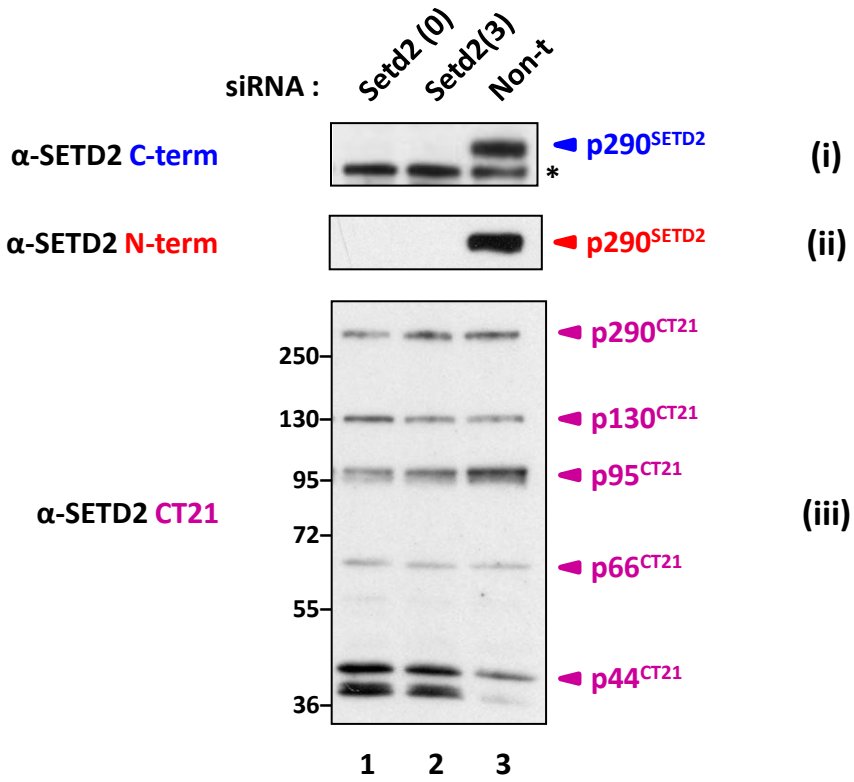

High-salt centrifugation:
16,000g
100,000g
Hypo-SN
NAse
Salt-S16
TX-S100
TD-S100
TD-P100
P16
Salt-S100
TX-S100
TD-S100
TD-P100
P100
α-SETD2
C-term
72
55
p66C
(i)
α-SETD2
CT21
72
55
p66CT21
(ii)
Ponceau S
250
130
95
72
55
36
(iii)
1 2 3 4 5 6 7 8 9 10 11 12

Figure B.2 α-SETD2 CT21 antibody does not recognise p290^{SETD2}

10T½ cells were transfected transiently with the following siRNAs: Setd2(0) against exon 13, Setd2(13) against exon 21 (encoding the CT21 epitope), or non-targeting control. TX–DOC extracts were prepared from quiesced cells and separated by 8% SDS-PAGE. Gels were transferred to PVDF in Towbin buffer before immunoblotting with α-SETD2 C-terminal **(i)**, N-terminal **(ii)**, or CT21 **(iii)** antibodies. The asterisk represents a background band, resulting from over-exposure.


siRNA :
Setd2 (0)
Setd2(3)
Non-t
α-SETD2 C-term
p290SETD2
*
(i)
α-SETD2 N-term
p290SETD2
(ii)
α-SETD2 CT21
250
130
95
72
55
36
p290CT21
p130CT21
p95CT21
p66CT21
p44CT21
(iii)
1
2
3

Appendix C Supplementary results for Chapter 5: *Identification of alternative* Setd2 *transcripts*

C-1 Exon–intron structure of alternatively spliced murine *Setd2* transcripts

C-1.1. Exon 12–Intron–Exon 12b

Splicing of exon 12 to exon 12b occurs by canonical sites for the major spliceosome:

Exon 12:

ATTATAAAGACTTTGGAGCATTTGCCCATTCCTACTAAAAATATGTTGGAAGAAAGCAAAGTACTTCCAATTATTCAGCGCTGGT
CTCAAACTAAGACTGCTGTTCCTCAGTTAAGTGAAGGAGACGGGTATTCTAGTGAGAATACATCACGTGCTCACACACCGCTTAA
TACACCGGACCCTTCTGCCAAGCCGAGCACGGAAATGGATACAGATACTCCCAAGAAACTCATATTCCGCAGACTGAAGATTATA
AGTGAAAATAGCATGGACAGTGCAGTCTCGGACGTCACTAGTGAGCTAGAATGCAAGGATGGCAAAGAGGACCTTGATCAGTTAG
AGACTGTCACTGTGGAAGAGGATGAAGAGCTGCAGTCCCAGCAGCTCCTCCCACAGCAGCTGTGCGAGTCCAAAGTTGAGAGTGA
AGCCACTATTGAAGTCAGTAAGTTACCCACATCTGAACCGGAGGCAGACACGGAGACAGAGCCCAAAGACAGCAATGGCACAAAA
CTAGAAGAAACTATTGCTGAGGAAACACCATCCCAAGATGAAGAAGAGGGAGTGTCTGATGTTGAAAGTGAGAGAAGCCAGGAGC
CACCAGATAAAACAGTGGATATAAGTGATTTGGCTACCAAGTTACTAGACAGTTGGAAAGATCTAAAG

Intron:

gtaagattaacctttttgtgttgt–[...intron...]–tgtaacttgaaccatttccatctag

Exon 12b:

CTTCTCAGGGCAGCAAGGAGCTTGTTAACATCACCAGCTGACCATGTGTCTCCTTTTCTCAACTCCATTTCCAACTTCGTGCTGA
AGAAAAACAAGACCCTAGCGTGCATTCTGTGAAGAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAAGTGTAATAGGT
GAAAAGAGGACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTG

C-1.2. Exon 12b–Intron–Exon 13

Splicing of exon 12b to exon 13 occurs by canonical sites for the major spliceosome:

Exon 12b:

CTTCTCAGGGCAGCAAGGAGCTTGTTAACATCACCAGCTGACCATGTGTCTCCTTTTCTCAACTCCATTTCCAACTTCGTGCTGA
AGAAAAACAAGACCCTAGCGTGCATTCTGTGAAGAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAAGTGTAATAGGT
GAAAAGAGGACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTG

Intron:

gtgcgttccttgtcaacaattgttg–[...intron...]–gtgtgtgttttttgttgttttgcag

Exon 13:

GAGGTGTATCGGATTCCAAAGAAAAGTCAAACTGAAAAGGAGAGCACAG

C-1.3. Exon 12b–Intron–Exon 12c

Splicing of exon 12b to exon 12c occurs by canonical sites for the major spliceosome:

Exon 12b:

CTTCTCAGGGCAGCAAGGAGCTTGTTAACATCACCAGCTGACCATGTGTCTCCTTTTCTCAACTCCATTTCCAACTTCGTGCTGA
AGAAAAACAAGACCCTAGCGTGCATTCTGTGAAGAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAAGTGTAATAGGT
GAAAAGAGGACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTG

Intron:

gtgcgttccttgtcaacaattgttgt–[...intron...]–ctccttttctcttatttcatctttttttttag

Exon 12c:

GCACTGCTTTTACAATACACTGTTTTAGCCAGGCCTGGTGGTGCACAACTTTAGCCTCAGCACTCAGAAGACAGAGGCAGGAGAG
TCTGTGAGTTTAAGGCCAGCCTGGTCTACATAGTGAGTTCCTGGAAATCCAGGGCTAGGTAGAGAGACCCTGTCTCAAAACCAAA
CCGAACTACATTTGGTTAGAAAATTCAACATTAAACTTTGCATAAGGCTTGAAATATGATTACCTAAACTAATGGTTTAGAGTTG
ATACTTGGAATTTAACTCCACCACTAATATTGTCTGTAAAAAAGATTTGTTAGCCAGGGCTATACAGAGAAGCCCTGTCTCAAAA
AACA...

C-2 *Setd2* RT-PCR sequencing data

C-2.1. Sequence of 12b–13

```
                Exon 12b
Predicted       TTGGGAAGTTGGATGATGCATGAAAAGTGTAATAGGTGAAAGAGGACAGAGGTGGTTAA 59
Forward         ----------------------------TNATANGNTAAANAAGACAGAGGTGGTTAA 30
Reverse         TTNAGAAGTTA-ATGAN-CNTGAAAAGTGTAATAGGTGAAAGAGGACAGAGGTGGTTAA 57
                **  ******  ****  * ********** *** *  *** * ***************

                                    |Exon 13
Predicted       CTCACTGGCGCCGGCCTTGTTTTGGAGGTGTATCGGATTCCAAAGAAAAGTCAAACTG 117
Forward         CTCACTGGCGCCGCCCTCGTTTTGGAGGTGTATCGGATTCCNAAGAAAAGNAAAANNC 88
Reverse         CTCACTGGCGCC------------------------------------------- 69
                ************* *** ********************** ********  ***
```

C-2.2. Sequence of 12b–13–14

```
                Exon 12b
Predicted       TTGGGAAGTTGGATGATGCATGAAAAGTGTAATAGGTGAAAAGAGGACAGAGGTGGTTAA 60
Forward         ---------TGGATGATGCATGAAAAGTGTAATAGGTGAAAAGAGGACAGAGGTGGTTAA 51
Reverse         TTGAGAAGTTA-ATGATGCNTGNAANTAATAATAGGTGAAAAGAGGACAGAGGTGGTTAA 59
                *** ****** ******** ** **    ********************************

                                    |Exon 13
Predicted       CTCACTGGCGCCGGCCTTGTTTTGGAGGTGTATCGGATTCCAAAGAAAAGTCAAACTGAA 120
Forward         CTCACTGGCGCCGGCCTTGTTTTGGAGGTNNATCGCATTCCANAAGAAAAGNAAACNGAA 111
Reverse         CTCACTGGCGCCGCCCTCGTTTTGGAGGTGTATCGGATTCCAAAGAAAAGTCAAACTGAA 119
                ****************************  *********** *****************

                            |Exon 14
Predicted       AAGGAGAGCACAGTAGCTGAACGAGGAAGAGATGCTGCTGCCTTCAGAGATCAAACAGCT 180
Forward         AAGAAGAGCACAGTAGCTGAACGAGGAAGAGATGCTGCTGCCTTCAGAGATCAAACAGCT 171
Reverse         AAGAAGAGCACAGTAGCTGAACGAGGAAGAGATGCTGCTGCCTTCAGAGATCAAACAGCT 179
                ************************************************************

Predicted       CCAAAGACTCCTAACAGGTCTAGAGAGAGAGACCCAGACAAGCAGTCTCAAAATAAAGAG 240
Forward         CCAAAGACTCCTAACAGGTCTAGAGAGAGAGACCCAGACAAGCAGTCTCAAAATAAAGAG 231
Reverse         CCAAAGACTCCTAACAGGTCTAGAGAGAGAGACCCAGACAAGCAGTCTCAAAATAAAGAG 239
                ************************************************************

Predicted       AAAAGGAAACGACGGGGCTCTCTGTCACC 269
Forward         AAAANNAN-CGACGNG-CTC--------- 249
Reverse         AAAAGGAAACGACGGGGCTCTCTGTCNNC 299
                ****  *  ***** * *********  *
```

C-2.3. Sequence of 12b–13–14–15

```
               Exon 12b
Predicted      TCAACTCCATTTCCAACTTCGTGCTGAAGAAAAACAAGACCCTAGCGTGCATTCTGTGAA  60
Forward        ------------------------------------------------------TGAA   4
Reverse        TTNTATGTCTCTCCACCTTCGTGGTGAAGAAAAACAACACCCCCGCGGGCATCCTTTGAA  60
               *  ****  * **** ******* ************ ****  *** **** ** ****

Predicted      GAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAAGTGTAATAGGTGAAAAGAG 120
Forward        GAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAAGTGTAATAGGTGAAAAGAG  64
Reverse        GAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAAGTGTAATAGGTGAAAAGAG 120
               ************************************************************
```

```
                                                        |Exon 13
Predicted       GACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTGGAGGTGTATCGGATTCCAAAG 180
Forward         GACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTGGAGGTGTATCGGATTCCAAAG 124
Reverse         GACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTGGAGGTGTATCGGATTCCAAAG 180
                ************************************************************

                                              |Exon 14
Predicted       AAAAGTCAAACTGAAAAGGAGAGCACAGTAGCTGAACGAGGAAGAGATGCTGCTGCCTTC 240
Forward         AAAAGTCAAACTGAAAAGGAGAGCACAGTAGCTGAACGAGGAAGAGATGCTGCTGCCTTC 184
Reverse         AAAAGTCAAACTGAAAAGGAGAGCACAGTAGCTGAACGAGGAAGAGATGCTGCTGCCTTC 240
                ************************************************************

Predicted       AGAGATCAAACAGCTCCAAAGACTCCTAACAGGTCTAGAGAGAGAGACCCAGACAAGCAG 300
Forward         AGAGATCAAACAGCTCCAAAGACTCCTAACAGGTCTAGAGAGAGAGACCCAGACAAGCAG 244
Reverse         AGAGATCAAACAGCTCCAAAGACTCCTAACAGGTCTAGAGAGAGAGACCCAGACAAGCAG 300
                ************************************************************

Predicted       TCTCAAAATAAAGAGAAAAGGAAACGACGGGGCTCTCTGTCACCACCCTCTTCTGCATAT 360
Forward         TCTCAAAATAAAGAGAAAAGGAAACGACGGGGCTCTCTGTCACCACCCTCTTCTGCATAT 304
Reverse         TCTCAAAATAAAGAGAAAAGGAAACGACGGGGCTCTCTGTCACCACCCTCTTCTGCATAT 360
                ************************************************************

                                                |Exon 15
Predicted       GAGCGGGGAACAAAAAGGCCAGATGACAGATATGATACACCAACTTCTAAAAAGAAAGTA 420
Forward         GAGCGGGGAACAAAAAGGCCAGATGACAGATATGATACACCAACTTCTAAAAAGAAAGTA 364
Reverse         GAGCGGGGAACAAAAAGGCCAGATGACAGATATGATACACCAACTTCTAAAAAGAAAGTA 420
                ************************************************************

Predicted       CGAATTAAAGACCGAAACAAACTTTCTACAGAGGAGCGCAGGAAGTTGTTTGAACAAGAG 480
Forward         CGAATTAAACACCGCAACAAACTTTCTACAGAGGAGCGCAGGAAGTTGTTTGAACAAGAG 424
Reverse         CGAATTAAAGACCGCAACAAACTTTCTACAGAGGAGCGCAGGAAGTTGTTTGAACAAGAG 480
                ************************************************************

Predicted       GTGGCACAGAGGGAAGCTCAGAAGCAACAGCAACAGATGCAGAACTTGGGGATGACAT 540
Forward         GTGGCACAGAGGGAAGCTCAGAAGCAACAGCAACAGATGCAGAACTTGGTGAC-ATAN 483
Reverse         GTN-CACAGAGGNAG---------------------------------------- 494
                ************ * ********************************** **  * *
```

C-2.4. Sequence of 12–12b

```
                Exon 12
Predicted       GAAACACCATCCCAAGATGAAGAAGAGGGAGTGTCTGATGTTGAAAGTGAGAGAAGCCAG 60
Forward         ----------------------------------------------ATGAGAGAAGCCAG 14
Reverse         GAAACACCATCCCAAGATGAAGAAGAGGGAGTGTCTGATGTTGAAAGTGAGAGAAGCCAG 60
                ************************************************************

Predicted       GAGCCACCAGATAAAACAGTGGATATAAGTGATTTGGCTACCAAGTTACTAGACAGTTGG 120
Forward         GAGCCACCAGATAAAACAGTGGATATAAGTGATTTGGCTACCAAGTCACTAGACAGTTGG 74
Reverse         GAGCCACCAGATAAAACAGTGGATATAAGTGATTTGGCTACCAAGTTACTAGACAGTTGG 120
                ******************************************** ************

                        |Exon 12b
Predicted       AAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTTGTTAACATCACCAGCTGACCATGTG 180
Forward         AAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTGGTTAACATCACCAGCTGACCATGTG 134
Reverse         AAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTTGTTAACATCACCAGCTGACCATGTG 180
                ********************************** ************************

Predicted       TCTCCTTTTCTCAACTCCATTTCCAACTTCGTGCTGAAGAAAAACAAGACCCTAGCGTGC 240
Forward         TCTCCTTTTCTCAACTCCATTTCCAACTTCNGGCTGAAAAAAAACAAGACCCTANCGGGN 194
Reverse         TCTCCTTTTCTCAACTCCATTTC----------------------------------- 203
                ****************************  ********************** ** *

Predicted       ATTCTGTG 248
Forward         NTTCTGGG 202
Reverse         --------
                 ***** *
```

C-2.5. Sequence of 11–12–12b

```
                Exon 11
Predicted       -TCATTTCTGGAACGTCATGGGTTGTCACTGTTGTGGATCTGGATGGCAGAGCTTGGCGA 59
Forward         ------------------------------------------------------------
Reverse         TTCATTTCTGGAACGTCATGGGTTGTCACTGTTGTGGATCTGGATGGCAGAGCTTGGCGA 60
                 ************************************************************

                                          | Exon 12
Predicted       CGGCCGGGAAAGTAACCAGAAGCTTCAGGAAGAGATTATAAAGACTTTGGAGCATTTGCC 119
Forward         ------------------------------------TAAAGACTTTGGAGCATTTGCC 22
Reverse         CGGCCGGGAAAGTAACCAGAAGCTTCAGGAAGAGATTATAAAGACTTTGGAGCATTTGCC 120
                ************************************************************

Predicted       CATTCCTACTAAAAATATGTTGGAAGAAAGCAAAGTACTTCCAATTATTCAGCGCTGGTC 179
Forward         CATTCCTACTAAAAATATGTTGGAAGAAAGCAAAGTACTTCCAATTATTCAGCGGTGGTC 81
Reverse         CATTCCTACTAAAAATATGTTGGAAGAAAGCAAAGTACTTCCAATTATTCAGCGCTGGTC 180
                ************************************************************

Predicted       TCAAACTAAGACTGCTGTTCCTCAGTTAAGTGAAGGAGACGGGTATTCTAGTGAGAATAC 239
Forward         TCAAACTAAGACTGCTGTTCCTCAGTTAAGTGAAGGAGACGGGTATTCTAGTGAGAATAC 140
Reverse         TCAAACTAAGACTGCTGTTCCTCAGTTAAGTGAAGGAGACGGGTATTCTAGTGAGAATAC 240
                ************************************************************

Predicted       ATCACGTGCTCACACACCGCTTAATACACCGGACCCTTCTGCCAAGCCGAGCACGGAAAT 299
Forward         ATCACGTGCTCACACACCGCTTAATACACCGGACCCTTCTGCCAAGCCGTGCACGGAAAT 199
Reverse         ATCACGTGCTCACACACCGCTTAATACACCGGACCCTTCTGCCAAGCCGAGCACGGAAAT 299
                ************************************************************

Predicted       GGATACAGATACTCCCAAGAAACTCATATTCCGCAGACTGAAGATTATAAGTGAAAATAG 359
Forward         GGATACAGATACTCCCAAGAAACTCATATTCCGCAGACTGAAGATTATAAGTGAAAATAG 259
Reverse         GGATACAGATACTCCCAAGAAACTCATATTCCGCAGACTGAAGATTATAAGTGAAAATAG 359
                ************************************************************

Predicted       CATGGACAGTGCAGTCTCGGACGTCACTAGTGAGCTAGAATGCAAGGATGGCAAAGAGGA 419
Forward         CATGGACAGTGCAGTCTCGGACGTCACTAGTGAGCTAGAATGCAAGGATGGCAAAGAGGA 319
Reverse         CATGGACAGTGCAGTCTCGGACGTCACTAGTGAGCTAGAATGCAAGGATGGCAAAGAGGA 419
                ************************************************************

Predicted       CCTTGATCAGTTAGAGACTGTCACTGTGGAAGAGGATGAAGAGCTGCAGTCCCAGCAGCT 479
Forward         CCTTGATCAGTTAGAGACTGTCACTGTGGAAGAGGATGAAGAGCTGCAGTCCCAGCAGCT 379
Reverse         CCTTGATCAGTTAGAGACTGTCACTGTGGAAGAGGATGAAGAGCTGCAGTCCCAGCAGCT 479
                ************************************************************

Predicted       CCTCCCACAGCAGCTGTGCGAGTCCAAAGTTGAGAGTGAAGCCACTATTGAAGTCAGTAA 539
Forward         CCTCCCACAGCAGCTGTGCGAGTCCAAAGTTGAGAGTGAAGCCACTATTGAAGTCAGTAA 439
Reverse         CCTCCCACAGCAGCTGTGCGAGTCCAAAGTTGAGAGTGAAGCCACTATTGAAGTCAGTAA 539
                ************************************************************

Predicted       GTTACCCACATCTGAACCGGAGGCAGACACGGAGACAGAGCCCAAAGACAGCAATGGCAC 599
Forward         GTTACCCACATCTGAACCGGAGGCAGACACGGAGACAGAGCCCAAAGACAGCAATGGCAC 499
Reverse         GTTACCCACATCTGAACCGGAGGCAGACACGGAGACAGAGCCCAAAGACAGCAATGGCAC 599
                ************************************************************

Predicted       AAAACTAGAAGAAACTATTGCTGAGGAAACACCATCCCAAGATGAAGAAGAGGGAGTGTC 659
Forward         AAAACTAGAAGAAACTATTGCTGAGGAAACACCATCCCAAGATGAAGAAGAGGGAGTGTC 559
Reverse         AAAACTAGAAGAAACTATTGCTGAGGAAACACCATCCCAAGATGAAGAAGAGGGAGTGTC 659
                ************************************************************

Predicted       TGATGTTGAAAGTGAGAGAAGCCAGGAGCCACCAGATAAAACAGTGGATATAAGTGATTT 719
Forward         TGATGTTGAAAGTGAGAGAAGCCAGGAGCCACCAGATAAAACAGTGGATATAAGTGATTT 619
Reverse         TGATGTTGAAAGTGAGAGAAGCCAGGAGCCACCAGATAAAACAGTGGATATAAGTGATTT 719
                ************************************************************

                                              | Exon 12b
Predicted       GGCTACCAAGTTACTAGACAGTTGGAAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTT 779
Forward         GGCTACCAAGTTACTAGACAGTTGGAAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTT 679
Reverse         GGCTACCAAGTTACTAGACAGTTGGAAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTT 779
                ************************************************************

Predicted       GTTAACATCACCAGCTGACCATGTGTCTCCTTTTCTCAACTCCATTTCCAACTTCGTGCT 839
Forward         GTTAACATCACCAGCTGACCATGTGTCTCCTTTTCTCAACTCCATTTCCAACTTCGTGCT 739
Reverse         GTTAACATCACCAGCTGACCATGTGTCTCCTTTTCTCAACTC------------------ 820
                ************************************************************


Predicted       GAAGAAAAACAAGACCCTAGCGTGCATTCTGTG---------------------------- 871
Forward         GAAGAAAAACAAGACCCTAGCGTGCATTCTNTGN--------------------------- 773
Reverse         ------------------------------------------------------------
                **************************** **
```

C-2.6. Sequence of 12–12b–12c

```
                Exon 12
Predicted       GAGGAAACACCATCCCAAGATGAAGAAGAGGGAGTGTCTGATGTTGAAAGTGAGAGAAGC 60
Forward         ---------------------------------GTCTGATGTTGAAAGTGAGAGAAGC 25
Reverse         GGGGAAACCCCATCCCAAGATGAAGAAGAGGGAGTGTCTGATGTTGAAAGTGAGAGAAGC 60
                ************************************************************

Predicted       CAGGAGCCACCAGATAAAACAGTGGATATAAGTGATTTGGCTACCAAGTTACTAGACAGT 120
Forward         CAGGAGCCACCAGATAAAACAGTGGATATAAGTGATTTGGCTACCAAGTTACTAGACAGT 85
Reverse         CAGGAGCCACCAGATAAAACAGTGGATATAAGTGATTTGGCTACCAAGTTACTAGACAGT 120
                ************************************************************

                          |Exon 12b
Predicted       TGGAAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTTGTTAACATCACCAGCTGACCAT 180
Forward         TGGAAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTTGTTAACATCACCAGCTGACCAT 145
Reverse         TGGAAAGATCTAAAGCTTCTCAGGGCAGCAAGGAGCTTGTTAACATCACCAGCTGACCAT 180
                ************************************************************

Predicted       GTGTCTCCTTTTCTCAACTCCATTTCCAACTTCGTGCTGAAGAAAAACAAGACCCTAGCG 240
Forward         GTGTCTCCTTTTCTCAACTCCATTTCCAACTTCGTGCTGAAGAAAAACAAGACCCTAGCG 205
Reverse         GTGTCTCCTTTTCTCAACTCCATTTCCAACTTCGTGCTGAAGAAAAACAAGACCCTAGCG 240
                ************************************************************

Predicted       TGCATTCTGTGAAGAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAAGTGTAA 300
Forward         TGCATTCTGTGAAGAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAAGTGTAA 265
Reverse         TGCATTCTGTGAAGAAACAGGAAGATGGTTGGGAAGTTGGATGATGCATGAAAAGTGTAA 300
                ************************************************************

                                                                  |Exon 12c
Predicted       TAGGTGAAAAGAGGACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTGGCACTGCT 360
Forward         TAGGTGAAAAGAGGACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTGGCACTGCT 325
Reverse         TAGGTGAAAAGAGGACAGAGGTGGTTAACTCACTGGCGCCGGCCTTGTTTTGGCACTGCT 360
                ************************************************************

Predicted       TTTACAATACACTGTTTTAGCCAGGCCTGGTGGTGCACAACTTTAGCCTCAGCA------ 444
Forward         TTTACAATACACTGTTTTAGCCAGGCCTGGTGGTGCACAACTTTAGCCTCNNN------- 378
Reverse         TTTACAATACAC---------------------------------------------- 372
                ************************************************
```

C-2.7. Sequence of 6–7–8–9–10.5

```
                Exon 6                                        |Exon 7
Predicted       ----------TACTACTTCATGGCCCTGAAAAATGACGAGATAATAGATGCCACTCAAAA 50
Forward         ------------------------------------------------------------
Reverse         NNNCNAGCNNTTCNNTTTCATGGCCCTGAAAAATGACGAGATAATAGATGCCACTCAAAA 60
                          ****************************************

                                                                      |Exon 8
Predicted       AGGGAATTGCTCTCGTTTCATGAATCATAGCTGTGAACCAAACTGTGAAACCCAGAAATG 110
Forward         NNNNNATTGCTCTCGTTTCATGAATCATAGCTGTGAACCAAACTGTGAAACCCAGAAATG 60
Reverse         AGGGAATTGCTCTCGTTTCATGAATCATAGCTGTGAACCAAACTGTGAAACCCAGAAATG 120
                ************************************************************

Predicted       GACTGTGAATGGACAGCTGAGGGTTGGATTTTTTACCACCAAACTAGTTCCTTCAGGCTC 170
Forward         GACTGTGAATGGACAGCTGAGGGTTGGATTTTTTACCACCAAACTAGTTCCTTCAGGCTC 120
Reverse         GACTGTGAATGGACAGCTGAGGGTTGGATTTTTTACCACCAAACTAGTTCCTTCAGGCTC 180
                ************************************************************

                                                |Exon 9
Predicted       AGAATTAACTTTTGACTACCAGTTCCAAAGATATGGCAAAGAAGCTCAGAAGTGTTTCTG 230
Forward         AGAATTAACTTTTGACTACCAGTTCCAAAGATATGGCAAAGAAGCTCAGAAGTGTTTCTG 180
Reverse         AGAATTAACTTTTGACTACCAGTTCCAAAGATATGGCAAAGAAGCTCAGAAGTGTTTCTG 240
                ************************************************************

Predicted       TGGGTCAGCCAACTGCCGGGGCTACTTGGGAGGAGAAAACAGAGTCAGTATCAGAGCTGC 290
Forward         TGGGTCAGCCAACTGCCGGGGCTACTTGGGAGGAGAAAACAGAGTCAGTATCAGAGCTGC 240
Reverse         TGGGTCAGCCAACTGCCGGGGCTACTTGGGAGGAGAAAACAGAGTCAGTATCAGAGCTGC 300
                ************************************************************

                                                          |Exon 10
Predicted       AGGAGGGAAGATGAAAAAGGAACGCTCTCGAAAGAAGGATTCAGTGGATGGAGAACTTGA 350
Forward         AGGAGGGAAGATGAAAAAGGAACGCTCTCGAAAGAAGGATTCAGTGGATGGAGAACTTGA 300
Reverse         AGGAGGGAAGATGAAAAAGGAACGCTCTCGAAAGAAGGATTCAGTGGATGGAGAACTTGA 360
                ************************************************************

Predicted       AGCACTGATGGAAAATGGTGAAGGCCTCTCTGATAAGAACCAGGTTCTTAGTTTATCCCG 410
Forward         AGCACTGATGGAAAATGGTGAAGGCCTCTCTGATAAGAACCAGGTTCTTAGTTTATCCCG 360
Reverse         AGCACTGATGGAAAATGGTGAAGGCCTCTCTGATAAGAACCAGGTTCTTAGTTTATCCCG 420
                ************************************************************

                                                                  |Intron 10-11
Predicted       GCTCATGGTTAGAATTGAAACTTTGGAACAGAAACTTACCTGTCTTAAGCTCATTCAGGT 470
Forward         GCTCATGGTTAGAATTGAAACTTTGGAACAGAAACTTACCTGTCTTAAGCTCATTCAGGT 420
Reverse         GCTCATGGTTAGAATTGAAACTTTGGAACAGAAACTTACCTGTCTTAAGCTCATTCAGGT 480
                ************************************************************

Predicted       AAGCACTTCAATACTAATATATTTCTGAATTAAGGGTAGGACATTTAAACTAAAATCTGG 530
Forward         AAGCACTTCAATACTAATATATTTCTGAATTAAGGGTAGGACATTTAAACTAAAATCTGG 480
Reverse         AAGCACTTCAATACTAATATATTTCTGAATTAAGGGTAGGACATTTAAACTAAAATCTGG 540
                ************************************************************

Predicted       CTAATGGAGAAAAATTAATAATGTTTTAAATTGTCATGCAGGTTCTTTAGAACTTTGAAA 590
Forward         CTAATGGAGAAAAATTAATAATGTTTTAAATTGTCATGCAGGTTCTTTAGAACTTTGAAA 540
Reverse         CTAATGGAGAAAAATTAATAATGTTTTAAATTGTCATGCAGGTTCTTTAGAACTTTGAAA 600
                ************************************************************

Predicted       GCTTATACCCTTGACTGAATAGTTTAATATGAATAGCAAAGCTCGTTTG-CTTACCCCTT 649
Forward         GCTTATACCCTTGACTGAATAGTTTAATATGAATAGCAAAGCTCGTTTNGCTTNCCCCTN 600
Reverse         GCTTATACCCTTGANNNANNGN--TAA------------------------------- 625
                ********************************************  *** ******
```

C-2.8. Sequence of 8–9.5

```
                Exon 8
Predicted       GACAGCTGAGGGTTGGATTTTTTACCACCAAACTAGTTCCTTCAGGCTCAGAATTAACTT 60
Forward         ----------------------------------------------TCAGAATTAACTT 13
Reverse         NNNAGCTGAGGGTTGGATTTTTTACCACCAAACTAGTTCCTTCAGGCTCAGAATTAACTT 60
                   ************************************************************

                                        |Exon 9
Predicted       TTGACTACCAGTTCCAAAGATATGGCAAAGAAGCTCAGAAGTGTTTCTGTGGGTCAGCCA 120
Forward         TTGACTACCAGTTCCAAAGATATGGCAAAGAAGCTCAGAAGTGTTTCTGTGGGTCAGCCA 73
Reverse         TTGACTACCAGTTCCAAAGATATGGCAAAGAAGCTCAGAAGTGTTTCTGTGGGTCAGCCA 120
                ************************************************************

Predicted       ACTGCCGGGGCTACTTGGGAGGAGAAAACAGAGTCAGTATCAGAGCTGCAGGAGGGAAGA 180
Forward         ACTGCCGGGGCTACTTGGGAGGAGAAAACAGAGTCAGTATCAGAGCTGCAGGAGGGAAGA 133
Reverse         ACTGCCGGGGCTACTTGGGAGGAGAAAACAGAGTCAGTATCAGAGCTGCAGGAGGGAAGA 180
                ************************************************************

                                              |Intron 9-10
Predicted       TGAAAAAGGAACGCTCTCGAAAGAAGGATTCAGTAAGTTTTCGGCTCTCTGTTCAAGTTT 240
Forward         TGAAAAAGGAACGCTCTCGAAAGAAGGATTCAGTAAGTTTTCGGCTCTCTGTTCAAGTTT 193
Reverse         TGAAAAAGGAACGCTCTCGAAAGAAGGATTCAGTAAGTTTTCGGCTCTCTGTTCAAGTTT 238
                ************************************************************

Predicted       TAAGGAAATAGCTTTTCTGATTAATTTAACAGGAAAATCCAGGTTTGGA----------- 289
Forward         TAAGGAAATAGCTTTTCTGATTAATTTAACAGGAAAATCCAGGTTTGGA----------- 242
Reverse         ------------------------------------------------------------
                ************************************************
```

C-2.9. Sequence of 17–18.5

```
                Exon 17
Predicted       TTTTACCTCCCAACTGGAAGACAGCCCGAGACCCTGAGGGGAAGATCTACTACTACCACG 60
Forward         ------------------------------------GGAAGATCTACTACTACCACG 21
Reverse         TTTTACCTCCCAACTGGAAGACAGCCCGAGACCCTGAGGGGAAGATCTACTACTACCACG 60
                ************************************************************

                        |Exon 18
Predicted       TGATCACAAGACAGACTCAGTGGGATCCTCCTACTTGGGAAAGCCCAGGAGATGATGCCA 120
Forward         TGATCACAAGACAGACTCAGTGGGATCCTCCTACTTGGGAAAGCCCAGGAGATGATGCCA 81
Reverse         TGATCACAAGACAGACTCAGTGGGATCCTCCTACTTGGGAAAGCCCAGGAGATGATGCCA 120
                ************************************************************

Predicted       GCCTTGAGCATGAAGCTGAAATGGACCTGGGAACCCCAACCTATGATGAAAACCCCATGA 180
Forward         GCCTTGAGCATGAAGCTGAAATGGACCTGGGAACCCCAACCTATGATGAAAACCCCATGA 141
Reverse         GCCTTGAGCATGAAGCTGAAATGGACCTGGGAACCCCAACCTATGATGAAAACCCCATGA 180
                ************************************************************

                |Intron 18-19
Predicted       AGGTGAGTGTAGCTTTTGTGTCCTCATCATTTGAGCATCTGTGTCCTTTCATTCAGTGAG 240
Forward         AGGTGAGTGTAGCTTTTGTGTCCTCATCATTTGAGCATCTGTGTCCTTTCATTCAGTGAG 201
Reverse         AGGTGAGTGTAGCTTTTGTGTCCTCATCATTTGAGCATCTGTGTCCTTTCATTCAGTGAG 240
                ************************************************************

Predicted       GTTAGTAACCATATTGTTCTGGGTTTTTTGTTTTCTTTCTTTGGAAAAATATGTCAAAGT 300
Forward         GTTAGTAACCATATTGTTCTGGGTTTTTTGTTTTCTTTCTTTGGAAAAATATGTCAAAGT 261
Reverse         GTTAGTAACCATATTGTTCTGGGTTTTTTGTTTTCTTTCTTTGGAAAAATATGTCAAAGT 300
                ************************************************************

Predicted       TCAGAGCTGGAGAGATGGCTTAGCAGTTACAAGCACTGGCACCAGAGGACCCAGGTTCAA 360
Forward         TCAGAGCTGGAGAGATGGCTTAGCAGTTACAAGCACTGGCACCAGAGGACCCAGGTTCAA 321
Reverse         TCAGAGCTGGAGAGATGGCTTAGCAGTTACAAGCACTGGCACCAGAGGACCCAGGTTCAA 360
                ************************************************************

Predicted       TTTCTAGTCCTACGTGACAGGTTACAACTGCCTATAACTCCAATTTCAGGTGATCTGATA 420
Forward         TTTCTAGTCCTACGTGACAGGTTACAACTGCCTATAACTCCAATTTCAGGTGATCTGATA 381
Reverse         TTTCTAGTCCTACGTGACAGGTTACAACTGCCTATAACTCCAATTTCAGGTGATCTGATA 420
                ************************************************************

Predicted       CCTTTACACAGACATATATGCAGGAAACATACCAGTGTGTGTGAAATTTTAAGAAATGTC 480
Forward         CCTTTACACAGACATATATGCAGGAAACATACCAGTGTGTGTGAAATTTTAAGAAATGTC 441
Reverse         CCTTTACACAGACATATATGCAGGAAACATACCAGTGTGTGTGAAATTTTAAGAAATGTC 480
                ************************************************************

Predicted       AAAGTTCTGCTGTATCAGAGGAGAGACTGAAGTCAAAGACCATTTTTAGACATTGAAGGG 540
Forward         AAAGTTCTGCTGTATCAGAGGAGAGACTGAAGTCAAAGACCATTTTTAGACATTGAAGGG 501
Reverse         AAAGTTCTGCTGTATCAGAGGAGAGA---------------------------------- 506
                ************************************************************

Predicted       TCAGATGCACC---------------------------------------------- 551
Forward         TCAGATGCACC---------------------------------------------- 512
Reverse         ------------------------------------------------------------
                ***********
```

References

Abramoff MD, Magalhaes PJ & Ram SJ (2004) Image Processing with ImageJ. *Biophotonics International,* **11** (7), 36-42.

Adhvaryu KK, Morris SA, Strahl BD & Selker EU (2005) Methylation of histone H3 lysine 36 is required for normal development in *Neurospora crassa*. *Eukaryot Cell,* **4** (8), 1455-1464.

Akhmedov AT & Lopez BS (2000) Human 100-kDa homologous DNA-pairing protein is the splicing factor PSF and promotes DNA strand invasion. *Nucleic Acids Res,* **28** (16), 3022-3030.

Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D & Ansari AZ (2009) TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell,* **34** (3), 387-393.

Al Sarakbi W, Sasi W, Jiang WG, Roberts T, Newbold RF & Mokbel K (2009) The mRNA expression of SETD2 in human breast cancer: correlation with clinico-pathological parameters. *BMC Cancer,* **9**, 290.

Allegra P, Sterner R, Clayton DF & Allfrey VG (1987) Affinity chromatographic purification of nucleosomes containing transcriptionally active DNA sequences. *J Mol Biol,* **196** (2), 379-388.

Allemand E, Batsche E & Muchardt C (2008) Splicing, transcription, and chromatin: a menage à trois. *Curr Opin Genet Dev,* **18** (2), 145-151.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* **25** (17), 3389-3402.

An S, Yeo KJ, Jeon YH & Song JJ (2011) Crystal Structure of the Human Histone Methyltransferase ASH1L Catalytic Domain and Its Implications for the Regulatory Mechanism. *J Biol Chem,* **286** (10), 8369-8374.

Andersen EC & Horvitz HR (2007) Two *C. elegans* histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. *Development,* **134** (16), 2991-2999.

Andersson R, Enroth S, Rada-Iglesias A, Wadelius C & Komorowski J (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res,* **19** (10), 1732-1741.

Angrand PO, Apiou F, Stewart AF, Dutrillaux B, Losson R & Chambon P (2001) *NSD3*, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. *Genomics,* **74** (1), 79-88.

Apostol BL, Illes K, Pallos J, Bodai L, Wu J, Strand A, Schweitzer ES, Olson JM, Kazantsev A, Marsh JL & Thompson LM (2006) Mutant huntingtin alters MAPK signaling pathways in PC12 and striatal cells: ERK1/2 protects against mutant huntingtin-associated toxicity. *Hum Mol Genet,* **15** (2), 273-285.

Avvakumov N, Nourani A & Cote J (2011) Histone chaperones: modulators of chromatin marks. *Mol Cell,* **41** (5), 502-514.

Bae BI, Xu H, Igarashi S, Fujimuro M, Agrawal N, Taya Y, Hayward SD, Moran TH, Montell C, Ross CA, Snyder SH & Sawa A (2005) p53 mediates cellular dysfunction and behavioral abnormalities in Huntington's disease. *Neuron,* **47** (1), 29-41.

Bannister AJ, Schneider R, Myers FA, Thorne AW, Crane-Robinson C & Kouzarides T (2005) Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem,* **280** (18), 17732-17736.

Bantignies F & Cavalli G (2006) Cellular memory and dynamic regulation of polycomb group proteins. *Curr Opin Cell Biol,* **18** (3), 275-283.

Barrand S, Andersen IS & Collas P (2010) Promoter-exon relationship of H3 lysine 9, 27, 36 and 79 methylation on pluripotency-associated genes. *Biochem Biophys Res Commun,* **401** (4), 611-617.

Barrand S & Collas P (2010) Chromatin states of core pluripotency-associated genes in pluripotent, multipotent and differentiated cells. *Biochem Biophys Res Commun,* **391** (1), 762-767.

Barratt MJ, Hazzalin CA, Cano E & Mahadevan LC (1994a) Mitogen-stimulated phosphorylation of histone H3 is targeted to a small hyperacetylation-sensitive fraction. *Proceedings of the National Academy of Sciences of the United States of America,* **91** (11), 4781-4785.

Barratt MJ, Hazzalin CA, Zhelev N & Mahadevan LC (1994b) A mitogen- and anisomycin-stimulated kinase phosphorylates HMG-14 in its basic amino-terminal domain *in vivo* and on isolated mononucleosomes. *EMBO J,* **13** (19), 4524-4535.

Barrera LO, Li Z, Smith AD, Arden KC, Cavenee WK, Zhang MQ, Green RD & Ren B (2008) Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res,* **18** (1), 46-59.

Barski A, Chepelev I, Liko D, Cuddapah S, Fleming AB, Birch J, Cui K, White RJ & Zhao K (2010) Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. *Nat Struct Mol Biol,* **17** (5), 629-634.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I & Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell,* **129** (4), 823-837.

Barski A, Jothi R, Cuddapah S, Cui K, Roh TY, Schones DE & Zhao K (2009) Chromatin poises miRNA- and protein-coding genes for expression. *Genome Res,* **19** (10), 1742-1751.

Barski A & Zhao K (2009) Genomic location analysis by ChIP-Seq. *J Cell Biochem,* **107** (1), 11-18.

Basu SK, Malik R, Huggins CJ, Lee S, Sebastian T, Sakchaisri K, Quinones OA, Alvord WG & Johnson PF (2011) 3' UTR elements inhibit Ras-induced C/EBPβ post-translational activation and senescence in tumour cells. *EMBO J,* **30**, 3714-3728.

Bauer WR, Hayes JJ, White JH & Wolffe AP (1994) Nucleosome structural changes due to acetylation. *J Mol Biol,* **236** (3), 685-690.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS & Haussler D (2004) Ultraconserved elements in the human genome. *Science,* **304** (5675), 1321-1325.

Bell O, Conrad T, Kind J, Wirbelauer C, Akhtar A & Schubeler D (2008) Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Mol Cell Biol,* **28** (10), 3401-3409.

Bell O, Wirbelauer C, Hild M, Scharf AN, Schwaiger M, MacAlpine DM, Zilbermann F, van Leeuwen F, Bell SP, Imhof A, Garza D, Peters AH & Schubeler D (2007) Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in *Drosophila*. *EMBO J,* **26** (24), 4974-4984.

Belle A, Tanay A, Bitincka L, Shamir R & O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A,* **103** (35), 13004-13009.

Bender LB, Suh J, Carroll CR, Fong Y, Fingerman IM, Briggs SD, Cao R, Zhang Y, Reinke V & Strome S (2006) MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the *C. elegans* germ line. *Development,* **133** (19), 3907-3917.

Benn CL, Sun T, Sadri-Vakili G, McFarland KN, DiRocco DP, Yohrling GJ, Clark TW, Bouzou B & Cha JH (2008) Huntingtin modulates transcription, occupies gene promoters *in vivo*, and binds directly to DNA in a polyglutamine-dependent manner. *J Neurosci,* **28** (42), 10720-10733.

Bennion BJ & Daggett V (2003) The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci U S A,* **100** (9), 5142-5147.

Berezney R & Coffey DS (1974) Identification of a nuclear protein matrix. *Biochem Biophys Res Commun,* **60** (4), 1410-1417.

Bergmann JH, Rodriguez MG, Martins NM, Kimura H, Kelly DA, Masumoto H, Larionov V, Jansen LE & Earnshaw WC (2011) Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO J,* **30**, 328-340.

Biswas D, Dutta-Biswas R, Mitra D, Shibata Y, Strahl BD, Formosa T & Stillman DJ (2006) Opposing roles for Set2 and yFACT in regulating TBP binding at promoters. *EMBO J,* **25** (19), 4479-4489.

Biswas D, Takahata S, Xin H, Dutta-Biswas R, Yu Y, Formosa T & Stillman DJ (2008) A role for Chd1 and Set2 in negatively regulating DNA replication in *Saccharomyces cerevisiae*. *Genetics,* **178** (2), 649-659.

Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ & Klose RJ (2010) CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell,* **38** (2), 179-190.

Blahnik KR, Dou L, Echipare L, Iyengar S, O'Geen H, Sanchez E, Zhao Y, Marra MA, Hirst M, Costello JF, Korf I & Farnham PJ (2011) Characterization of the contradictory chromatin signatures at the 3' exons of zinc finger genes. *PLoS One,* **6** (2), e17121.

Bohmann D, Bos TJ, Admon A, Nishimura T, Vogt PK & Tjian R (1987) Human proto-oncogene c-*jun* encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. *Science,* **238** (4832), 1386-1392.

Boucher L, Ouzounis CA, Enright AJ & Blencowe BJ (2001) A genome-wide survey of RS domain proteins. *RNA,* **7** (12), 1693-1701.

Brookes E & Pombo A (2009) Modifications of RNA polymerase II are pivotal in regulating gene expression states. *EMBO Rep,* **10** (11), 1213-1219.

Brown MA, Sims RJ, 3rd, Gottlieb PD & Tucker PW (2006) Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Mol Cancer,* **5**, 26.

Brown SE, Campbell RD & Sanderson CM (2001) Novel NG36/G9a gene products encoded within the human and mouse MHC class III regions. *Mamm Genome,* **12** (12), 916-924.

Burgess RR (2009) Important but little known (or forgotten) artifacts in protein biochemistry. *Methods Enzymol,* **463**, 813-820.

Buro LJ, Chipumuro E & Henriksen MA (2010) Menin and RNF20 recruitment is associated with dynamic histone modifications that regulate signal transducer and activator of transcription 1 (STAT1)-activated transcription of the interferon regulatory factor 1 gene (IRF1). *Epigenetics Chromatin,* **3** (1), 16.

Caiafa P & Zampieri M (2005) DNA methylation and chromatin structure: the puzzling CpG islands. *J Cell Biochem,* **94** (2), 257-265.

Cappai R, Leck SL, Tew DJ, Williamson NA, Smith DP, Galatis D, Sharples RA, Curtain CC, Ali FE, Cherny RA, Culvenor JG, Bottomley SP, Masters CL, Barnham KJ & Hill AF (2005) Dopamine promotes α-synuclein aggregation into SDS-resistant soluble oligomers via a distinct folding pathway. *FASEB J,* **19** (10), 1377-1379.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest AR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, Ambesi-Impiombato A, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, Della Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T, Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminiecki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasawa Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SP, Kruger A, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Madan Babu M, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rost B, Ruan Y, Salzberg SL, Sandelin A, Schneider C, Schonbach C, Sekiguchi K, Semple CA, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammoja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegen E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusic V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M, Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J & Hayashizaki Y (2005) The transcriptional landscape of the mammalian genome. *Science,* **309** (5740), 1559-1563.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP & Workman JL (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell,* **123** (4), 581-592.

Caviston JP, Ross JL, Antony SM, Tokito M & Holzbaur EL (2007) Huntingtin facilitates dynein/dynactin-mediated vesicle transport. *Proc Natl Acad Sci U S A,* **104** (24), 10045-10050.

Cazalla D, Zhu J, Manche L, Huber E, Krainer AR & Caceres JF (2002) Nuclear export and retention signals in the RS domain of SR proteins. *Mol Cell Biol,* **22** (19), 6871-6882.

Chait BT (2011) Mass spectrometry in the postgenomic era. *Annu Rev Biochem,* **80**, 239-246.

Chang T & Abraham CR (1996) A novel brain cysteine protease forms an SDS stable complex with the β-amyloid precursor protein. *Ann N Y Acad Sci,* **777**, 183-188.

Chantalat S, Depaux A, Hery P, Barral S, Thuret JY, Dimitrov S & Gerard M (2011) Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Res,* **21** (9), 1426-1437.

Chen ES, Zhang K, Nicolas E, Cam HP, Zofall M & Grewal SI (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature,* **451** (7179), 734-737.

Chen TA & Allfrey VG (1987) Rapid and reversible changes in nucleosome structure accompany the activation, repression, and superinduction of murine fibroblast protooncogenes c-*fos* and c-*myc*. *Proc Natl Acad Sci U S A,* **84** (15), 5252-5256.

Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, Hughes TR & Winston F (2008) Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. *PLoS Biol,* **6** (11), e277.

Cho EJ, Kobor MS, Kim M, Greenblatt J & Buratowski S (2001) Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev,* **15** (24), 3319-3329.

Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Kramer U, Merchant SS, Zhang X, Jacobsen SE & Pellegrini M (2010) Relationship between nucleosome positioning and DNA methylation. *Nature,* **466** (7304), 388-392.

Choi JK (2010) Contrasting chromatin organization of CpG islands and exons in the human genome. *Genome Biol,* **11** (7), R70.

Chomczynski P & Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem,* **162** (1), 156-159.

Christy BA, Lau LF & Nathans D (1988) A gene activated in mouse 3T3 cells by serum growth factors encodes a protein with "zinc finger" sequences. *Proc Natl Acad Sci U S A,* **85** (21), 7857-7861.

Chu Y, Simic R, Warner MH, Arndt KM & Prelich G (2007) Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. *EMBO J,* **26** (22), 4646-4656.

Chu Y, Sutton A, Sternglanz R & Prelich G (2006) The Bur1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by Set2. *Mol Cell Biol,* **26** (8), 3029-3038.

Chuikov S, Kurash JK, Wilson JR, Xiao B, Justin N, Ivanov GS, McKinney K, Tempst P, Prives C, Gamblin SJ, Barlev NA & Reinberg D (2004) Regulation of p53 activity through lysine methylation. *Nature,* **432** (7015), 353-360.

Church GM & Gilbert W (1984) Genomic sequencing. *Proc Natl Acad Sci U S A,* **81** (7), 1991-1995.

Churchman LS & Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature,* **469** (7330), 368-373.

Clapier CR & Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem,* **78**, 273-304.

Clayton AL, Hazzalin CA & Mahadevan LC (2006) Enhanced histone acetylation and transcription: a dynamic perspective. *Mol Cell,* **23** (3), 289-296.

Clayton AL, Rose S, Barratt MJ & Mahadevan LC (2000) Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. *EMBO J,* **19** (14), 3714-3726.

Colin C, Tobaruella FS, Correa RG, Sogayar MC & Demasi MA (2010) Cloning and characterization of a novel alternatively spliced transcript of the human CHD7 putative helicase. *BMC Res Notes,* **3**, 252.

Cross SH & Bird AP (1995) CpG islands and genes. *Curr Opin Genet Dev,* **5** (3), 309-314.

Cui X, De Vivo I, Slany R, Miyamoto A, Firestein R & Cleary ML (1998) Association of SET domain and myotubularin-related proteins modulates growth control. *Nat Genet,* **18** (4), 331-337.

Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, Teague J, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Forbes S, Jia M, Jones D, Knott H, Kok CY, Lau KW, Leroy C, Lin ML, McBride DJ, Maddison M, Maguire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, O'Meara S, Pleasance E, Rajasingham A, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turrell K, Dykema KJ, Khoo SK, Petillo D, Wondergem B, Anema J, Kahnoski RJ, Teh BT, Stratton MR & Futreal PA (2010) Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature,* **463** (7279), 360-363.

Das M, Harvey I, Chu LL, Sinha M & Pelletier J (2001) Full-length cDNAs: more than just reaching the ends. *Physiol Genomics,* **6** (2), 57-80.

Davies SW, Turmaine M, Cozens BA, DiFiglia M, Sharp AH, Ross CA, Scherzinger E, Wanker EE, Mangiarini L & Bates GP (1997) Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell,* **90** (3), 537-548.

de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, Levezinho H, Gut M, Eick D, Gut I, Andrau JC, Ferrier P & Carmo-Fonseca M (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat Struct Mol Biol,* **18** (9), 977-983.

de Chassey B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaugue S, Meiffren G, Pradezynski F, Faria BF, Chantier T, Le Breton M, Pellet J, Davoust N, Mangeot PE, Chaboud A, Penin F, Jacob Y, Vidalain PO, Vidal M, Andre P, Rabourdin-Combe C & Lotteau V (2008) Hepatitis C virus infection protein network. *Mol Syst Biol,* **4**, 230.

De Vivo I, Cui X, Domen J & Cleary ML (1998) Growth stimulation of primary B cell precursors by the anti-phosphatase Sbf1. *Proc Natl Acad Sci U S A,* **95** (16), 9471-9476.

Deaton AM & Bird A (2011) CpG islands and the regulation of transcription. *Genes Dev,* **25** (10), 1010-1022.

Dephoure N, Zhou C, Villen J, Beausoleil SA, Bakalarski CE, Elledge SJ & Gygi SP (2008) A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A,* **105** (31), 10762-10767.

Dettmann A, Jaschke Y, Triebel I, Bogs J, Schroder I & Schuller HJ (2010) Mediator subunits and histone methyltransferase Set2 contribute to Ino2-dependent transcriptional activation of phospholipid biosynthesis in the yeast *Saccharomyces cerevisiae*. *Mol Genet Genomics,* **283** (3), 211-221.

Dhami P, Saffrey P, Bruce AW, Dillon SC, Chiang K, Bonhoure N, Koch CM, Bye J, James K, Foad NS, Ellis P, Watkins NA, Ouwehand WH, Langford C, Andrews RM, Dunham I & Vetrie D (2010) Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. *PLoS One,* **5** (8), e12339.

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S & Jeltsch A (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem,* **285** (34), 26114-26120.

Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, Magen A, Canidio E, Pagani M, Peluso I, Lin-Marq N, Koch M, Bilio M, Cantiello I, Verde R, De Masi C, Bianchi SA, Cicchini J, Perroud E, Mehmeti S, Dagand E, Schrinner S, Nurnberger A, Schmidt K, Metz K, Zwingmann C, Brieske N, Springer C, Hernandez AM, Herzog S, Grabbe F, Sieverding C, Fischer B, Schrader K, Brockmeyer M, Dettmer S, Helbig C, Alunni V, Battaini MA, Mura C, Henrichsen CN, Garcia-Lopez R, Echevarria D, Puelles E, Garcia-Calero E, Kruse S, Uhr M, Kauck C, Feng G, Milyaev N, Ong CK, Kumar L, Lam M, Semple CA, Gyenesei A, Mundlos S, Radelof U, Lehrach H, Sarmientos P, Reymond A, Davidson DR, Dolle P, Antonarakis SE, Yaspo ML, Martinez S, Baldock RA, Eichele G & Ballabio A (2011) A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol,* **9** (1), e1000582.

DiFiglia M, Sapp E, Chase K, Schwarz C, Meloni A, Young C, Martin E, Vonsattel JP, Carraway R, Reeves SA & et al. (1995) Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. *Neuron,* **14** (5), 1075-1081.

Dillon SC, Zhang X, Trievel RC & Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol,* **6** (8), 227.

Drouin S, Laramee L, Jacques PE, Forest A, Bergeron M & Robert F (2010) DSIF and RNA Polymerase II CTD Phosphorylation Coordinate the Recruitment of Rpd3S to Actively Transcribed Genes. *PLoS Genet,* **6** (10), e1001173.

Du HN & Briggs SD (2010) A nucleosome surface formed by histone H4, H2A, and H3 residues is needed for proper histone H3 Lys36 methylation, histone acetylation, and repression of cryptic transcription. *J Biol Chem,* **285** (15), 11704-11713.

Du HN, Fingerman IM & Briggs SD (2008) Histone H3 K36 methylation is mediated by a *trans*-histone methylation pathway involving an interaction between Set2 and histone H4. *Genes Dev,* **22** (20), 2786-2798.

Duns G, van den Berg E, van Duivenbode I, Osinga J, Hollema H, Hofstra RM & Kok K (2010) Histone methyltransferase gene *SETD2* is a novel tumor suppressor gene in clear cell renal cell carcinoma. *Cancer Res,* **70** (11), 4287-4291.

Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, Folstein S, Ross C, Franz M, Abbott M, Gray J, Conneally P, Young A, Penney J, Hollingsworth Z, Shoulson I, Lazzarini A, Falek A, Koroshetz W, Sax D, Bird E, Vonsattel J, Bonilla E, Alvir J, Bickham-Conde J, Cha J-H, Dure L, Gomez F, Ramos M, Sanchez-Ramos J, Snodgrass S, de Young M, Wexler N, Moscowitz C, Penchaszadeh G, MacFarlane H, Anderson M, Jenkins B, Srinidhi J, Barnes G, Gusella J & MacDonald M (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet,* **4** (4), 387-392.

Duyao MP, Auerbach AB, Ryan A, Persichetti F, Barnes GT, McNeil SM, Ge P, Vonsattel JP, Gusella JF, Joyner AL & et al. (1995) Inactivation of the mouse Huntington's disease gene homolog *Hdh*. *Science,* **269** (5222), 407-410.

Dwyer N & Blobel G (1976) A modified procedure for the isolation of a pore complex-lamina fraction from rat liver nuclei. *J Cell Biol,* **70** (3), 581-591.

Edmunds JW, Mahadevan LC & Clayton AL (2008) Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J,* **27** (2), 406-420.

Edwards DR & Mahadevan LC (1992) Protein synthesis inhibitors differentially superinduce c-*fos* and c-*jun* by three distinct mechanisms: lack of evidence for labile repressors. *EMBO J,* **11** (7), 2415-2424.

Egelhofer TA, Minoda A, Klugman S, Lee K, Kolasinska-Zwierz P, Alekseyenko AA, Cheung MS, Day DS, Gadel S, Gorchakov AA, Gu T, Kharchenko PV, Kuan S, Latorre I, Linder-Basso D, Luu Y, Ngo Q, Perry M,

Rechtsteiner A, Riddle NC, Schwartz YB, Shanower GA, Vielle A, Ahringer J, Elgin SC, Kuroda MI, Pirrotta V, Ren B, Strome S, Park PJ, Karpen GH, Hawkins RD & Lieb JD (2011) An assessment of histone-modification antibody quality. *Nat Struct Mol Biol,* **18** (1), 91-93.

Egloff S, Al-Rawaf H, O'Reilly D & Murphy S (2009) Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. *Mol Cell Biol,* **29** (14), 4002-4013.

Endoh M, Zhu W, Hasegawa J, Watanabe H, Kim DK, Aida M, Inukai N, Narita T, Yamada T, Furuya A, Sato H, Yamaguchi Y, Mandal SS, Reinberg D, Wada T & Handa H (2004) Human Spt6 stimulates transcription elongation by RNA polymerase II *in vitro*. *Mol Cell Biol,* **24** (8), 3324-3336.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M & Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* **473** (7345), 43-49.

Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultchi AI, Wutz A & Bickmore WA (2010) Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell,* **38** (3), 452-464.

Eyras E, Caccamo M, Curwen V & Clamp M (2004) ESTGenes: alternative splicing from ESTs in Ensembl. *Genome Res,* **14** (5), 976-987.

Faber PW, Barnes GT, Srinidhi J, Chen J, Gusella JF & MacDonald ME (1998) Huntingtin interacts with a family of WW domain proteins. *Hum Mol Genet,* **7** (9), 1463-1474.

Fang HY, Chang CL, Hsu SH, Huang CY, Chiang SF, Chiou SH, Huang CH, Hsiao YT, Lin TY, Chiang IP, Hsu WH, Sugano S, Chen CY, Lin CY, Ko WJ & Chow KC (2010) ATPase family AAA domain-containing 3A is a novel anti-apoptotic factor in lung adenocarcinoma cells. *J Cell Sci,* **123** (7), 1171-1180.

Farah CA, Liazoghli D, Perreault S, Desjardins M, Guimont A, Anton A, Lauzon M, Kreibich G, Paiement J & Leclerc N (2005) Interaction of microtubule-associated protein-2 and p63: a new link between microtubules and rough endoplasmic reticulum membranes in neurons. *J Biol Chem,* **280** (10), 9439-9449.

Feng Z, Jin S, Zupnick A, Hoh J, de Stanchina E, Lowe S, Prives C & Levine AJ (2006) p53 tumor suppressor protein regulates the levels of huntingtin gene expression. *Oncogene,* **25** (1), 1-7.

Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ & van Steensel B (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell,* **143** (2), 212-224.

Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD & Khorasanizadeh S (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev,* **17** (15), 1870-1881.

Fiumara F, Fioriti L, Kandel ER & Hendrickson WA (2010) Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell,* **143** (7), 1121-1135.

Flanagan JM (2007) Host epigenetic modifications by oncogenic viruses. *Br J Cancer,* **96** (2), 183-188.

Foster LJ, De Hoog CL & Mann M (2003) Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. *Proc Natl Acad Sci U S A,* **100** (10), 5813-5818.

Frescas D, Guardavaccaro D, Kuchay SM, Kato H, Poleshko A, Basrur V, Elenitoba-Johnson KS, Katz RA & Pagano M (2008) KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. *Cell Cycle,* **7** (22), 3539-3547.

Frohman MA, Dush MK & Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A,* **85** (23), 8998-9002.

Fu J, Yang Z, Wei J, Han J & Gu J (2006) Nuclear protein NP60 regulates p38 MAPK activity. *J Cell Sci,* **119** (1), 115-123.

Fujiki R, Chikanishi T, Hashiba W, Ito H, Takada I, Roeder RG, Kitagawa H & Kato S (2009) GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. *Nature,* **459** (7245), 455-459.

Furuhashi H, Takasaki T, Rechtsteiner A, Li T, Kimura H, Checchi PM, Strome S & Kelly WG (2010) Trans-generational epigenetic regulation of *C. elegans* primordial germ cells. *Epigenetics Chromatin,* **3** (1), 15.

Garcia A, Prabhakar S, Brock CJ, Pearce AC, Dwek RA, Watson SP, Hebestreit HF & Zitzmann N (2004) Extensive analysis of the human platelet proteome by two-dimensional gel electrophoresis and mass spectrometry. *Proteomics,* **4** (3), 656-668.

Garcia BA, Thomas CE, Kelleher NL & Mizzen CA (2008) Tissue-specific expression and post-translational modification of histone H3 variants. *J Proteome Res,* **7** (10), 4225-4236.

Gardner KE, Allis CD & Strahl BD (2011) Operating on chromatin, a colorful language where context matters. *J Mol Biol,* **409** (1), 36-46.

Garlisi CG, Uss AS, Xiao H, Tian F, Sheridan KE, Wang L, Motasim Billah M, Egan RW, Stranick KS & Umland SP (2001) A unique mRNA initiated within a middle intron of *WHSC1/MMSET* encodes a DNA binding protein that suppresses human *IL-5* transcription. *Am J Respir Cell Mol Biol,* **24** (1), 90-98.

Gauci S, Helbig AO, Slijper M, Krijgsveld J, Heck AJ & Mohammed S (2009) Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal Chem,* **81** (11), 4493-4501.

Gerashchenko MV, Su D & Gladyshev VN (2010) CUG start codon generates thioredoxin/glutathione reductase isoforms in mouse testes. *J Biol Chem,* **285** (7), 4595-4602.

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorrakrai K, Agarwal A, Alexander RP, Barber G, Brdlik CM, Brennan J, Brouillet JJ, Carr A, Cheung MS, Clawson H, Contrino S, Dannenberg LO, Dernburg AF, Desai A, Dick L, Dose AC, Du J, Egelhofer T, Ercan S, Euskirchen G, Ewing B, Feingold EA, Gassmann R, Good PJ, Green P, Gullier F, Gutwein M, Guyer MS, Habegger L, Han T, Henikoff JG, Henz SR, Hinrichs A, Holster H, Hyman T, Iniguez AL, Janette J, Jensen M, Kato M, Kent WJ, Kephart E, Khivansara V, Khurana E, Kim JK, Kolasinska-Zwierz P, Lai EC, Latorre I, Leahey A, Lewis S, Lloyd P, Lochovsky L, Lowdon RF, Lubling Y, Lyne R, MacCoss M, Mackowiak SD, Mangone M, McKay S, Mecenas D, Merrihew G, Miller DM, Muroyama A, Murray JI, Ooi SL, Pham H, Phippen T, Preston EA, Rajewsky N, Ratsch G, Rosenbaum H, Rozowsky J, Rutherford K, Ruzanov P, Sarov M, Sasidharan R, Sboner A, Scheid P, Segal E, Shin H, Shou C, Slack FJ, Slightam C, Smith R, Spencer WC, Stinson EO, Taing S, Takasaki T, Vafeados D, Voronina K, Wang G, Washington NL, Whittle CM, Wu B, Yan KK, Zeller G, Zha Z, Zhong M, Zhou X, Ahringer J, Strome S, Gunsalus KC, Micklem G, Liu XS, Reinke V, Kim SK, Hillier LW, Henikoff S, Piano F, Snyder M, Stein L, Lieb JD & Waterston RH (2010) Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science,* **330** (6012), 1775-1787.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK & Weissman JS (2003) Global analysis of protein expression in yeast. *Nature,* **425** (6959), 737-741.

Glover CV (1982) Heat shock induces rapid dephosphorylation of a ribosomal protein in *Drosophila*. *Proc Natl Acad Sci U S A,* **79** (6), 1781-1785.

Gobeil S, Zhu X, Doillon CJ & Green MR (2008) A genome-wide shRNA screen identifies *GAS1* as a novel melanoma metastasis suppressor gene. *Genes Dev,* **22** (21), 2932-2940.

Godovac-Zimmermann J, Kleiner O, Brown LR & Drukier AK (2005) Perspectives in spicing up proteomics with splicing. *Proteomics,* **5** (3), 699-709.

Goehler H, Lalowski M, Stelzl U, Waelter S, Stroedicke M, Worm U, Droege A, Lindenberg KS, Knoblich M, Haenig C, Herbst M, Suopanki J, Scherzinger E, Abraham C, Bauer B, Hasenbank R, Fritzsche A, Ludewig AH, Bussow K, Coleman SH, Gutekunst CA, Landwehrmeyer BG, Lehrach H & Wanker EE (2004) A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol Cell,* **15** (6), 853-865.

Goulet I, Gauvin G, Boisvenue S & Cote J (2007) Alternative splicing yields protein arginine methyltransferase 1 isoforms with distinct activity, substrate specificity, and subcellular localization. *J Biol Chem,* **282** (45), 33009-33021.

Govind CK, Qiu H, Ginsburg DS, Ruan C, Hofmeyer K, Hu C, Swaminathan V, Workman JL, Li B & Hinnebusch AG (2010) Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. *Mol Cell,* **39** (2), 234-246.

Gozani O, Patton JG & Reed R (1994) A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. *EMBO J,* **13** (14), 3356-3367.

Graveley BR (2000) Sorting out the complexity of SR protein functions. *RNA,* **6** (9), 1197-1211.

Guenther MG, Levine SS, Boyer LA, Jaenisch R & Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell,* **130** (1), 77-88.

Guerrera IC & Kleiner O (2005) Application of mass spectrometry in proteomics. *Biosci Rep,* **25** (1-2), 71-93.

Gutekunst CA, Levey AI, Heilman CJ, Whaley WL, Yi H, Nash NR, Rees HD, Madden JJ & Hersch SM (1995) Identification and localization of huntingtin in brain and human lymphoblastoid cell lines with anti-fusion protein antibodies. *Proc Natl Acad Sci U S A,* **92** (19), 8710-8714.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL & Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature,* **458** (7235), 223-227.

Haan C & Behrmann I (2007) A cost effective non-commercial ECL-solution for Western blot detections yielding strong signals and low background. *J Immunol Methods,* **318** (1-2), 11-19.

Hagiwara M & Nojima T (2007) Cross-talks between transcription and post-transcriptional events within a 'mRNA factory'. *J Biochem,* **142** (1), 11-15.

Hahn MA, Wu X, Li AX, Hahn T & Pfeifer GP (2011) Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PLoS One,* **6** (4), e18844.

Hainer SJ & Martens JA (2011) Identification of histone mutants that are defective for transcription-coupled nucleosome occupancy. *Mol Cell Biol,* **31** (17), 3557-3568.

Hainer SJ, Pruneski JA, Mitchell RD, Monteverde RM & Martens JA (2010) Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev,* **25** (1), 29-40.

Hamatani T, Carter MG, Sharov AA & Ko MS (2004) Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell,* **6** (1), 117-131.

Hancock R (2000) A new look at the nuclear matrix. *Chromosoma,* **109** (4), 219-225.

Hannon GJ (ed.) (2003) *RNAi: A Guide to Gene Silencing,* Cold Spring Harbour Laboratory Press.

Hansen JC, Tse C & Wolffe AP (1998a) Structure and function of the core histone N-termini: more than meets the eye. *Biochemistry,* **37** (51), 17637-17641.

Hansen S, Lane DP & Midgley CA (1998b) The N terminus of the murine p53 tumour suppressor is an independent regulatory domain affecting activation and thermostability. *J Mol Biol,* **275** (4), 575-588.

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson JA, Ecker JR & Ren B (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell,* **6** (5), 479-491.

Hazzalin CA, Cuenda A, Cano E, Cohen P & Mahadevan LC (1997) Effects of the inhibition of p38/RK MAP kinase on induction of five *fos* and *jun* genes by diverse stimuli. *Oncogene,* **15** (19), 2321-2331.

Hazzalin CA & Mahadevan LC (2002) MAPK-regulated transcription: a continuously variable gene switch? *Nat Rev Mol Cell Biol,* **3** (1), 30-40.

Hazzalin CA & Mahadevan LC (2005) Dynamic acetylation of all lysine 4-methylated histone H3 in the mouse nucleus: analysis at c-*fos* and c-*jun*. *PLoS biology,* **3** (12), e393.

He J, Mao CC, Reyes A, Sembongi H, Di Re M, Granycome C, Clippingdale AB, Fearnley IM, Harbour M, Robinson AJ, Reichelt S, Spelbrink JN, Walker JE & Holt IJ (2007) The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization. *J Cell Biol,* **176** (2), 141-146.

Hebbes TR, Thorne AW & Crane-Robinson C (1988) A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J,* **7** (5), 1395-1402.

Heber S, Alekseyev M, Sze SH, Tang H & Pevzner PA (2002) Splicing graphs and EST assembly problem. *Bioinformatics,* **18** (Suppl 1), S181-188.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M & Ren B (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature,* **459** (7243), 108-112.

Hellen CU & Sarnow P (2001) Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev,* **15** (13), 1593-1612.

Hilditch-Maguire P, Trettel F, Passani LA, Auerbach A, Persichetti F & MacDonald ME (2000) Huntingtin: an iron-regulated protein essential for normal nuclear and perinuclear organelles. *Hum Mol Genet,* **9** (19), 2789-2797.

Hill MM, Bastiani M, Luetterforst R, Kirkham M, Kirkham A, Nixon SJ, Walser P, Abankwa D, Oorschot VM, Martin S, Hancock JF & Parton RG (2008) PTRF-Cavin, a conserved cytoplasmic protein required for caveola formation and function. *Cell,* **132** (1), 113-124.

Hoffner G, Kahlem P & Djian P (2002) Perinuclear localization of huntingtin as a consequence of its binding to microtubules through an interaction with beta-tubulin: relevance to Huntington's disease. *J Cell Sci,* **115** (Pt 5), 941-948.

Hollstein M, Sidransky D, Vogelstein B & Harris CC (1991) p53 mutations in human cancers. *Science,* **253** (5015), 49-53.

Hon G, Wang W & Ren B (2009) Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput Biol,* **5** (11), e1000566.

Houseley J, Rubbi L, Grunstein M, Tollervey D & Vogelauer M (2008) A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Mol Cell,* **32** (5), 685-695.

Hu H, Columbus J, Zhang Y, Wu D, Lian L, Yang S, Goodwin J, Luczak C, Carter M, Chen L, James M, Davis R, Sudol M, Rodwell J & Herrero JJ (2004) A map of WW domain family interactions. *Proteomics,* **4** (3), 643-655.

Hu M, Sun XJ, Zhang YL, Kuang Y, Hu CQ, Wu WL, Shen SH, Du TT, Li H, He F, Xiao HS, Wang ZG, Liu TX, Lu H, Huang QH, Chen SJ & Chen Z (2010) Histone H3 lysine 36 methyltransferase *Hypb*/*Setd2* is required for embryonic vascular remodeling. *Proc Natl Acad Sci U S A,* **107** (7), 2956-2961.

Huang J, Dorsey J, Chuikov S, Perez-Burgos L, Zhang X, Jenuwein T, Reinberg D & Berger SL (2010) G9a and Glp methylate lysine 373 in the tumor suppressor p53. *J Biol Chem,* **285** (13), 9636-9641.

Huang J, Perez-Burgos L, Placek BJ, Sengupta R, Richter M, Dorsey JA, Kubicek S, Opravil S, Jenuwein T & Berger SL (2006) Repression of p53 activity by Smyd2-mediated methylation. *Nature,* **444** (7119), 629-632.

Huff JT, Plocik AM, Guthrie C & Yamamoto KR (2010) Reciprocal intronic and exonic histone modification regions in humans. *Nat Struct Mol Biol,* **17** (12), 1495-1499.

Hwang WW & Madhani HD (2009) Nonredundant requirement for multiple histone modifications for the early anaphase release of the mitotic exit regulator Cdc14 from nucleolar chromatin. *PLoS Genet,* **5** (8), e1000588.

Illuzzi JL, Vickers CA & Kmiec EB (2011) Modifications of p53 and the DNA Damage Response in Cells Expressing Mutant Form of the Protein Huntingtin. *J Mol Neurosci,* **45** (2), 256-268.

Ivanov IP, Firth AE, Michel AM, Atkins JF & Baranov PV (2011) Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res,* **39** (10), 4220-4234.

Jaehning JA (2010) The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta,* **1799** (5-6), 379-388.

Jansa P, Burek C, Sander EE & Grummt I (2001) The transcript release factor PTRF augments ribosomal gene transcription by facilitating reinitiation of RNA polymerase I. *Nucleic Acids Res,* **29** (2), 423-429.

Jansa P, Mason SW, Hoffmann-Rohrer U & Grummt I (1998) Cloning and functional characterization of PTRF, a novel protein which induces dissociation of paused ternary transcription complexes. *EMBO J,* **17** (10), 2855-2864.

Jelinic P, Pellegrino J & David G (2011) A novel mammalian complex containing Sin3B mitigates histone acetylation and RNAPII progression within transcribed loci. *Mol Cell Biol,* **31** (1), 54-62.

Jia G, Wang W, Li H, Mao Z, Cai G, Sun J, Wu H, Xu M, Yang P, Yuan W, Chen S & Zhu B (2009) A systematic evaluation of the compatibility of histones containing methyl-lysine analogues with biochemical reactions. *Cell Res,* **19** (10), 1217-1220.

Jin S & Levine AJ (2001) The p53 functional circuit. *J Cell Sci,* **114** (Pt 23), 4139-4140.

Jonson L, Vikesaa J, Krogh A, Nielsen LK, Hansen T, Borup R, Johnsen AH, Christiansen J & Nielsen FC (2007) Molecular composition of IMP1 ribonucleoprotein granules. *Mol Cell Proteomics,* **6** (5), 798-811.

Joshi AA & Struhl K (2005) Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell,* **20** (6), 971-978.

Kaltenbach LS, Romero E, Becklin RR, Chettier R, Bell R, Phansalkar A, Strand A, Torcassi C, Savage J, Hurlburt A, Cha GH, Ukani L, Chepanoske CL, Zhen Y, Sahasrabudhe S, Olson J, Kurschner C, Ellerby LM, Peltier JM, Botas J & Hughes RE (2007) Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS Genet,* **3** (5), e82.

Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F & Reid G (2008) Transient cyclical methylation of promoter DNA. *Nature,* **452** (7183), 112-115.

Kaplan CD, Laprade L & Winston F (2003) Transcription elongation factors repress transcription initiation from cryptic sites. *Science,* **301** (5636), 1096-1099.

Karlovich CA, John RM, Ramirez L, Stainier DY & Myers RM (1998) Characterization of the Huntington's disease (HD) gene homologue in the zebrafish *Danio rerio*. *Gene,* **217** (1-2), 117-125.

Kato Y, Nagata K, Takahashi M, Lian L, Herrero JJ, Sudol M & Tanokura M (2004) Common mechanism of ligand recognition by group II/III WW domains: redefining their functional classification. *J Biol Chem,* **279** (30), 31833-31841.

Kegel KB, Meloni AR, Yi Y, Kim YJ, Doyle E, Cuiffo BG, Sapp E, Wang Y, Qin ZH, Chen JD, Nevins JR, Aronin N & DiFiglia M (2002) Huntingtin is present in the nucleus, interacts with the transcriptional corepressor C-terminal binding protein, and represses transcription. *J Biol Chem,* **277** (9), 7466-7476.

Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ, Boone C, Emili A, Weissman JS, Hughes TR, Strahl BD, Grunstein M, Greenblatt JF, Buratowski S & Krogan NJ (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell,* **123** (4), 593-605.

Keren H, Lev-Maor G & Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet,* **11** (5), 345-355.

Kim A, Kiefer CM & Dean A (2007) Distinctive signatures of histone methylation in transcribed coding and noncoding human beta-globin sequences. *Mol Cell Biol,* **27** (4), 1271-1279.

Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, Megee PC & Bentley DL (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol,* **17** (10), 1279-1286.

Kim JY, Kee HJ, Choe NW, Kim SM, Eom GH, Baek HJ, Kook H & Seo SB (2008) Multiple-myeloma-related WHSC1/MMSET isoform RE-IIBP is a histone methyltransferase with transcriptional repression activity. *Mol Cell Biol,* **28** (6), 2023-2034.

Kim S, Kim H, Fong N, Erickson B & Bentley DL (2011) Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc Natl Acad Sci U S A,* **108** (33), 13564-13569.

Kim SM, Kee HJ, Eom GH, Choe NW, Kim JY, Kim YS, Kim SK, Kook H, Kook H & Seo SB (2006) Characterization of a novel WHSC1-associated SET domain protein with H3K4 and H3K27 methyltransferase activity. *Biochem Biophys Res Commun,* **345** (1), 318-323.

Kim YK, Furic L, Desgroseillers L & Maquat LE (2005) Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell,* **120** (2), 195-208.

Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL & Strahl BD (2005) A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol Cell Biol,* **25** (8), 3305-3316.

Klenova EM, Nicolas RH, U S, Carne AF, Lee RE, Lobanenkov VV & Goodwin GH (1997) Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in SDS-PAGE and the size of the expressed protein is affected by the UTRs and sequences within the coding region of the CTCF gene. *Nucleic Acids Res,* **25** (3), 466-474.

Klopfenstein DR, Kappeler F & Hauri HP (1998) A novel direct interaction of endoplasmic reticulum with microtubules. *Embo J,* **17** (21), 6168-6177.

Klose RJ & Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci,* **31** (2), 89-97.

Klose RJ, Gardner KE, Liang G, Erdjument-Bromage H, Tempst P & Zhang Y (2007) Demethylation of histone H3K36 and H3K9 by Rph1: a vestige of an H3K9 methylation system in *Saccharomyces cerevisiae*? *Mol Cell Biol,* **27** (11), 3951-3961.

Kobayashi R, Patenia R, Ashizawa S & Vykoukal J (2009) Targeted mass spectrometric analysis of N-terminally truncated isoforms generated via alternative translation initiation. *FEBS Lett,* **583** (14), 2441-2445.

Koch CM, Andrews RM, Flicek P, Dillon SC, Karaoz U, Clelland GK, Wilcox S, Beare DM, Fowler JC, Couttet P, James KD, Lefebvre GC, Bruce AW, Dovey OM, Ellis PD, Dhami P, Langford CF, Weng Z, Birney E, Carter NP, Vetrie D & Dunham I (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res,* **17** (6), 691-707.

Kochetov AV (2008) Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. *Bioessays,* **30** (7), 683-691.

Kochetov AV, Ahmad S, Ivanisenko V, Volkova OA, Kolchanov NA & Sarai A (2008) uORFs, reinitiation and alternative translation start sites in human mRNAs. *FEBS Lett,* **582** (9), 1293-1297.

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS & Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet,* **41** (3), 376-381.

Kolodziejski PJ, Rashid MB & Eissa NT (2003) Intracellular formation of "undisruptable" dimers of inducible nitric oxide synthase. *Proc Natl Acad Sci U S A,* **100** (24), 14263-14268.

Komarnitsky P, Cho EJ & Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev,* **14** (19), 2452-2460.

Kontaki H & Talianidis I (2010) Lysine methylation regulates E2F1-induced cell death. *Mol Cell,* **39** (1), 152-160.

Kouskouti A, Scheer E, Staub A, Tora L & Talianidis I (2004) Gene-specific modulation of TAF10 function by SET9-mediated methylation. *Mol Cell,* **14** (2), 175-182.

Kouzarides T (2007) Chromatin modifications and their function. *Cell,* **128** (4), 693-705.

Kowit JD & Maloney J (1982) Protein cleavage by boiling in sodium dodecyl sulfate prior to electrophoresis. *Anal Biochem,* **123** (1), 86-93.

Krajewski WA, Nakamura T, Mazo A & Canaani E (2005) A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Mol Cell Biol,* **25** (5), 1891-1899.

Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, Richards DP, Beattie BK, Emili A, Boone C, Shilatifard A, Buratowski S & Greenblatt J (2003) Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol,* **23** (12), 4207-4218.

Kurotaki N, Harada N, Yoshiura K, Sugano S, Niikawa N & Matsumoto N (2001) Molecular characterization of *NSD1*, a human homologue of the mouse *Nsd1* gene. *Gene,* **279** (2), 197-204.

Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Hasegawa T, Chinen Y, Tomita Ha HA, Kinoshita A, Mizuguchi T, Yoshiura Ki K, Ohta T, Kishino T, Fukushima Y, Niikawa N & Matsumoto N (2002) Haploinsufficiency of *NSD1* causes Sotos syndrome. *Nat Genet,* **30** (4), 365-366.

Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P & Reinberg D (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev,* **16** (22), 2893-2905.

Lachner M, O'Carroll D, Rea S, Mechtler K & Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature,* **410** (6824), 116-120.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* **227** (259), 680-685.

Lam YC, Bowman AB, Jafar-Nejad P, Lim J, Richman R, Fryer JD, Hyun ED, Duvick LA, Orr HT, Botas J & Zoghbi HY (2006) ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell,* **127** (7), 1335-1347.

Lander ES (2011) Initial impact of the sequencing of the human genome. *Nature,* **470** (7333), 187-197.

Landry J, Sutton A, Hesman T, Min JR, Xu RM, Johnston M & Sternglanz R (2003) Set2-catalyzed methylation of histone H3 represses basal expression of *GAL4* in *Saccharomyces cerevisiae*. *Mol Cell Biol,* **23** (17), 5972-5978.

Lareau LF, Inada M, Green RE, Wengrod JC & Brenner SE (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature,* **446** (7138), 926-929.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ & Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics,* **23** (21), 2947-2948.

Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P, Workman JL, Park PJ & Kuroda MI (2007) MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell,* **28** (1), 121-133.

Laue K, Daujat S, Crump JG, Plaster N, Roehl HH, Kimmel CB, Schneider R & Hammerschmidt M (2008) The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity. *Development,* **135** (11), 1935-1946.

Lee BM & Mahadevan LC (2009) Stability of histone modifications across mammalian genomes: implications for 'epigenetic' marking. *J Cell Biochem,* **108** (1), 22-34.

Lee JS, Smith E & Shilatifard A (2010) The language of histone crosstalk. *Cell,* **142** (5), 682-685.

Lee SH, Oshige M, Durant ST, Rasila KK, Williamson EA, Ramsey H, Kwan L, Nickoloff JA & Hromas R (2005) The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. *Proc Natl Acad Sci U S A,* **102** (50), 18075-18080.

Li B, Carey M & Workman JL (2007a) The role of chromatin during transcription. *Cell,* **128** (4), 707-719.

Li B, Gogol M, Carey M, Lee D, Seidel C & Workman JL (2007b) Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science,* **316** (5827), 1050-1054.

Li B, Howe L, Anderson S, Yates JR, 3rd & Workman JL (2003) The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem,* **278** (11), 8897-8903.

Li B, Jackson J, Simon MD, Fleharty B, Gogol M, Seidel C, Workman JL & Shilatifard A (2009a) Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *J Biol Chem,* **284** (12), 7970-7976.

Li J, Moazed D & Gygi SP (2002) Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *J Biol Chem,* **277** (51), 49383-49388.

Li M, Phatnani HP, Greenleaf AL & Zhou P (2006) NMR assignment of the SRI domain of human Set2/HYPB. *Journal of Biomolecular NMR,* **36**, 5.

Li M, Phatnani HP, Guan Z, Sage H, Greenleaf AL & Zhou P (2005) Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proc Natl Acad Sci U S A,* **102** (49), 17636-17641.

Li SH & Li XJ (2004) Huntingtin-protein interactions and the pathogenesis of Huntington's disease. *Trends Genet,* **20** (3), 146-154.

Li Y, Trojer P, Xu CF, Cheung P, Kuo A, Drury WJ, 3rd, Qiao Q, Neubert TA, Xu RM, Gozani O & Reinberg D (2009b) The target of the NSD family of histone lysine methyl transferases depends on the nature of the substrate. *J Biol Chem,* **284** (49), 34283-34295.

Lian Z, Karpikov A, Lian J, Mahajan MC, Hartman S, Gerstein M, Snyder M & Weissman SM (2008) A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation. *Genome Res,* **18** (8), 1224-1237.

Lim J, Hao T, Shaw C, Patel AJ, Szabo G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE, Barabasi AL, Vidal M & Zoghbi HY (2006) A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell,* **125** (4), 801-814.

Lim WK, Rosgen J & Englander SW (2009) Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc Natl Acad Sci U S A,* **106** (8), 2595-2600.

Lin LJ, Minard LV, Johnston GC, Singer RA & Schultz MC (2010) Asf1 can promote trimethylation of H3 K36 by Set2. *Mol Cell Biol,* **30** (5), 1116-1129.

Liu L, Brown D, McKee M, Lebrasseur NK, Yang D, Albrecht KH, Ravid K & Pilch PF (2008) Deletion of *Cavin/PTRF* causes global loss of caveolae, dyslipidemia, and glucose intolerance. *Cell Metab,* **8** (4), 310-317.

Lopez AJ (1995) Developmental role of transcription factor isoforms generated by alternative splicing. *Dev Biol,* **172** (2), 396-411.

Lu T, Jackson MW, Wang B, Yang M, Chance MR, Miyagi M, Gudkov AV & Stark GR (2010) Regulation of NF-κB by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proc Natl Acad Sci U S A,* **107** (1), 46-51.

Lu X, Simon MD, Chodaparambil JV, Hansen JC, Shokat KM & Luger K (2008) The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat Struct Mol Biol,* **15** (10), 1122-1124.

Lucio-Eterovic AK, Singh MM, Gardner JE, Veerappan CS, Rice JC & Carpenter PB (2010) Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc Natl Acad Sci U S A,* **107** (39), 16952-16957.

Luco RF, Allo M, Schor IE, Kornblihtt AR & Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. *Cell,* **144** (1), 16-26.

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM & Misteli T (2010) Regulation of alternative splicing by histone modifications. *Science,* **327** (5968), 996-1000.

Luders J, Pyrowolakis G & Jentsch S (2003) The ubiquitin-like protein HUB1 forms SDS-resistant complexes with cellular proteins in the absence of ATP. *EMBO Rep,* **4** (12), 1169-1174.

Luger K, Mader AW, Richmond RK, Sargent DF & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature,* **389** (6648), 251-260.

Lumsden AL, Henshall TL, Dayan S, Lardelli MT & Richards RI (2007) Huntingtin-deficient zebrafish exhibit defects in iron utilization and development. *Hum Mol Genet,* **16** (16), 1905-1920.

MacConnell WP & Verma IM (1983) Expression of FBJ-MSV oncogene (*fos*) product in bacteria. *Virology,* **131** (2), 367-374.

Macdonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, Barnes G, Taylor SA, James M, Groot N, MacFarlane H, Jenkins B, Anderson M, Wexler NS & Gusella JF (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell,* **72** (6), 971-983.

Mahadevan LC, Heath JK, Leichtfried FE, Cumming DV, Hirst EM & Foulkes JG (1988) Rapid appearance of novel phosphoproteins in the nuclei of mitogen-stimulated fibroblasts. *Oncogene,* **2** (3), 249-257.

Mahadevan LC, Targett K & Heath JK (1989) 2-Aminopurine abolishes epidermal growth factor-stimulated phosphorylation of complexed and chromatin-associated forms of a 33 kDa phosphoprotein. *Oncogene,* **4** (6), 699-706.

Malik S, Chaurasia P, Lahudkar S, Durairaj G, Shukla A & Bhaumik SR (2010) Rad26p, a transcription-coupled repair factor, is recruited to the site of DNA lesion in an elongating RNA polymerase II-dependent manner *in vivo*. *Nucleic Acids Res,* **38** (5), 1461-1477.

Mann M & Jensen ON (2003) Proteomic analysis of post-translational modifications. *Nat Biotechnol,* **21** (3), 255-261.

Margariti A, Xiao Q, Zampetaki A, Zhang Z, Li H, Martin D, Hu Y, Zeng L & Xu Q (2009) Splicing of HDAC7 modulates the SRF-myocardin complex during stem-cell differentiation towards smooth muscle cells. *J Cell Sci,* **122** (Pt 4), 460-470.

Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ, 3rd, Voigt P, Martin SR, Taylor WR, De Marco V, Pirrotta V, Reinberg D & Gamblin SJ (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature,* **461** (7265), 762-767.

Martin DG, Baetz K, Shi X, Walter KL, MacDonald VE, Wlodarski MJ, Gozani O, Hieter P & Howe L (2006a) The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. *Mol Cell Biol,* **26** (21), 7871-7879.

Martin DG, Grimes DE, Baetz K & Howe L (2006b) Methylation of histone H3 mediates the association of the NuA3 histone acetyltransferase with chromatin. *Mol Cell Biol,* **26** (8), 3018-3028.

Martinez-Garcia E, Popovic R, Min DJ, Sweet SM, Thomas PM, Zamdborg L, Heffner A, Will C, Lamy L, Staudt LM, Levens DL, Kelleher NL & Licht JD (2011) The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood,* **117** (1), 211-220.

Marton LS & Stefansson K (1986) Exposure of binding sites for antibodies and concanavalin A on collagen by solubilization in hot urea. An immunoblot analysis. *J Immunol Methods,* **91** (2), 187-194.

Mathur M, Tucker PW & Samuels HH (2001) PSF is a novel corepressor that mediates its effect through Sin3A and the DNA binding domain of nuclear hormone receptors. *Mol Cell Biol,* **21** (7), 2298-2311.

Matzke MA & Birchler JA (2005) RNAi-mediated pathways in the nucleus. *Nat Rev Genet,* **6** (1), 24-35.

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T & Costello JF (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature,* **466** (7303), 253-257.

May P & May E (1999) Twenty years of p53 research: structural and functional aspects of the p53 protein. *Oncogene,* **18** (53), 7621-7636.

Mayya V, Lundgren DH, Hwang SI, Rezaul K, Wu L, Eng JK, Rodionov V & Han DK (2009) Quantitative phosphoproteomic analysis of T cell receptor signaling reveals system-wide modulation of protein-protein interactions. *Sci Signal,* **2** (84), ra46.

McCarthy J, Hopwood F, Oxley D, Laver M, Castagna A, Righetti PG, Williams K & Herbert B (2003) Carbamylation of proteins in 2-D electrophoresis—myth or reality? *J Proteome Res,* **2** (3), 239-242.

McIlwain DR, Pan Q, Reilly PT, Elia AJ, McCracken S, Wakeham AC, Itie-Youten A, Blencowe BJ & Mak TW (2010) Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. *Proc Natl Acad Sci U S A,* **107** (27), 12186-12191.

Melamud E & Moult J (2009a) Stochastic noise in splicing machinery. *Nucleic Acids Res,* **37** (14), 4873-4886.

Melamud E & Moult J (2009b) Structural implication of splicing stochastics. *Nucleic Acids Res,* **37** (14), 4862-4872.

Merker JD, Dominska M, Greenwell PW, Rinella E, Bouck DC, Shibata Y, Strahl BD, Mieczkowski P & Petes TD (2008) The histone methylase Set2p and the histone deacetylase Rpd3p repress meiotic recombination at the *HIS4* meiotic recombination hotspot in *Saccharomyces cerevisiae*. *DNA Repair,* **7** (8), 1298-1308.

Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F & Salbert G (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature,* **452** (7183), 45-50.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES & Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature,* **448** (7153), 553-560.

Moore MJ & Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell,* **136** (4), 688-700.

Morillon A, Karabetsou N, Nair A & Mellor J (2005) Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. *Mol Cell,* **18** (6), 723-734.

Morishita M & di Luccio E (2011) Cancers and the NSD family of histone lysine methyltransferases. *Biochim Biophys Acta,* **1816** (2), 158-163.

Morris SA, Shibata Y, Noma K, Tsukamoto Y, Warren E, Temple B, Grewal SI & Strahl BD (2005) Histone H3 K36 methylation is associated with transcription elongation in *Schizosaccharomyces pombe*. *Eukaryot Cell,* **4** (8), 1446-1454.

Muller WG, Rieder D, Karpova TS, John S, Trajanoski Z & McNally JG (2007) Organization of chromatin and histone modifications at a transcription site. *J Cell Biol,* **177** (6), 957-967.

Mumberg D, Lucibello FC, Schuermann M & Muller R (1991) Alternative splicing of *fosB* transcripts results in differentially expressed mRNAs encoding functionally antagonistic proteins. *Genes Dev,* **5** (7), 1212-1223.

Munro S, Khaire N, Inche A, Carr S & La Thangue NB (2010) Lysine methylation regulates the pRb tumour suppressor protein. *Oncogene,* **29** (16), 2357-2367.

Nahkuri S, Taft RJ & Mattick JS (2009) Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle,* **8** (20), 3420-3424.

Nakayama J, Rice JC, Strahl BD, Allis CD & Grewal SI (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science,* **292** (5514), 110-113.

Nasir J, Floresco SB, O'Kusky JR, Diewert VM, Richman JM, Zeisler J, Borowski A, Marth JD, Phillips AG & Hayden MR (1995) Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell,* **81** (5), 811-823.

Neckelmann N, Warner CK, Chung A, Kudoh J, Minoshima S, Fukuyama R, Maekawa M, Shimizu Y, Shimizu N, Liu JD & et al. (1989) The human ATP synthase beta subunit gene: sequence analysis, chromosome assignment, and differential expression. *Genomics,* **5** (4), 829-843.

Nelson CJ, Santos-Rosa H & Kouzarides T (2006) Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell,* **126** (5), 905-916.

Newbold RF & Mokbel K (2010) Evidence for a tumour suppressor function of SETD2 in human breast cancer: a new hypothesis. *Anticancer Res,* **30** (9), 3309-3311.

Newell-Price J, Clark AJ & King P (2000) DNA methylation and silencing of gene expression. *Trends Endocrinol Metab,* **11** (4), 142-148.

Ng HH, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y & Struhl K (2002) Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev,* **16** (12), 1518-1527.

Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE & Ares M, Jr. (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev,* **21** (6), 708-718.

Nikonov AV, Hauri HP, Lauring B & Kreibich G (2007) Climp-63-mediated binding of microtubules to the ER affects the lateral mobility of translocon complexes. *J Cell Sci,* **120** (13), 2248-2258.

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem,* **250** (10), 4007-4021.

Oler AJ, Alla RK, Roberts DN, Wong A, Hollenhorst PC, Chandler KJ, Cassiday PA, Nelson CA, Hagedorn CH, Graves BJ & Cairns BR (2010) Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nat Struct Mol Biol,* **17** (5), 620-628.

Orphanides G & Reinberg D (2002) A unified theory of gene expression. *Cell,* **108** (4), 439-451.

Ovcharenko I, Nobrega MA, Loots GG & Stubbs L (2004) ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res,* **32** (Web Server issue), W280-286.

Ozsolak F & Milos PM (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet,* **12** (2), 87-98.

Pan Q, Shai O, Lee LJ, Frey BJ & Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet,* **40** (12), 1413-1415.

Panyim S & Chalkley R (1969) High resolution acrylamide gel electrophoresis of histones. *Arch Biochem Biophys,* **130** (1), 337-346.

Parada CA & Roeder RG (1999) A novel RNA polymerase II-containing complex potentiates Tat-enhanced HIV-1 transcription. *EMBO J,* **18** (13), 3688-3701.

Passani LA, Bedford MT, Faber PW, McGinnis KM, Sharp AH, Gusella JF, Vonsattel JP & MacDonald ME (2000) Huntingtin's WW domain partners in Huntington's disease post-mortem brain fulfill genetic criteria for direct involvement in Huntington's disease pathogenesis. *Hum Mol Genet,* **9** (14), 2175-2182.

Pattenden SG, Gogol MM & Workman JL (2010) Features of cryptic promoters and their varied reliance on bromodomain-containing factors. *PLoS One,* **5** (9), e12927.

Paulsen M & Ferguson-Smith AC (2001) DNA methylation in genomic imprinting, development, and disease. *J Pathol,* **195** (1), 97-110.

Peters J, Baumeister W & Lupas A (1996) Hyperthermostable surface layer protein tetrabrachion from the archaebacterium *Staphylothermus marinus*: evidence for the presence of a right-handed coiled coil derived from the primary structure. *J Mol Biol,* **257** (5), 1031-1041.

Peters J, Nitsch M, Kuhlmorgen B, Golbik R, Lupas A, Kellermann J, Engelhardt H, Pfander JP, Muller S, Goldie K & et al. (1995) Tetrabrachion: a filamentous archaebacterial surface protein assembly of unusual structure and extreme stability. *J Mol Biol,* **245** (4), 385-401.

Petrie K, Guidez F, Howell L, Healy L, Waxman S, Greaves M & Zelent A (2003) The histone deacetylase 9 gene encodes multiple protein isoforms. *J Biol Chem,* **278** (18), 16059-16072.

Phatnani HP & Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev,* **20** (21), 2922-2936.

Phatnani HP, Jones JC & Greenleaf AL (2004) Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry,* **43** (50), 15702-15719.

Philipps D, Celotto AM, Wang QQ, Tarng RS & Graveley BR (2003) Arginine/serine repeats are sufficient to constitute a splicing activation domain. *Nucleic Acids Res,* **31** (22), 6502-6508.

Pirngruber J, Shchebet A & Johnsen SA (2009) Insights into the function of the human P-TEFb component CDK9 in the regulation of chromatin modifications and co-transcriptional mRNA processing. *Cell Cycle,* **8** (22), 3636-3642.

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK & Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell,* **122** (4), 517-527.

Pryde F, Jain D, Kerr A, Curley R, Mariotti FR & Vogelauer M (2009) H3 k36 methylation helps determine the timing of cdc45 association with replication origins. *PLoS One,* **4** (6), e5882.

Prymakowska-Bosak M, Misteli T, Herrera JE, Shirakawa H, Birger Y, Garfield S & Bustin M (2001) Mitotic phosphorylation prevents the binding of HMGN proteins to chromatin. *Mol Cell Biol,* **21** (15), 5169-5178.

Psathas JN, Zheng S, Tan S & Reese JC (2009) Set2-dependent K36 methylation is regulated by novel intra-tail interactions within H3. *Mol Cell Biol,* **29** (24), 6413-6426.

Qian C & Zhou MM (2006) SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell Mol Life Sci,* **63** (23), 2755-2763.

Qiao Q, Li Y, Chen Z, Wang M, Reinberg D & Xu RM (2011) The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J Biol Chem,* **286** (10), 8361-8368.

Rabilloud T, Chevallet M, Luche S & Lelong C (2010) Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J Proteomics,* **73** (11), 2064-2077.

Raha D, Wang Z, Moqtaderi Z, Wu L, Zhong G, Gerstein M, Struhl K & Snyder M (2010) Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proc Natl Acad Sci U S A,* **107** (8), 3639-3644.

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA & Young RA (2010) c-Myc regulates transcriptional pause release. *Cell,* **141** (3), 432-445.

Rampersaud N (2003) *Role of the nuclear substructure in immediate-early gene induction*. DPhil, University of Oxford.

Ramsby ML & Makowski GS (1999) Differential detergent fractionation of eukaryotic cells. Analysis by two-dimensional gel electrophoresis. *Methods Mol Biol,* **112**, 53-66.

Ran Q & Pereira-Smith OM (2000) Identification of an alternatively spliced form of the Tat interactive protein (Tip60), Tip60β. *Gene,* **258** (1-2), 141-146.

Rao B, Shibata Y, Strahl BD & Lieb JD (2005) Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol Cell Biol,* **25** (21), 9447-9459.

Raser JM & O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. *Science,* **309** (5743), 2010-2013.

Rath A, Glibowicka M, Nadeau VG, Chen G & Deber CM (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci U S A,* **106** (6), 1760-1765.

Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, Lerouge T, Hager GL, Chambon P & Losson R (2003) NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J,* **22** (12), 3153-3163.

Rechtsteiner A, Ercan S, Takasaki T, Phippen TM, Egelhofer TA, Wang W, Kimura H, Lieb JD & Strome S (2010) The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. *PLoS Genet,* **6** (9), e1001091.

Rega S, Stiewe T, Chang DI, Pollmeier B, Esche H, Bardenheuer W, Marquitan G & Putzer BM (2001) Identification of the full-length huntingtin-interacting protein p231$^{HBP/HYPB}$ as a DNA-binding factor. *Mol Cell Neurosci,* **18** (1), 68-79.

Richards EJ & Elgin SC (2002) Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell,* **108** (4), 489-500.

Ripaud L, Maillet L, Immel-Torterotot F, Durand F & Cullin C (2004) The [URE3] yeast prion results from protein aggregates that differ from amyloid filaments formed *in vitro*. *J Biol Chem,* **279** (49), 50962-50968.

Roh TY, Cuddapah S & Zhao K (2005) Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev,* **19** (5), 542-552.

Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S & Papavasiliou FN (2011) Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. *Nat Struct Mol Biol,* **18** (2), 230-236.

Roudier F, Ahmed I, Berard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, Giraut L, Despres B, Drevensek S, Barneche F, Derozier S, Brunaud V, Aubourg S, Schnittger A, Bowler C, Martin-Magniette ML, Robin S, Caboche M & Colot V (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J,* **30** (10), 1928-1938.

Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF, Washietl S, Arshinoff BI, Ay F, Meyer PE, Robine N, Washington NL, Di Stefano L, Berezikov E, Brown CD, Candeias R, Carlson JW, Carr A, Jungreis I, Marbach D, Sealfon R, Tolstorukov MY, Will S, Alekseyenko AA, Artieri C, Booth BW, Brooks AN, Dai Q, Davis CA, Duff MO, Feng X, Gorchakov AA, Gu T, Henikoff JG, Kapranov P, Li R, MacAlpine HK, Malone J, Minoda A, Nordman J, Okamura K, Perry M, Powell SK, Riddle NC, Sakai A, Samsonova A, Sandler JE, Schwartz YB, Sher N, Spokony R, Sturgill D, van Baren M, Wan KH, Yang L, Yu C, Feingold E, Good P, Guyer M, Lowdon R, Ahmad K, Andrews J, Berger B, Brenner SE, Brent MR, Cherbas L, Elgin SC, Gingeras TR, Grossman R, Hoskins RA, Kaufman TC, Kent W, Kuroda MI, Orr-Weaver T, Perrimon N, Pirrotta V, Posakony JW, Ren B, Russell S, Cherbas P, Graveley BR, Lewis S, Micklem G, Oliver B, Park PJ, Celniker SE, Henikoff S, Karpen GH, Lai EC, MacAlpine DM, Stein LD, White KP & Kellis M (2010) Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science,* **330** (6012), 1787-1797.

Rozen S & Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol,* **132**, 365-386.

Saddic LA, West LE, Aslanian A, Yates JR, 3rd, Rubin SM, Gozani O & Sage J (2010) Methylation of the retinoblastoma tumor suppressor by SMYD2. *J Biol Chem,* **285** (48), 37733-37740.

Saint-Andre V, Batsche E, Rachez C & Muchardt C (2011) Histone H3 lysine 9 trimethylation and HP1γ favor inclusion of alternative exons. *Nat Struct Mol Biol,* **18** (3), 337-344.

Sakamoto I, Kishida S, Fukui A, Kishida M, Yamamoto H, Hino S, Michiue T, Takada S, Asashima M & Kikuchi A (2000) A novel β-catenin-binding protein inhibits β-catenin-dependent Tcf activation and axis formation. *J Biol Chem,* **275** (42), 32871-32878.

Saksouk N, Avvakumov N, Champagne KS, Hung T, Doyon Y, Cayrou C, Paquet E, Ullah M, Landry AJ, Cote V, Yang XJ, Gozani O, Kutateladze TG & Cote J (2009) HBO1 HAT complexes target chromatin throughout gene coding regions via multiple PHD finger interactions with histone H3 tail. *Mol Cell,* **33** (2), 257-265.

Saltzman AL, Kim YK, Pan Q, Fagnani MM, Maquat LE & Blencowe BJ (2008) Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Mol Cell Biol,* **28** (13), 4320-4330.

Sambrook J, Fritsch EF & Maniatis T (1989) *Molecular cloning: a laboratory manual,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanchez R & Zhou MM (2009) The role of human bromodomains in chromatin biology and gene transcription. *Curr Opin Drug Discov Devel,* **12** (5), 659-665.

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J & Kouzarides T (2002) Active genes are tri-methylated at K4 of histone H3. *Nature,* **419** (6905), 407-411.

Santos-Rosa H, Valls E, Kouzarides T & Martinez-Balbas M (2003) Mechanisms of P/CAF auto-acetylation. *Nucleic Acids Res,* **31** (15), 4285-4292.

Sayer JA, Manczak M, Akileswaran L, Reddy PH & Coghlan VM (2005) Interaction of the nuclear matrix protein NAKAP with HypA and huntingtin: implications for nuclear toxicity in Huntington's disease pathogenesis. *Neuromolecular Med,* **7** (4), 297-310.

Schaft D, Roguev A, Kotovic KM, Shevchenko A, Sarov M, Neugebauer KM & Stewart AF (2003) The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. *Nucleic Acids Res,* **31** (10), 2475-2482.

Scharnhorst V, Dekker P, van der Eb AJ & Jochemsen AG (1999) Internal translation initiation generates novel WT1 protein isoforms with distinct biological properties. *J Biol Chem,* **274** (33), 23456-23462.

Schmitges FW, Prusty AB, Faty M, Stutzer A, Lingaraju GM, Aiwazian J, Sack R, Hess D, Li L, Zhou S, Bunker RD, Wirth U, Bouwmeester T, Bauer A, Ly-Hartig N, Zhao K, Chan H, Gu J, Gut H, Fischle W, Muller J & Thoma NH (2011) Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell,* **42** (3), 330-341.

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE & Zipursky SL (2000) *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell,* **101** (6), 671-684.

Schor IE, Rascovan N, Pelisch F, Allo M & Kornblihtt AR (2009) Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc Natl Acad Sci U S A,* **106** (11), 4325-4330.

Schotta G, Ebert A, Dorn R & Reuter G (2003) Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Semin Cell Dev Biol,* **14** (1), 67-75.

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W & Selbach M (2011) Global quantification of mammalian gene expression control. *Nature,* **473** (7347), 337-342.

Schwartz S, Meshorer E & Ast G (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol,* **16** (9), 990-995.

Schwartz YB & Pirrotta V (2007) Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet,* **8** (1), 9-22.

Schweizer A, Ericsson M, Bachi T, Griffiths G & Hauri HP (1993) Characterization of a novel 63 kDa membrane protein. Implications for the organization of the ER-to-Golgi pathway. *J Cell Sci,* **104** (3), 671-683.

Schweizer A, Rohrer J, Slot JW, Geuze HJ & Kornfeld S (1995) Reassessment of the subcellular localization of p63. *J Cell Sci,* **108** (6), 2477-2485.

Seong IS, Woda JM, Song JJ, Lloret A, Abeyrathne PD, Woo CJ, Gregory G, Lee JM, Wheeler VC, Walz T, Kingston RE, Gusella JF, Conlon RA & MacDonald ME (2010) Huntingtin facilitates polycomb repressive complex 2. *Hum Mol Genet,* **19** (4), 573-583.

Shaw BF, Lelie HL, Durazo A, Nersissian AM, Xu G, Chan PK, Gralla EB, Tiwari A, Hayward LJ, Borchelt DR, Valentine JS & Whitelegge JP (2008) Detergent-insoluble aggregates associated with amyotrophic lateral sclerosis in transgenic mice contain primarily full-length, unmodified superoxide dismutase-1. *J Biol Chem,* **283** (13), 8340-8350.

Shaw G, Morse S, Ararat M & Graham FL (2002) Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB Journal,* **16** (8), 869-871.

Shen Z, St-Denis A & Chartrand P (2010) Cotranscriptional recruitment of She2p by RNA pol II elongation factor Spt4-Spt5/DSIF promotes mRNA localization to the yeast bud. *Genes Dev,* **24** (17), 1914-1926.

Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ & Shi Y (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA,* **17** (4), 761-772.

Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Pena P, Lan F, Kaadige MR, Lacoste N, Cayrou C, Davrazou F, Saha A, Cairns BR, Ayer DE, Kutateladze TG, Shi Y, Cote J, Chua KF & Gozani O (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature,* **442** (7098), 96-99.

Shi X, Kachirskaia I, Walter KL, Kuo JH, Lake A, Davrazou F, Chan SM, Martin DG, Fingerman IM, Briggs SD, Howe L, Utz PJ, Kutateladze TG, Lugovskoy AA, Bedford MT & Gozani O (2007a) Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J Biol Chem,* **282** (4), 2450-2455.

Shi X, Kachirskaia I, Yamaguchi H, West LE, Wen H, Wang EW, Dutta S, Appella E & Gozani O (2007b) Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol Cell,* **27** (4), 636-646.

Shibata Y, Shemesh T, Prinz WA, Palazzo AF, Kozlov MM & Rapoport TA (2010) Mechanisms determining the morphology of the peripheral ER. *Cell,* **143** (5), 774-788.

Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR & Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science,* **311** (5762), 844-847.

Sims RJ, 3rd, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL & Reinberg D (2007) Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol Cell,* **28** (4), 665-676.

Smith CL, Debouck C, Rosenberg M & Culp JS (1989) Phosphorylation of serine residue 89 of human adenovirus E1A proteins is responsible for their characteristic electrophoretic mobility shifts, and its mutation affects biological function. *J Virol,* **63** (4), 1569-1577.

Smith JA, Francis SH & Corbin JD (1993) Autophosphorylation: a salient feature of protein kinases. *Mol Cell Biochem,* **127-128**, 51-70.

Smith R, Bacos K, Fedele V, Soulet D, Walz HA, Obermuller S, Lindqvist A, Bjorkqvist M, Klein P, Onnerfjord P, Brundin P, Mulder H & Li JY (2009) Mutant huntingtin interacts with β-tubulin and disrupts vesicular transport and insulin secretion. *Hum Mol Genet,* **18** (20), 3942-3954.

Smyth DG (1967) Carbamylation of amino and tyrosine hydroxyl groups. Preparation of an inhibitor of oxytocin with no intrinsic activity on the isolated uterus. *J Biol Chem,* **242** (7), 1579-1591.

Spies N, Nielsen CB, Padgett RA & Burge CB (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell,* **36** (2), 245-254.

Stabell M, Larsson J, Aalen RB & Lambertsson A (2007) *Drosophila* dSet2 functions in H3-K36 methylation and is required for development. *Biochem Biophys Res Commun,* **359** (3), 784-789.

Stec I, Wright TJ, van Ommen GJ, de Boer PA, van Haeringen A, Moorman AF, Altherr MR & den Dunnen JT (1998) *WHSC1*, a 90 kb SET domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to *IgH* in t(4;14) multiple myeloma. *Hum Mol Genet,* **7** (7), 1071-1082.

Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, Vakoc AL, Kim JE, Chen J, Lazar MA, Blobel GA & Vakoc CR (2008) DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol,* **28** (8), 2825-2839.

Steiner AW, Helander ER, Fujitaki JM, Smith LS & Smith RA (1980) High-performance liquid chromatography of acid-stable and acid-labile phosphoamino acids. *J Chromatogr,* **202** (2), 263-269.

Sterner DE, Lee JM, Hardin SE & Greenleaf AL (1995) The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex. *Mol Cell Biol,* **15** (10), 5716-5724.

Stewart SA, Dykxhoorn DM, Palliser D, Mizuno H, Yu EY, An DS, Sabatini DM, Chen IS, Hahn WC, Sharp PA, Weinberg RA & Novina CD (2003) Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA,* **9** (4), 493-501.

Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, Brockdorff N, Fisher AG & Pombo A (2007) Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol,* **9** (12), 1428-1435.

Stone MJ, Chuang S, Hou X, Shoham M & Zhu JZ (2009) Tyrosine sulfation: an increasingly recognised post-translational modification of secreted proteins. *N Biotechnol,* **25** (5), 299-317.

Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature,* **403** (6765), 41-45.

Strahl BD, Grant PA, Briggs SD, Sun ZW, Bone JR, Caldwell JA, Mollah S, Cook RG, Shabanowitz J, Hunt DF & Allis CD (2002) Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol,* **22** (5), 1298-1306.

Sun XJ, Wei J, Wu XY, Hu M, Wang L, Wang HH, Zhang QH, Chen SJ, Huang QH & Chen Z (2005) Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J Biol Chem,* **280** (42), 35261-35271.

Sun XJ, Xu PF, Zhou T, Hu M, Fu CT, Zhang Y, Jin Y, Chen Y, Chen SJ, Huang QH, Liu TX & Chen Z (2008) Genome-wide survey and developmental expression mapping of zebrafish SET domain-containing genes. *PLoS One,* **3** (1), e1499.

Sural TH, Peng S, Li B, Workman JL, Park PJ & Kuroda MI (2008) The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. *Nat Struct Mol Biol,* **15** (12), 1318-1325.

Takagaki Y & Manley JL (2000) Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol,* **20** (5), 1515-1525.

Talbert PB & Henikoff S (2010) Histone variants — ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol,* **11** (4), 264-275.

Tamura T, Smith M, Kanno T, Dasenbrock H, Nishiyama A & Ozato K (2009) Inducible deposition of the histone variant H3.3 in interferon-stimulated genes. *J Biol Chem,* **284** (18), 12217-12225.

Taverna SD, Li H, Ruthenburg AJ, Allis CD & Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol,* **14** (11), 1025-1040.

Thierry-Mieg D & Thierry-Mieg J (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol,* **7** (Suppl 1), S12: 11-14.

Thompson PR, Wang D, Wang L, Fulco M, Pediconi N, Zhang D, An W, Ge Q, Roeder RG, Wong J, Levrero M, Sartorelli V, Cotter RJ & Cole PA (2004) Regulation of the p300 HAT domain via a novel activation loop. *Nat Struct Mol Biol,* **11** (4), 308-315.

Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J & Guigo R (2009) Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol,* **16** (9), 996-1001.

Todoerti K, Ronchetti D, Agnelli L, Castellani S, Marelli S, Deliliers GL, Zanella A, Lombardi L & Neri A (2005) Transcription repression activity is associated with the type I isoform of the *MMSET* gene involved in t(4;14) in multiple myeloma. *Br J Haematol,* **131** (2), 214-218.

Tompa R & Madhani HD (2007) Histone H3 lysine 36 methylation antagonizes silencing in *Saccharomyces cerevisiae* independently of the Rpd3S histone deacetylase complex. *Genetics,* **175** (2), 585-593.

Tong JK, Hassig CA, Schnitzler GR, Kingston RE & Schreiber SL (1998) Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature,* **395** (6705), 917-921.

Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats AC & Vagner S (2003) Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol Cell,* **95** (3-4), 169-178.

Towbin H, Staehelin T & Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A,* **76** (9), 4350-4354.

Tripic T, Edmondson DG, Davie JK, Strahl BD & Dent SY (2006) The Set2 methyltransferase associates with Ssn6 yet Tup1-Ssn6 repression is independent of histone methylation. *Biochem Biophys Res Commun,* **339** (3), 905-914.

Tsang LW, Hu N & Underhill DA (2010) Comparative analyses of SUV420H1 isoforms and SUV420H2 reveal differences in their cellular localization and effects on myogenic differentiation. *PLoS One,* **5** (12), e14447.

Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P & Zhang Y (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature,* **439** (7078), 811-816.

Turner BM (2005) Reading signals on the nucleosome with a new nomenclature for modified histones. *Nat Struct Mol Biol,* **12** (2), 110-112.

Turner BM, Birley AJ & Lavender J (1992) Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell,* **69** (2), 375-384.

Vakoc CR, Mandat SA, Olenchock BA & Blobel GA (2005) Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. *Mol Cell,* **19** (3), 381-391.

Vakoc CR, Sachdeva MM, Wang H & Blobel GA (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol Cell Biol,* **26** (24), 9185-9195.

van Bakel H & Hughes TR (2009) Establishing legitimacy and function in the new transcriptome. *Brief Funct Genomic Proteomic,* **8** (6), 424-436.

Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, Davies H, Jones D, Lin ML, Teague J, Bignell G, Butler A, Cho J, Dalgliesh GL, Galappaththige D, Greenman C, Hardy C, Jia M, Latimer C, Lau KW, Marshall J, McLaren S, Menzies A, Mudie L, Stebbings L, Largaespada DA, Wessels LF, Richard S, Kahnoski RJ, Anema J, Tuveson DA, Perez-Mancera PA, Mustonen V, Fischer A, Adams DJ, Rust A, Chan-on W, Subimerb C, Dykema K, Furge K, Campbell PJ, Teh BT, Stratton MR & Futreal PA (2011) Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature,* **469** (7331), 539-542.

Venters BJ & Pugh BF (2009) How eukaryotic genes are transcribed. *Crit Rev Biochem Mol Biol,* **44** (2-3), 117-141.

Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG & Mann M (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell,* **142** (6), 967-980.

Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M & Timmers HT (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell,* **131** (1), 58-69.

Verrier L, Escaffit F, Chailleux C, Trouche D & Vandromme M (2011) A new isoform of the histone demethylase JMJD2A/KDM4A is required for skeletal muscle differentiation. *PLoS Genet,* **7** (6), e1001390.

Vezzoli A, Bonadies N, Allen MD, Freund SM, Santiveri CM, Kvinlaug BT, Huntly BJ, Gottgens B & Bycroft M (2010) Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nat Struct Mol Biol,* **17** (5), 617-619.

Villard J (2004) Transcription regulation and human diseases. *Swiss Med Wkly,* **134** (39-40), 571-579.

Vojnic E, Simon B, Strahl BD, Sattler M & Cramer P (2006) Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys(36) methylation to transcription. *J Biol Chem,* **281** (1), 13-15.

Volkel P & Angrand PO (2007) The control of histone lysine methylation in epigenetic regulation. *Biochimie,* **89** (1), 1-20.

Wada Y, Ohta Y, Xu M, Tsutsumi S, Minami T, Inoue K, Komura D, Kitakami J, Oshida N, Papantonis A, Izumi A, Kobayashi M, Meguro H, Kanki Y, Mimura I, Yamamoto K, Mataki C, Hamakubo T, Shirahige K, Aburatani H, Kimura H, Kodama T, Cook PR & Ihara S (2009) A wave of nascent transcription on activated human genes. *Proc Natl Acad Sci U S A,* **106** (43), 18357-18361.

Walker FO (2007) Huntington's disease. *Lancet,* **369** (9557), 218-228.

Wang CL (2001) Caldesmon and smooth-muscle regulation. *Cell Biochem Biophys,* **35** (3), 275-288.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP & Burge CB (2008a) Alternative isoform regulation in human tissue transcriptomes. *Nature,* **456** (7221), 470-476.

Wang XQ & Rothnagel JA (2004) 5'-untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. *Nucleic Acids Res,* **32** (4), 1382-1391.

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ & Zhao K (2008b) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet,* **40** (7), 897-903.

Watt RA, Shatzman AR & Rosenberg M (1985) Expression and characterization of the human c-*myc* DNA-binding protein. *Mol Cell Biol,* **5** (3), 448-456.

Weber K & Osborn M (1969) The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J Biol Chem,* **244** (16), 4406-4412.

Wegrzyn JL, Bark SJ, Funkelstein L, Mosier C, Yap A, Kazemi-Esfarjani P, La Spada AR, Sigurdson C, O'Connor DT & Hook V (2010) Proteomics of dense core secretory vesicles reveal distinct protein categories for secretion of neuroeffectors for cell-cell communication. *J Proteome Res,* **9** (10), 5002-5024.

Wegrzyn JL, Drudge TM, Valafar F & Hook V (2008) Bioinformatic analyses of mammalian 5'-UTR sequence properties of mRNAs predicts alternative translation initiation sites. *BMC Bioinformatics,* **9**, 232.

Wernersson R (2006) Virtual Ribosome—a comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Res,* **34** (Web Server issue), W385-388.

Wesierska-Gadek J, Gueorguieva M, Kramer MP, Ranftler C, Sarg B & Lindner H (2007) A new, unexpected action of olomoucine, a CDK inhibitor, on normal human cells: up-regulation of CLIMP-63, a cytoskeleton-linking membrane protein. *J Cell Biochem,* **102** (6), 1405-1419.

Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, Chen Z, Spooner E, Li E, Zhang G, Colaiacovo M & Shi Y (2006) Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell,* **125** (3), 467-481.

Whitehead KA, Langer R & Anderson DG (2009) Knocking down barriers: advances in siRNA delivery. *Nat Rev Drug Discov,* **8** (2), 129-138.

Wilkinson FL, Nguyen TM, Manilal SB, Thomas P, Neal JW, Harper PS, Jones AL & Morris GE (1999) Localization of rabbit huntingtin using a new panel of monoclonal antibodies. *Brain Res Mol Brain Res,* **69** (1), 10-20.

Wong P, Hattangadi SM, Cheng AW, Frampton GM, Young RA & Lodish HF (2011) Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes. *Blood,* **118** (16), e128-e138.

Wu H, Zeng H, Lam R, Tempel W, Amaya MF, Xu C, Dombrovski L, Qiu W, Wang Y & Min J (2011a) Structural and histone binding ability characterizations of human PWWP domains. *PLoS One,* **6** (6), e18919.

Wu J, Smith LT, Plass C & Huang TH (2006) ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res,* **66** (14), 6899-6902.

Wu SF, Zhang H & Cairns BR (2011b) Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. *Genome Res,* **21** (4), 578-589.

Xhemalce B & Kouzarides T (2010) A chromodomain switch mediated by histone H3 Lys 4 acetylation regulates heterochromatin assembly. *Genes Dev,* **24** (7), 647-652.

Xiao T, Hall H, Kizer KO, Shibata Y, Hall MC, Borchers CH & Strahl BD (2003) Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev,* **17** (5), 654-663.

Xie L, Pelz C, Wang W, Bashar A, Varlamova O, Shadle S & Impey S (2011) KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO J,* **30** (8), 1473-1484.

Xie P, Tian C, An L, Nie J, Lu K, Xing G, Zhang L & He F (2008) Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. *Cell Signal,* **20** (9), 1671-1678.

Xie W, Song C, Young NL, Sperling AS, Xu F, Sridharan R, Conway AE, Garcia BA, Plath K, Clark AT & Grunstein M (2009) Histone h3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. *Mol Cell,* **33** (4), 417-427.

Xu CR, Schaffer L, Head SR & Feeney AJ (2008a) Reciprocal patterns of methylation of H3K36 and H3K27 on proximal vs. distal IgV$_H$ genes are modulated by IL-7 and Pax5. *Proc Natl Acad Sci U S A,* **105** (25), 8685-8690.

Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou JP, Steinmetz A & Shen WH (2008b) Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in Arabidopsis thaliana. *Mol Cell Biol,* **28** (4), 1348-1360.

Xu W, Edmondson DG & Roth SY (1998) Mammalian GCN5 and P/CAF acetyltransferases have homologous amino-terminal domains important for recognition of nucleosomal substrates. *Mol Cell Biol,* **18** (10), 5659-5669.

Yamada-Okabe T, Imamura K, Kawaguchi N, Sakai H, Yamashita M & Matsumoto N (2010) Functional characterization of the zebrafish WHSC1-related gene, a homolog of human NSD2. *Biochem Biophys Res Commun,* **402** (2), 335-339.

Yang B, Gonzalez L, Jr., Prekeris R, Steegmaier M, Advani RJ & Scheller RH (1999) SNARE interactions are not selective. Implications for membrane fusion specificity. *J Biol Chem,* **274** (9), 5649-5653.

Yang J & Everett AD (2007) Hepatoma-derived growth factor binds DNA through the N-terminal PWWP domain. *BMC Mol Biol,* **8**, 101.

Yang J, Huang J, Dasgupta M, Sears N, Miyagi M, Wang B, Chance MR, Chen X, Du Y, Wang Y, An L, Wang Q, Lu T, Zhang X, Wang Z & Stark GR (2010) Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. *Proc Natl Acad Sci U S A,* **107** (50), 21499-21504.

Yang ZR, Thomson R, McNeil P & Esnouf RM (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics,* **21** (16), 3369-3376.

Yee SP & Branton PE (1985) Analysis of multiple forms of human adenovirus type 5 E1A polypeptides using an anti-peptide antiserum specific for the amino terminus. *Virology,* **146** (2), 315-322.

Yoh SM, Cho H, Pickle L, Evans RM & Jones KA (2007) The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev,* **21** (2), 160-174.

Yoh SM, Lucas JS & Jones KA (2008) The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev,* **22** (24), 3422-3434.

Youdell ML, Kizer KO, Kisseleva-Romanova E, Fuchs SM, Duro E, Strahl BD & Mellor J (2008) Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Mol Cell Biol,* **28** (16), 4915-4926.

Yuan W, Xie J, Long C, Erdjument-Bromage H, Ding X, Zheng Y, Tempst P, Chen S, Zhu B & Reinberg D (2009) Heterogeneous nuclear ribonucleoprotein L is a subunit of human KMT3a/Set2 complex required for H3 Lys-36 trimethylation activity *in vivo*. *J Biol Chem,* **284** (23), 15701-15707.

Yuan W, Xu M, Huang C, Liu N, Chen S & Zhu B (2011) H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem,* **286** (10), 7983-7989.

Yun M, Wu J, Workman JL & Li B (2011) Readers of histone modifications. *Cell Res,* **21** (4), 564-578.

Zentner GE, Saiakhova A, Manaenkov P, Adams MD & Scacheri PC (2011) Integrative genomic analysis of human ribosomal DNA. *Nucleic Acids Res,* **39** (12), 4949-4960.

Zhang P, Du J, Sun B, Dong X, Xu G, Zhou J, Huang Q, Liu Q, Hao Q & Ding J (2006) Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res,* **34** (22), 6621-6628.

Zhou K, Kuo WH, Fillingham J & Greenblatt JF (2009) Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc Natl Acad Sci U S A,* **106** (17), 6956-6961.

Zhou X, Marks PA, Rifkind RA & Richon VM (2001) Cloning and characterization of a histone deacetylase, HDAC9. *Proc Natl Acad Sci U S A,* **98** (19), 10572-10577.

Zhu S, Jiang Q, Wang G, Liu B, Teng M & Wang Y (2011) Chromatin structure characteristics of pre-miRNA genomic sequences. *BMC Genomics,* **12**, 329.

Zibetti C, Adamo A, Binda C, Forneris F, Toffolo E, Verpelli C, Ginelli E, Mattevi A, Sala C & Battaglioli E (2010) Alternative splicing of the histone demethylase LSD1/KDM1 contributes to the modulation of neurite morphogenesis in the mammalian nervous system. *J Neurosci*, **30** (7), 2521-2532.

Zuccato C, Tartari M, Crotti A, Goffredo D, Valenza M, Conti L, Cataudella T, Leavitt BR, Hayden MR, Timmusk T, Rigamonti D & Cattaneo E (2003) Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nat Genet*, **35** (1), 76-83.