DNA Methylation Analysis of the Evolution of Wilms Tumour from its Precursor Nephrogenic Rests

Jocelyn Charlton

University College London

A dissertation submitted to University College London in candidature for the degree of Doctor of Philosophy

September 2014

Declaration

This dissertation entitled '*DNA methylation Analysis of the Evolution of Wilms tumour from its Precursor Nephrogenic Rests'* is the result of my own work and includes nothing which is the outcome of work done in collaboration except were specifically indicated in the text. No part of my research has already been submitted, or is concurrently being submitted, for any other research degree, diploma or qualification.

Jocelyn Charlton

_____

Date: 3rd September 2014

Acknowledgements

I would like to thank Professor Kathy Pritchard-Jones for giving me the opportunity to undertake this PhD and all her support and guidance over the last three years, including supporting travel to international conferences to share my research and for giving me the freedom to pursue my own independent work. I would also like to thank the members of the Pritchard-Jones research team, past and present, and in particular Mariana Maschietto, who I would like to thank for her consistent support, advice and friendship throughout my PhD; I couldn't have done it without you! I would like to thank Richard Williams for sparking my initial interest in bioinformatics, Tasnim Chagtai and Marisa Alcaide-German for their initial support when I first started, Becky West for always being so precise and Mark Weeks for providing humour when necessary! Thank you for joining in when I forced you to dress up as Santa and run around Victoria Park, thank you for eating many cakes with me, drinking many coffees and listening to my many strange stories.

A huge thank you also goes to Professor Stephan Beck, as my secondary supervisor, as well as the members of his research team, Tiffany Morris, Lee Butcher and Paul Guilhamon. You all gave essential help when I needed it the most, driving my passion for epigenetics!

I am privileged to have received support from a team of experienced paediatric histopatholgists. In particular, Neil Sebire and Sergey Popov, who spent hours looking through stacks of tumour sections to identify nephrogenic rests. I admire your patience and hope that one day, there will be a NR wonder-marker. Also thank

you to Gordan Vujanic and William Misfud who will be sad they missed out on the initial NR hunt, but provided essential support afterwards.

I extend my thanks to the UCL Genomics team, in particular, Kerra Pearce and Tony Brooks for running my arrays and sequencing for me. And also to the UCL Grand Challenges Scheme for funding my PhD.

On a more personal note, I wish to thank my parents for their constant love, support and encouragement, and Alana, Conrad and Verity for not only being my siblings, but my friends.

Abstract

Recurrent loss of imprinting at 11p15, paucity of recurrent genetic mutations and associated nephrogenic rests (NR; precursor lesions that resemble embryonic kidney (EK)) implicate aberrant DNA methylation in tumourigenesis of paediatric Wilms tumour (WT) and predict that interrogation of the methylome, rather than the genome, is more likely to reveal tumour-specific biomarkers

To test if aberrant DNA methylation is implicated in tumourigenesis, methylome analysis was performed on 36 normal kidney (NK), 22 NR, 36 WT and 4 EK, including 20 matched trios and 34 matched NK-WT pairs, using Illumina 450k arrays. Findings were validated with bisulfite-sequencing and RNA sequencing.

This thesis describes the successful identification of changes in methylation that distinguish between tissue types. Through analysis of DNA methylation, NR formation was associated with gain of methylation at developmental loci related to Polycomb target binding sites. Evolution to WT was associated with increase in methylation variability in a subset of WTs (group-1), which also showed common changes in methylation in comparison to their associated NR, including silencing of novel tumour suppressor genes. Group-1 WTs were significantly enriched for bilateral cases whereas those in group-2 showed no differences in methylation compared to their associated NR.

Comparison between NK and WT identified three DMRs of genome-wide significance ($P<5x10^{-8}$) for use as tumour-specific biomarkers. As proof of principle for clinical utility, DMR-2 was successfully used in a case study to monitor tumour burden during treatment in cell-free serum DNA.

This thesis concludes that methylation levels vary during WT evolution. As group-1 WT included all bilateral cases, our data suggests that methylation analysis could aid treatment planning in bilateral disease and that some WT may be candidates for epigenetic-modifier therapy. These findings define the first cell-free epigenetic biomarker for WT with potential for clinical utility.

Table of contents

List of Figures

List of Tables

List of Abbreviations

AdoMet	S-adenosyl-L-methionine
ANOVA	Analysis of variance
ATRA	All-trans retinoic acid
Δβ	Change in β-value
β_{mean}	Average β value (across CpGs)
BWS	Beckwith Wiedemann syndrome
cfDNA	Cell free DNA
CGI	CpG Island
CIC	Chemotherapy induced changes
CIMP	CpG island methylator phenotype
CPDN	Cystic partially differentiated nephroblastoma
CTCF	CCCTC-binding factor
DMR	Differentially methylated region
DNMT	DNA methyl transferase
Dox	Doxycycline
ESCs	Embryonic stem cells
EK	Embryonic kidney
EMT	Epithelial to mesenchymal transition
EWAS	Epigenome-wide association study
FDR	False discovery rate
FFPE	Formalin fixed paraffin embedded
GO	Gene ontology
GOSH	Great Ormond Street Hospital
GREAT	Genomic regions enrichment of annotations tool
H&E	Haematoxylin and eosin

H3K27ac	Histone 3 lysine 27 acetylation
H2K27me3	Histone 3 lysine 27 trimethylation
H3K36me3	Histone 3 lysine 36 trimethylation
H3K4me1	Histone 3 lysine 4 monomethylation
H3K4me3	Histone 3 lysine 4 trimethylation
H3K9Me2	Histone 3 lysine 9 dimethylation
H-me	Hydroxymethylation
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDBR	Human Developmental Biology Resource
hyper-DMRs	Hypermethylated differentially methylated region
hyper-MVPs	Hypermethylated methylation variable positions
hypo-DMRs	Hypomethylated differentially methylated region
hypo-MVPs	Hypometylated methylation variable positions
ICR	Imprinting control region
IGF	Insulin-like growth factor
IGF1R	Insulin-like growth factor receptor 1
ILNR	Intralobar nephrogenic rest
IM	Intermediate mesenchyme
IMPORT	Improving Outcomes for Patients with Renal Tumours
IPS	Induced pluripotent stem cell lines
IRS	Insulin receptor substrates
KDM	Lysine demethylase
KMT	Lysine methyltransferase
KR-DMRs	Differentially methylated regions in NR compared to NK
LMR	Low methylated region
LOI	Loss of imprinting

MDS	Multidimensional scaling
MET	Mesenchymal to epithelial transition
MHC	Major histocompatibility complex
miRNA	Micro RNA
MM	Metanephric mesenchyme
mRNA	Messenger RNA
MVP	Methylation variable position
NK	Normal kidney
NR	Nephrogenic rest
PCDH	Protocadherin gene cluster
PCR	Polymerase chain reaction
pDMR	Putative differentially methylated region
PLNR	Perilobar nephrogenic rest
PRC	Polycomb repressive complex
pUPD	Paternal uniparental disomy
RA	Retinoic acid
RAR	Retinoic acid receptor
RAREs	Retinoic acid response elements
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RXR	Retinoic X receptor
SIOP	Society of Paediatric Oncology
SWAN	Subset within quantile normalisation
TF	Transcription factor
WAGR	Wilms, anaridia, genitouriniary abnormalities and mental retardation
WT	Wilms tumour

WT-DMRs	Differentially methylated regions in WT compared to NR

Chapter 1: Introduction and background

1.1 Thesis overview

The work described in this thesis aims to investigate the hypothesis that epigenetic changes (in particular, DNA methylation) underlie:

- the mechanisms by which embryonic kidney may sometimes fail to fully differentiate, leading to persistence of 'nephrogenic rests',
- the transformation of these presumed precursor lesions into fully malignant Wilms tumour,
- the subsequent clinical phenotype of the Wilms tumour, in relation to presenting clinical and histological features.

Chapter 1 provides a broad overview of the molecular biology of Wilms tumour (WT) aetiology with particular focus on the process of nephrogenesis (kidney development) and how its disruption is associated with persistence of regions of incompletely or aberrantly differentiated tissue, termed WT precursors, nephrogenic rests (NR). Next, the array of epigenetic modifications and their influence on gene expression is discussed in the context of both development and cancer to give an in depth understanding of the similarities between both processes that are represented in this childhood cancer that results from arrested normal development. Recent publications have been included in the introductory background description, that although were not considered for generating the thesis hypotheses, incorporate new knowledge that substantially supports the contribution of epigenetic changes to WT evolution. For example, a recent mouse model of *in vivo* reprogramming identified

WT-like lesions with altered epigenomic architecture yet no identifiable somatic genetic mutations (Ohnishi, Semi et al. 2014).

1.2 Overview of Wilms tumour

WT is a paediatric kidney cancer with a prevalence of 1 in 10,000 children younger than 15 years of age (Breslow, Olshan et al. 1993). WT is an embryonal cancer, meaning it shows mimicry of fetal development processes in the corresponding organ of origin. Hence it comprises undifferentiated blastema cells, immature epithelia and stroma; structures that are seen during normal nephrogenesis (Figure 1.1). Embryonal childhood cancers in general were the first models described by the Knudson two-hit hypothesis for cancer generation whereby a tumour suppressor (TS) gene is silenced by germline loss-of-function mutation or deletion of one allele, then the second allele is lost as a somatic event post-natally leading to cancer. Mutations in TS genes such as Retinoblastoma 1 (*RB1)* in retinoblastoma and Wilms Tumour 1 (*WT1)* in WT have been identified in both germline and tumour DNA following this hypothesis; however genetic predisposition in WT is uncommon (~5% of all cases) and only a minority of such cases follow Knudson's 2-hit model, those involving the *WT1* gene.

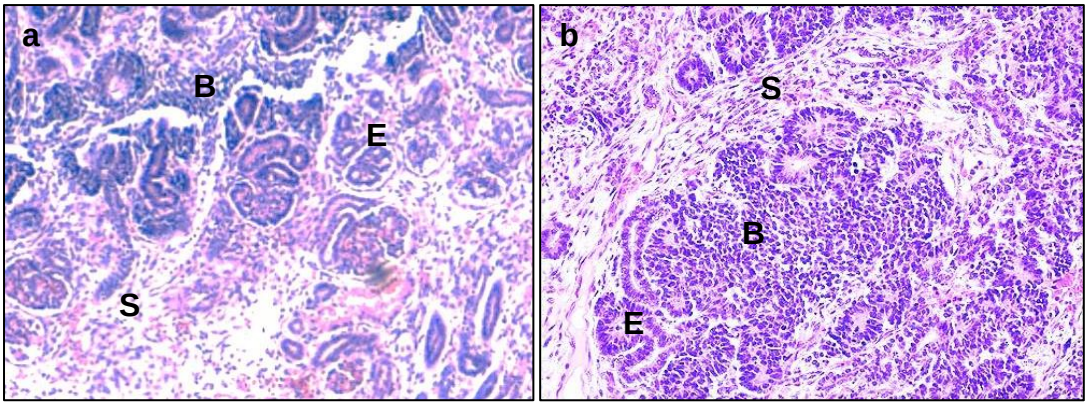


Figure 1.1: Wilms tumour shows mimicry of cell types in normal embryonic kidney. (**a**) A haematoxylin and eosin (H&E) stained 3µm section of an embryonic kidney at 14 weeks showing epithelial (E), stromal (S) and blastemal (B) elements. (**b**) An H&E stained section of a mixed Wilms tumour showing E, S and B elements. The epithelia observed in WT is similar to comma-shaped or s-shaped body or glomeruli in embryonic kidney whereas the blastema mimics condensing metanephric mesencyme (MM; discussed later).

Related to its embryonic composition, WTs are associated with disrupted renal development. From assessment of post-nephrectomy WT, 40% sporadic tumours and ~100% bilateral tumours can be associated with NRs which are considered precursor lesions to WT (Beckwith, Kiviat et al. 1990) (Figure 1.2). These appear morphologically comparable to embryonic kidney (EK) with disrupted architecture, however their presence within the post-natal kidney suggests these lesions have failed to successfully complete nephrogenesis (Beckwith, Kiviat et al. 1990). There are two types of NR; intralobar (ILNR), which are found within the renal medulla, and perilobar (PLNR) found towards the periphery of the kidney (Beckwith, Kiviat et al. 1990). ILNR are thought to arise at an earlier developmental stage as nephron formation begins in the renal medulla then continues outwards towards the periphery (Beckwith, Kiviat et al. 1990). NRs are found in approximately 1% of normal

neonatal kidneys at autopsy (Lonergan, Martinez-Leon et al. 1998) whereas only 1 in 10,000 children get WT giving a supposed malignant transformation rate of approximately 1 in 100. This rate is similar to that found in acute lymphoid leukemia (ALL) where the frequency of TEL-AML1 fusion genes in neonatal blood spots is 100-fold greater than the risk of the corresponding TEL-AML1 positive ALL (Mori, Colman et al. 2002). Despite accurate epidemiological characterisation of these lesions, the reasons behind tissue retention and the steps leading to transformation are unknown.

Similar to other paediatric malignancies, few genetic mutations have been found in WT and the majority of tumours do not carry any of the most common mutations (Scott, Murray et al. 2012), although limited genome-wide sequencing studies have been published. The most common aberration, that is syndrome-associated and is therefore linked to predisposition, occurs at the 11p15 locus. At 11p15, specific loss of the maternal allele over the paternal allele resulting in loss of heterozygosity (LOH) by duplication of the paternal allele puzzled scientists for years until the phenomenon of genomic imprinting at this loci controlling allele-specific expression of insulin-like growth factor 2 (*IGF2)* and imprinted maternally expressed transcript (*H19)* was discovered. As epigenetic alterations, defined as changes to DNA that alter gene expression without affecting the DNA coding sequence, are critical in development, and since one of the most commonly associated loci with WT (11p15) shows epigenetic dysregulation, it can be hypothesised that epigenetic mechanisms play a major role in WT tumourgienesis and that interrogation of the methylome, rather than the genome, is more likely to reveal tumour-specific biomarkers.

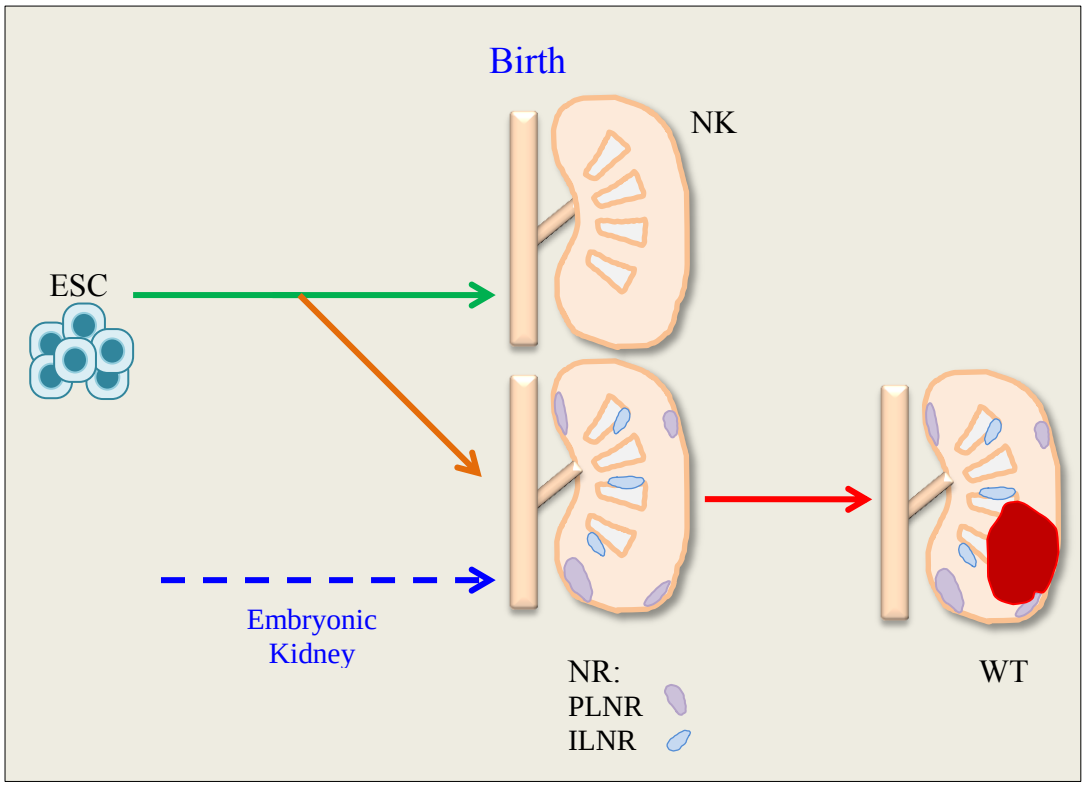


Figure 1.2: Schematic of Wilms tumour (WT) growth from its precursor lesion, the nephrogenic rest (NR). Normal kidney (NK) development (green arrow) is disrupted in rare cases (orange arrow) resulting in a post-natal kidney with perilobar and/or intralobar nephrogenic rests (purple and brown respectively). From these, WT can grow (red).

1.3 Epidemiology

The median age of WT onset is 38 months however age at diagnosis is affected by patient sex, the presence of NRs, underlying syndromes and laterality (Beckwith, Kiviat et al. 1990; Breslow, Olshan et al. 1993). Males are diagnosed on average 6 months earlier than females (Breslow, Olshan et al. 1993). Age at diagnosis is also decreased by the presence of ILNRs (16 months) over PLNRs (36 months) and even further with the presence of both (12 months; see Figure 1.3) (Breslow, Olshan et al. 1993).

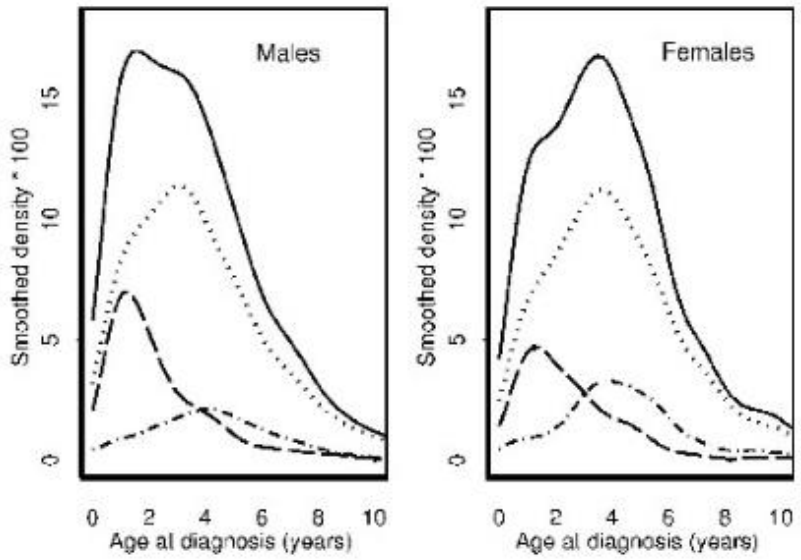


Figure 1.3: Frequency distribution and sub-distributions of age-at-onset by gender for patients with unifocal tumours, taken from (Breslow, Beckwith et al. 2006). Solid line = all patients, dotted line = NR negative, dashed line = ILNR (+/- PLNR), dashed and dotted line = PLNR only.

There are many syndromes described that predispose to WT (Scott, Stiller et al. 2006), however, only the two major subgroups of WT predisposition syndromes associated with known genetic mechanisms are described here. WAGR syndrome (Wilms tumour, aniridia, genitourinary anomalies and mental retardation) and Denys-Drash syndrome are associated with genitourinary malformation, *WT1* mutation and ILNRs (Francke, Holmes et al. 1979; Beckwith, Kiviat et al. 1990; Pelletier, Bruening et al. 1991). Beckwith-Wiedemann syndrome (BWS) and the overlapping hemi-hypertrophy associated WTs are associated with asymmetric overgrowth, PLNRs and loss of imprinting at 11p15 affecting expression of genes *IGF2* and *H19* (Beckwith, Kiviat et al. 1990; Reik, Brown et al. 1995). As both *WT1* and the IGF2/H19 locus are located on chromosome 11p, loss of the maternal allele may simultaneously cause reduction to homozygosity of a mutant *WT1* allele, acting

as a tumour suppressor gene and duplication of the paternally expressed *IGF2* and complete silencing of the maternally expressed *H19* genes. Both insults acting together have been shown to generate WT an *in vivo* mouse model (Hu, Gao et al. 2011). The more recently characterised Perlman syndrome, a rare autosomal recessive overgrowth syndrome with increased risk of WT, is associated with mutations in the gene DIS3 mitotic control homolog (S. Cerevisiae)-like 2 (*DIS3L2*) with reciprocal upregulation of Let-7 and its targets which promote oncogenesis (Chang, Triboulet et al. 2013; Morris, Astuti et al. 2013).

Familial predisposition to WT has also been associated with loci on 19q (McDonald, Douglass et al. 1998) and 17q (Rahman, Arbour et al. 1996) although specific genes have not been identified. Besides predisposition syndromes and familial studies, comparison of blood from WT patients to controls identified single nucleotide polymorphisms (SNPs) at 2p24 and 11q14 that confer susceptibility to WT (Turnbull, Perdeaux et al. 2012).

1.4 Pathology and classification

WTs show a mixed histology with varying proportions of stroma, epithelia and blastema. WTs can also be anaplastic, which is a morphological diagnosis comprising cells with large or multiple nuclei that are abnormally shaped with nucleus size variation at least 3-fold and atypical mitoses. This is a sign of increased proliferation and a hallmark of aggressive cancer. There are currently two main treatment approaches for WT that differ in their histological classification of WT. The Society of Paediatric Oncology (SIOP) protocol used in Europe and some other

countries performs pre-operative chemotherapy and then assesses the cellular composition according to the tumour response to determine the histological risk group and hence the level of post-operative chemotherapy. The Children's Oncology Group (COG) protocol used in the USA performs immediate nephrectomy then describes the cellular composition of the untreated tumour.

Based on histopathological review of a resection, tumours are classified and subdivided into (SIOP) low, intermediate and high risk groups or (COG) favourable or unfavourable groups (Table 1.1). SIOP tumours are also staged according to response to chemotherapy, completeness of tumour excision, whether the renal capsule is intact, tumour rupture, presence of tumour beyond excision margins, contamination of local vessels or lymph nodes and laterality (Table A1, Appendix 1). Currently, no molecular biomarkers aid clinical diagnosis or staging of WT for patients treated according to the SIOP protocol however there is new evidence that post-chemotherapy blastema volume at nephrectomy may be of valuable prognostic significance; this is yet to be fully explored. Conversely, patients treated according to the COG protocol are stratified according to loss of heterozygosity (LOH) at 1p and 16q (Grundy, Telzerow et al. 1994; Grundy, Breslow et al. 2005).

SIOP risk group	Histology (post-chemotherapy)	COG risk group	Histology (immediate nephrectomy)
Low	Completely necrotic	**Favourable**	Blastemal predominant
Intermediate	Epithelial type		Stromal predominant
	Stromal type		Epithelial predominant
	Mixed type		Mixed histology
	Regressive type		
	Focal anaplasia		
High	Blastema type	**Unfavourable**	Diffuse anaplasia
	Diffuse anaplasia		Focal anaplasia

Table 1.1 – Wilms tumour histological classification and risk groups according to the SIOP and COG protocols. Focal anaplasia refers to distinct focal regions of anaplastic cells whereas diffuse anaplasia refers to widespread anaplsia throughout the tumour.

1.5 Clinical management

WT can be diagnosed clinically from its typical presentation as an asymptomatic abdominal mass in an otherwise healthy child and typical imaging appearances of an intrarenal tumour. Occasionally, biopsy of the renal mass is used for histological confirmation. In the USA, immediate nephrectomy is performed for all patients. Although the SIOP protocol specifies pre-operative chemotherapy, patients younger than 6 months of age undergo immediate nephrectomy without pre-operative chemotherapy. In all other cases, a course of pre-operative chemotherapy is given of vincristine (V) and dactinomycin (A) for 4 weeks. The length and number of drugs given as post-operative chemotherapy is dependent of risk and stage and

typically includes V and A for stage I with additional doxorubicin (D) for stages II-V. Bilateral disease or cases with diffuse anaplasia may be given additional cyclophosphamide and etoposide and stage III patients have additional radiotherapy.

Studies on survivors of paediatric cancers show that around 60% of adolescents develop chronic health problems including, for WT, musculoskeletal effects, cardiac toxicity, reproductive problems, renal dysfunction and the development of second malignant neoplasms (Wright, Green et al. 2009). These are usually due to the further drugs and radiotherapy given to higher stage or relapsed patients (Wright, Green et al. 2009).

From 2001-2011 a Society of Paediatric Oncology (SIOP) study allowed for consented collection and storage of fresh frozen and formalin fixed paraffin embedded (FFPE) tissue. From 2012, this was superseded in the UK by the Improving Population Outcomes for Children with Renal Tumours (IMPORT) study which also collects serial aliquots of blood serum, blood plasma and urine at diagnosis, mid-way during pre-operative chemotherapy, immediately prior to nephrectomy, post-surgery prior to post-operative chemotherapy and at end-of-treatment (Figure 1.4).

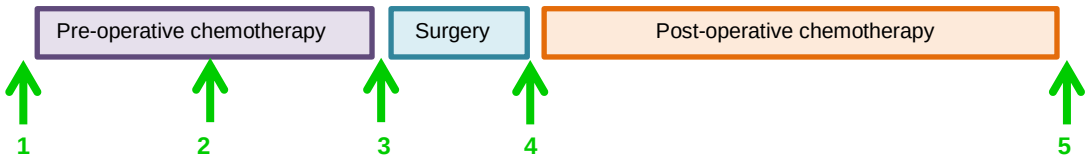


Figure 1.4: IMPORT study serial sample collection time-course. Samples are collected at 5 time points during treatment indicated by green arrows.

1.6 Renal development and Wilms tumour

WT is a developmental cancer, evidenced by its occurrence in early childhood, its association with precursor lesions that resemble EK and mixed histology of different embryonic cell types. To understand the transformation stages involved in Wilms tumourigenesis, it is important to understand the process of normal renal development, as the two are closely linked.

1.6.1 Renal development overview

Nephrogenesis describes the process of embryonic growth and maturation of the kidney. This process begins at 22 days and terminates at around 36 weeks gestation. During embryogenesis, the intermediate mesoderm (IM) gives rise to the metanephros along the anteroposterior axis. Within this structure, IM cells progress to form the metanephric mesenchyme (MM), or 'renal stem cells'. In the meantime, from the Wolffian duct, an epithelial structure termed the ureteric bud extends. This structure is essential to establish the initiation of nephrogenesis. Upon contact with the MM, reciprocal induction occurs, whereby the MM induce the ureteric bud to bifurcate and hence form the collecting duct system of the adult kidney whilst the ureteric bud simultaneously induces the MM to undergo differentiation (Bard, McConnell et al. 1994). Some of these stem cells differentiate into stromal cells that predominantly localise to the medulla and some condense into small aggregates that go through mesenchymal to epithelial transition (MET) to form comma-shaped then S-shaped bodies before forming a nephron (Figure 1.5). The distal end of the forming nephron fuses with the bifurcated duct and the proximal end forms a

glomerulus and localises with capillaries for ultrafiltration of the blood. Any uninduced MM undergo apoptosis (Bard, McConnell et al. 1994; Rivera and Haber 2005; Schedl 2007). The number of nephrons is species-specific and is acquired in utero; no new nephrons will grow postnatally (Merlet-Benichou, Gilbert et al. 1999).

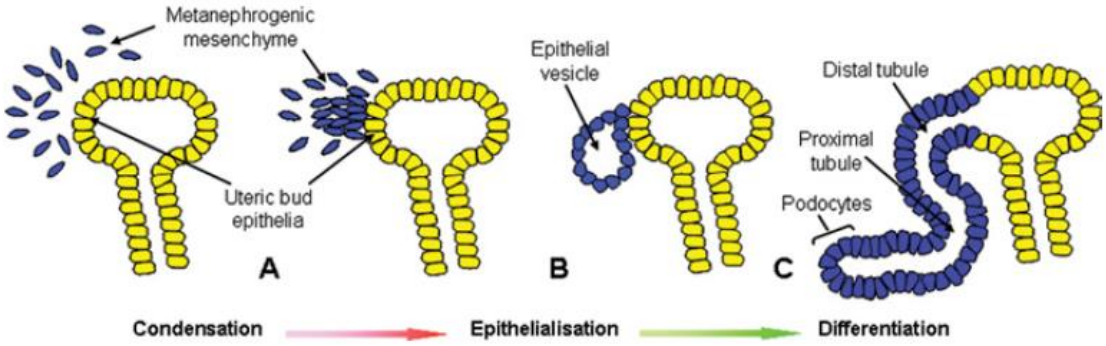


Figure 1.5: Reciprocal induction between the ureteric bud and the metanephric mesenchyme results in nephron formation. Taken from Burke & Toshe (Burke and Tosh 2005). (**A**) The ureteric bud (yellow) and the metanephric mesenchyme (MM, blue) undergo reciprocal induction upon contact. (**B**) Epithelialisation of the MM, (**C**) progression into a nephron connecting via the distal end to the ureteric bud which forms the collecting duct system.

Several genes, which can be considered markers of renal stem cells, have been shown to be specifically expressed in MM cells and are essential for their differentiation: *PAX2* (Torres, Gomez-Pardo et al. 1995), *WT1* (Kreidberg, Sariola et al. 1993), *EYA1* (Xu, Adams et al. 1999), *SIX2* (Brodbeck, Besenbeck et al. 2004), *HOX11* paralogous group (Wellik, Hawkes et al. 2002), *CITED1* (Boyle, Misfeldt et al. 2008) and *SALL1* (Nishinakamura 2003).

1.6.2 Disrupted development: nephrogenic rest formation

In very rare cases, normal nephrogenesis is disrupted and lesions termed nephrogenic rests (NRs) are formed. These morphologically resemble EK with immature nephron structures. Apart from the different physical location of PLNRs and ILNRs, there are other morphological differences. PLNRs are often numerous and diffuse whereas ILNRs are usually observed singularly. Microscopic differences include variation in the rest margin; well-defined and smooth in PLNRs (Figure 1.6) but irregular and often indistinct in ILNRs (Figure 1.7) and rest composition; PLNRs are predominantly blastemal in early lesions but develop epithelial structures or sclerosis with age and ILNRs are stromal predominant with the presence of occasional blastema and epithelia and often mature fat (Beckwith 1998).

NRs can follow a variety of fates, not all resulting in transformation to WT. Neoplastic NRs are considered as likely WT precursors whereas hyperplastic NRs are considered relatively stable despite their generalised growth. NRs can also remain dormant or may regress (by apoptosis) resulting in their obsolescence (Beckwith, Kiviat et al. 1990). Occasionally, the entire kidney is affected by 'nephroblastomatosis', the presence of diffuse NR. As 60% WT are not visibly associated with NR, it can be speculated that the evolving tumour may arise within a NR causing infiltration of the NR and loss of its identification in the subsequent nephrectomy specimen following tumour excision.

It is unclear what causes NR retention or persistence and as they are identified only from histological examination of formalin fixed paraffin embedded (FFPE) tissue, downstream analysis is restricted to methods that support analysis on such tissue. Because of this, genomic, transcriptomic and epigenomic data is limited to studies

assessing few genes or with small sample sizes and therefore the transformation steps required for tumour formation are unknown. Furthermore, although this method of detection is used diagnostically, pathologists have to deal with the major difficulty of distinguishing between a true NR and chemotherapy-treated WT, both of which look morphologically very similar.

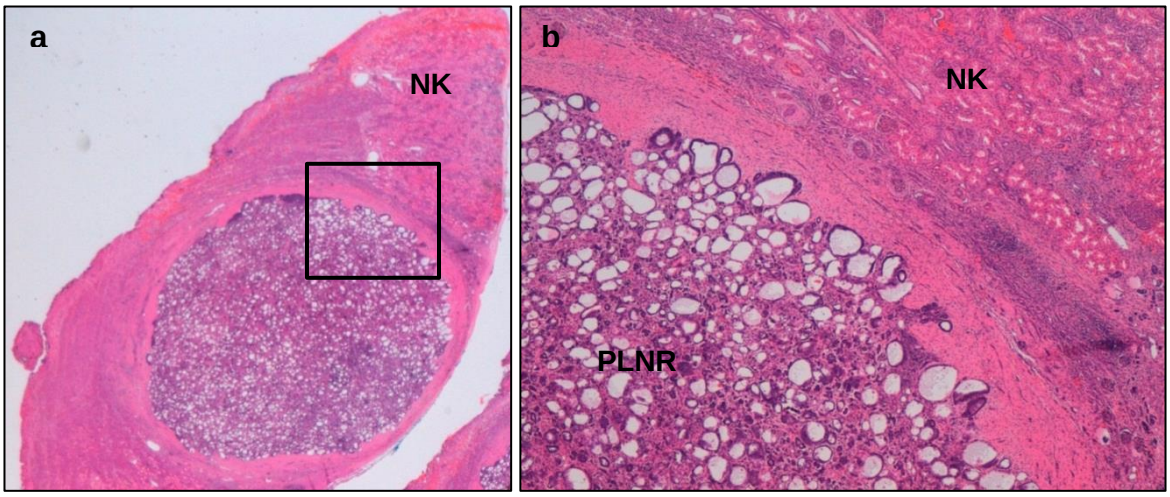


Figure 1.6: A perilobar nephrogenic rest surrounded by normal kidney (NK). (**a**) A 3μm section of FFPE nephrectomy sample stained with haematoxylin and eosin (100x). (**b**) magnification (400x) of the highlighted region in **(a)**. A well-defined smooth margin is seen between the PLNR and NK tissue. The PLNR tissue shows primitive epithelial structures within sclerotic stroma. In comparison, the NK shows tubules with well-defined lumen and mature glomeruli. The large quantity of glomeruli within this section of NK implies its location near the periphery of the lobule, where PLNR are common.

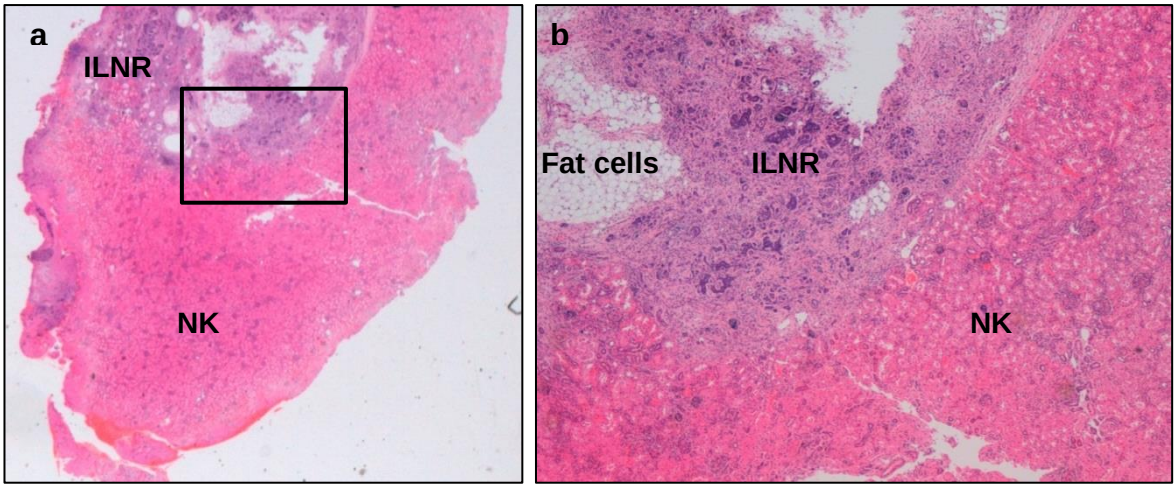


Figure 1.7: Intralobar nephrogenic rest surrounded by normal kidney. (**a**) A 3µm section of FFPE nephrectomy sample stained with haematoxylin and eosin (100x magnification). (**b**) The area highlighted in **(a)** at greater magnification (400x) showing the ILNR merging with the NK tissue at its border (indistinct rest margin). Within the ILNR blastema and immature epithelial structures with relatively large nuclei and scant cytoplasm are observed in a large amount of stroma. Regions that have differentiated into fat cells are also seen. The NK shows epithelial cells with large cytoplasm forming tubules with well-defined lumen. The reduced density of glomeruli within this section of NK implies its location near the central medulla, where ILNR are common.

1.6.3 Supporting molecular data: expression analysis

Several studies have used array-based methods to study mRNA expression in WT and found that expression profiles appear similar to those in EK (Li, Guo et al. 2002). In particular, of the transcription factors expressed in WT, *PAX2*, *EYA1*, *HBF2* and *HOXA11* are all key genes essential for MM proliferation in early renal development. Furthermore, expression of *SIX1* and *SALL2* was found to be high in WT and both genes are associated with MET but their expression decreases as cells further differentiate (Li, Guo et al. 2002). Taken together, this suggests that the cell-of-origin is the MM that has committed to MET and arrested development shortly

afterwards. This provides an explanation for the triphasic histology observed in WT which encompasses all cell-types seen during normal MM differentiation.

The same study identified genes expressed in WT that are not associated with EK including Wnt receptor *FZD7*, tumour suppressor gene *PRAME* and imprinted gene *NNAT* (Li, Guo et al. 2002). Further to this, overexpression of imprinted gene *IGF2* and loss of expression of imprinted *H19* is reported in ~74% WTs (Scott, Murray et al. 2012) suggesting an association between aberrant DNA methylation at imprinted loci and Wilms tumourigenesis. This is further supported by observed overexpression of imprinted genes *DLK1*, *RTL1*, *MEG3* and *MEST* in WT (Hubertus, Lacher et al. 2011).

A recent study with unexpected results gave insight into the properties of tumours formed when disruption in renal development occurred. Researchers constructed embryonic stem cells (ESCs) where expression of reprogramming factors (Oct3/4, Sox2, Klf4 and c-Myc) was inducible by addition of Doxycycline (Dox) (Ohnishi, Semi et al. 2014). From these, chimeric mice were generated. Mice treated with Dox for a short period of time (4-7 days) developed dysplastic lesions that resolved once Dox treatment was removed. Mice treated for longer periods of time (at least 7 days) developed tumours that were sustained once treatment was removed, including renal tumours that remarkably resembled WT including showing expression of Six2, Eya1 and Lgr5 (Ohnishi, Semi et al. 2014). These tumours also showed histological features of WT (the triphasic morphology) as well as Igf2 overexpression with gain of methylation at H19. Importantly, no genetic mutation in Wt1, Wtx, Ctnnb1, Trp53 or a panel of cancer-related genes (514 in total) was found and no prevalent chromosomal alteration, indicating this cancer phenotype could be entirely

epigenetic or could be associated with genetic features outside the exonic regions sequenced. The authors postulated that in human WT, a reprograming process might cause cell-fate conversion into progenitor-like states generating NRs that transform to WT.

1.7 Genetic features of Wilms tumour

The key genes mutated in WT are *WT1* (12%), *WTX* (32%), *CTNNB1* (15%), *TP53* (5%) and FBXW7 (4%) (Williams, Al-Saadi et al. 2010; Scott, Murray et al. 2012). Most cases of *WT1* mutation overlap with *CTNNB1* mutation whereas *WTX* mutation is more mutually exclusive (Figure 1.8). These common mutations were mostly discovered through association with familial disease or predisposition syndromes. As of yet, limited genome-wide mutational analyses have been published for WT however preliminary data (through personal communication) indicates mutations in *SIX2* and *MYCN* are common in WT. The only published study to report the results of genome-wide sequencing in WTs identified common mutations in *DROSHA* (12%) and other mutations in genes of the miRNA processing pathway (Torrezan, Ferreira et al. 2014).

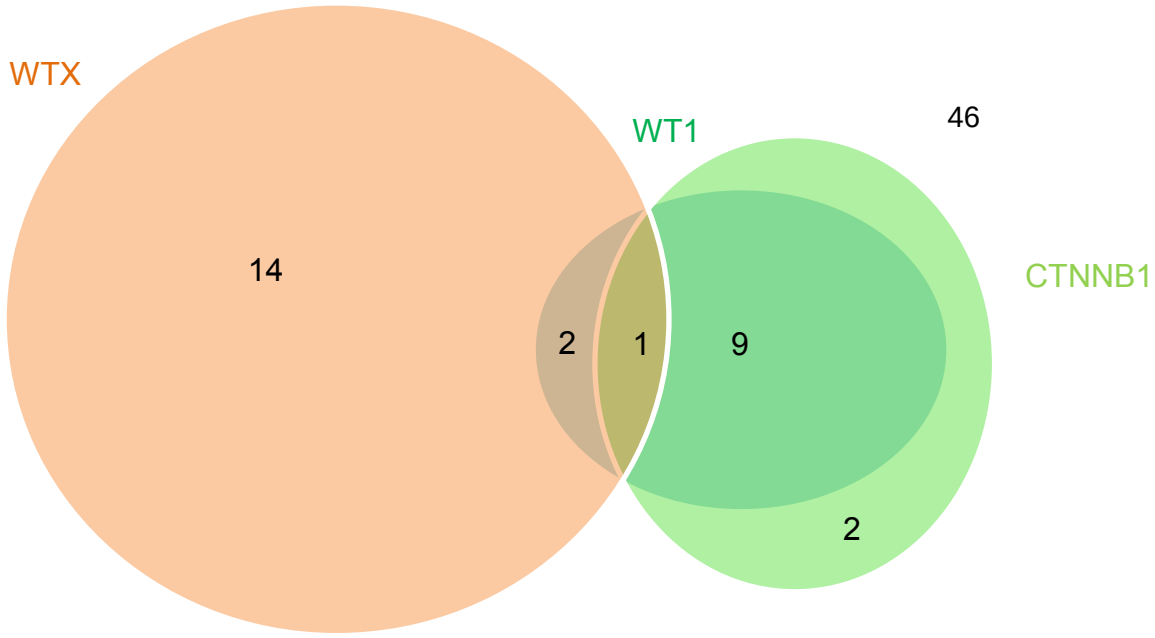


Figure 1.8: Mutational analysis of WTX, WT1 and CTNNB1 in WT (n=74). Venn diagram showing the proportions frequency of mutations in the most commonly mutated genes (WTX; orange, WT1; blue and CTNNB1; green) in a total of 74 WT where mutational analysis was carried out (from Scott et al (Scott, Murray et al. 2012)). 46 WT were poly-negative for mutation in all three genes.

WT1 encodes a zinc finger DNA binding protein that acts as a transcription factor mediating nephrogenesis with many target genes (Moore, McInnes et al. 1999). WT1 is expressed in the uninduced MM of the developing kidney (Mundlos, Pelletier et al. 1993) and is linked to survival and differentiation of the MM (Kreidberg, Sariola et al. 1993). WT1 knockout mice fail to form functional MM and fail to grow the ureteric bud resulting in embryonic lethality (Kreidberg, Sariola et al. 1993). *WT1* mutation is considered an 'early' event in WT as mutations are also seen in associated ILNRs (Fukuzawa, Heathcott et al. 2007).

CTNNB1 encodes β-catenin, a key protein involved in the Wnt signalling pathway, which is critical for epithelialisation post MET (Koesters, Ridder et al. 1999). *CTNNB1* knockout within renal progenitor cells in mice blocked renal vesicle formation and overexpression bypassed the requirement for signalling molecules Wnt9b and Wnt4 but also failed to form epithelial structures (Park, Valerius et al. 2007). In WT, the mutation is activating and therefore overexpression of β-catenin is seen which may therefore be associated with the immature nephron structures seen in NR and WT tissue. *CTNNB1* mutations have been defined as a 'late' event in WT tumourigenesis as they are present in WT but not NR, although only a small patient cohort was studied (Fukuzawa, Heathcott et al. 2007).

Of 74 WTs where mutational status for *WTX*, *WT1* and *CTNNB1* was determined, 88% tumours with either *WT1* or *CTNNB1* also had mutation in the other gene whereas only 5.8% *WTX* mutant tumours also had *CTNNB1* mutation and 17.6% had *WT1* mutation (Scott, Murray et al. 2012).

WTX (Wilms tumour on chromosome X, also known as *AMER1*) links both the *WT1* and *CTNNB1* genes together as the protein negatively regulates Wnt signalling (Major, Camp et al. 2007) and has been associated with WT1 transcriptional control (Rivera, Kim et al. 2009). In WT, *WTX* is mutated or deleted in around 30% tumours and has been described as a tumour suppressor gene. *WTX* is expressed during normal development in the MM and *WTX* knockout mice show defects in tissue derived from embryonic mesenchyme (Moisan, Rivera et al. 2011) with 64% of *WTX* null mice having uni- or bilateral kidney agenesis.

TP53 mutation or loss of copy number occurs in a low proportion of WT, but is particularly associated with WT that show anaplasia. These tumours show

characteristic copy number profiles with multiple focal areas of gain or loss which are generally absent in tumours with wild-type *TP53* showing that *TP53* loss associates with genomic instability; these cases often have a poor prognosis (Bardeesy, Falkoff et al. 1994).

FBXW7 is an E3 ubiquitin ligase that is deleted or mutated in ~4% WT. MycN, a target of FBXW7 ubiquitination and degradation, shows copy number gain in a subset of WTs (Williams, Al-Saadi et al. 2010), with several others showing activating mutation (unpublished personal communication). This suggests a common pathway is being targeted.

A further pathway with common mutations is the miRNA processing pathway with mutations in *DROSHA*, which encodes a protein with an RNaseIII domain for cleaving pri-miRNA into pre-mRNA, found in 12% WT (26/222). In total, 81% of these mutations cause E1147K mutation in the protein which alters a metal binding residue of the RNase IIIb domain. Furthermore, non-recurrent mutations in other genes in this pathway (*DGCR8*, *DICER1*, *XPO5* and *TARBP2*) were identified, indicating the significance of silencing the miRNA processing machinery in Wilms tumourigenesis.

Besides genetic mutations, common chromosomal alterations in WT occur at 16q and 1p (loss of heterozygosity), 1q (copy number gain), 17p (copy number loss; affecting TP53), 2p (copy number gain; affecting MycN) and 11p15 and 11p13 (loss of heterozygosity) (Grundy, Telzerow et al. 1994; Hing, Lu et al. 2001; Williams, Al-Saadi et al. 2010). Of these, 1q gain and 1p with 16q loss have shown prognostic significance associating with adverse outcome in favourable histology WT in more than one study (Hing, Lu et al. 2001; Grundy, Breslow et al. 2005; Messahel,

Williams et al. 2009; Gratias, Jennings et al. 2013). Further patient-specific chromosomal alterations were studied to discern the relationship between WT and its associated NR (Vuononvirta, Sebire et al. 2008). Gain of chr12, chr13 chr18 and loss of chr1p were observed in the NRs and termed ‘early’ events which were also seen in the associated WT. However in several cases, large scale genomic changes were observed in the precursor lesion but not the tumour, which showed different changes. This data supports the lesion as a precursor however reinstates that not all NR undergo malignant transformation.

1.8 Molecular biomarkers

There are several characteristic molecular markers of WT that either act as outcome-associated markers, markers of the disease or are associated with specific aspects of WT pathology or treatment (Table 1.2).

Association	Biomarker
Relapse	CACNA1E overexpression/amplification, specific to favourable histology WT (COG) (Natrajan, Little et al. 2006) IGF1R gain/overexpression in blastemal cells (Natrajan, Reis-Filho et al. 2006)
Poor prognosis/adverse outcome	1p and 16q loss (Grundy, Telzerow et al. 1994; Grundy, Breslow et al. 2005) Gain of 1q (Hing, Lu et al. 2001; Natrajan, Williams et al. 2006; Gratias, Jennings et al. 2013) Increased telomere activity (Dome, Bockhold et al. 2005) TrkB levels(Eggert, Grotzer et al. 2001)
WT diagnosis	p73 promoter methylation and p73 mRNA expression in peripheral blood (increase in WT patients compared to controls) (Song, Yue et al. 2013) Serum SAA1 and APO C-III levels (Wang, Wang et al. 2012)
Histology-associated markers	Gain of *MYCN* in unfavourable histology (Williams, Al-Saadi et al. 2010; Williams, Al-Saadi et al. 2011) miRNA profile expressed in high risk blastema (Watson, Bryan et al. 2013) TP53 mutation in anaplastic WT (Bardeesy, Falkoff et al. 1994) Reduced retinoic acid pathway activity in high risk tumours (Wegert, Bausenwein et al. 2011)

Table 1.2: Wilms tumour biomarkers

1.9 Common pathways

Combining the known genetic and epigenetic aberrations discovered in WT, it is clear that several pathways are commonly targeted in WT aetiology including the Wnt signalling pathway, IGF pathway and retinoic acid pathway.

1.9.1 Wnt signalling pathway

CTNNB1 mutations cause constitutive activation of the Wnt signalling pathway which is critical for reciprocal induction of the ureteric bud and MM as well as MET (Koesters, Ridder et al. 1999). Upon signalling by Wnt ligands through frizzled receptors, β-catenin is stabilised and localises to the nucleus where it forms a complex with LEF and TCF transcription factors to switch on expression of target genes. In the absence of Wnt signals, β-catenin is phosphorylated at multiple sites by CK1A, CK1E and GSK3β which enables recognition, ubiquitylation and hence degradation by proteasomes (Huelsken and Behrens 2002). The protein dishevelled is responsible for maintaining β-catenin in the presence of WNTs by inhibiting GSK3β activity. β-catenin has also been associated with renewal of stem cell pluripotency (Reya, Duncan et al. 2003; Sato, Meijer et al. 2004; Sokol 2011).

Epimutation at 5q31 (hypermethylation of the protocadherin cluster) resulted in loss of expression of protocadherin genes and reciprocal increase in Wnt signalling target gene expression (Dallosso, Hancock et al. 2009) indicating that other mechanisms may increase Wnt signalling in a larger subset of WTs. Despite its relatively common mutation rate, there has been little research undertaken into therapeutic targeting of the Wnt signalling pathway in WT. However, in one study, using an

antibody to Wnt receptor Frizzled-7 (FZD7) induced apoptosis in a primary WT cell lines that expressed FZD7 suggesting a subset of WT may respond well to such therapies (Pode-Shakked, Harari-Steinberg et al. 2011).

1.9.2 IGF pathway

IGF2, as mentioned, is implicated in the overgrowth syndrome BWS. In normal tissue, this gene is maternally imprinted causing monoallelic expression from the paternal allele. Overexpression can result from either loss of this imprinting mark or paternal uniparental disomy (pUPD). The mechanisms behind regulation of imprinted gene expression will be discussed later in chapter 1.10.3.3. The overexpression of *IGF2* results in activation of the insulin signalling pathway. Binding of insulin, IGF1 or IGF2 to the extracellular portion of the insulin receptor (IR), IGF1R or hybrid receptor leads to autophosphorylation of the β-subunit tyrosine kinase, followed by the phosphorylation of additional tyrosine residues. This leads to recruitment of insulin receptor substrates (IRS)1 to IRS4 and other proteins, allowing activation of the PI3K and mitogen activated protein kinase (MAPK) signalling pathways which lead to unregulated protein synthesis, cell cycle progression and cell growth, and prevention of apoptosis (Gallagher and LeRoith 2010).

In terms of targeting this pathway in WT, the IGF1R inhibitor BMS-754807, an ATP-competitive small molecule, was tested *in vivo* using xenograft models of WT and demonstrated significant growth delay (Kolb, Gorlick et al. 2011). The anti-IGF1R monoclonal antibody IMC-A12 was also used to treat the same WT models

but showed very limited efficacy (Houghton, Morton et al. 2010). In WT cell lines studied in the kidney environment, the use of the *IGF1R* inhibitor NVP-AEW541 resulted in growth inhibition associated with down-regulation of PI3K and MAPK pathways and down-regulation of cell cycle control genes *CCNA2* and *CCNB1*, with drug efficacy dependent on the levels of phosphorylated *IGF1R* (Bielen, Box et al. 2012). Targeting this pathway may be improved with combination therapy however it seems the cells are not dependent on it for survival so it may not be the best therapeutic target.

1.9.3 Retinoic acid pathway

A study of cDNA in WT identified several genes of the retinoic acid (RA) pathway (*RARRES2*, *RARRES3* and *CTGF* and also RA pathway targets *NK4*, *RAMP*, and *ENPP2*) that were downregulated in relapsed tumours (Zirn, Hartmann et al. 2006). This was further validated in a second study where RA pathway down regulation was seen in high risk WT (Wegert, Bausenwein et al. 2011). Retinoic acid (RA) is the biologically active form of vitamin A and binds to the retinoic acid receptor (RAR) which binds in the membrane to the retinoic X receptor (RXR) in regions called retinoic acid response elements (RAREs). Upon activation of RAR by binding RA, the classicial RA pathway induces differentiation, cell cycle arrest and eventual apoptosis.

Treating WT cells with retinoids or all-trans-RA (ATRA) to activate RA signalling showed positive effects including reduced proliferation, apoptosis however these

were readily reversible upon drug withdrawal (Zirn, Samans et al. 2005; Wegert, Bausenwein et al. 2011).

1.9.4 miRNA processing pathway

miRNAs are encoded for in human DNA. Expression of mature miRNA begins with RNA polymerase proteins transcribing the sequences into pri-miRNA which forms a long double-stranded hairpin which is then cleaved by a complex of Drosha and DGCR8 (termed Pasha) into a smaller pre-miRNA hairpin which is exported from the nucleus. They are then cleaved by Dicer (an RNase) and TRBP (with specificity for double stranded RNA) to remove the hairpin loop and leave two single stranded miRNAs. The functional strand binds to Argonaute (Ago2) proteins into the RNA-induced silencing complex (RISC) where it guides the complex to its target mRNA while the non-functional strand is degraded. Targeting of mRNAs by this method results in mRNA silencing by mRNA cleavage, translational repression or deadenylation.

Mutations in WTs have been found in genes *DROSHA*, *DGCR8*, *DICER1*, *XPO5* (a nuclear exporter of pre-miRNA) and *TARBP2* (required for RISC formation) (Torrezan, Ferreira et al. 2014), (Wu, Sabbaghian et al. 2013). WTs with mutations in this pathway showed downregulation of mature miRNAs as expected, and not in the primary miRNA transcripts indicating the machinery is failing to generate mature mRNA.

Further showing the importance of miRNA regulation in WT formation is the finding that overexpression of *Lin28* in renal precursor cells generates proliferative

WT-like lesions in mouse kidneys. Lin28 inhibits biogenesis of Let-7 family miRNAs, which are key determinants of pluripotency, differentiation and transformation (Urbach, Yermalovich et al. 2014). Furthermore, overexpression of Lin28 was found in the blastemal elements of human WT in comparison to normal kidney (NK) (Urbach, Yermalovich et al. 2014) and activating mutations in *DIS3L2*, which degrades Lin28-mediated uridylated pre-let-7 miRNAs, have been identified as predisposing to WT (Chang, Triboulet et al. 2013). Therefore, both LIN28 overexpression and activation of DIS3L2 result in loss of mature Let-7 miRNAs and are associated with causing WT. The full paper describing Lin28 overexpression in mouse and human WTs, to which I contributed, can be found in Appendix 2.

miRNA species play a major role in many human cancers by affecting expression of miRNA target genes (Calin and Croce 2006). This occurs by loss, amplification or mutation of the miRNA DNA sequence and can be associated with cancer predisposition (if germline) or progression (if somatic). In general, a down-regulation of miRNA is seen in tumour compared to normal tissue (Lu, Getz et al. 2005). miRNA expression signatures are also a common marker of solid cancers (Iorio, Ferracin et al. 2005; Volinia, Calin et al. 2006; Lee, Gusev et al. 2007). In WT, as this is a relatively recent discovery, no therapeutic avenues for targeting miRNA processing machinery have been explored.

1.10 Overview of epigenetics

An overview of epigenetics is given at this point due to the significance of WT being a developmental tumour as well as the association of specific epimutations

with WT, a topic that will be discussed in detail later. Epigenetics describes heritable modifications to DNA or chromatin that affect gene expression without changing the nucleotide sequence. Every cell type has a unique epigenome that gives cell identity. Unlike the genome, the epigenome can be remodelled at any stage; however in healthy, terminally differentiated cells it is stable. Epigenetic gene regulatory mechanisms can be broadly classified into covalent chromatin modification, DNA cytosine methylation and non-coding RNA.

1.10.1 Chromatin remodelling

Histones are responsible for maintenance of the structure of chromosomal DNA. Gene expression can be affected either by expression of histone protein variants or by modification of the histone amino acids. Within the histone complex, DNA wraps around histone proteins H2a, H2b, H3 and H4. Histone protein H1 binds like a clip on the outside of this structure. Each protein has a subfamily of proteins with multiple members that can combine together to form a nucleosome. Variants of these proteins exist for example H3.3, which is associated with actively transcribed genes (Ahmad and Henikoff 2002), H2A.Z which flanks both active and inactive gene promoters (Raisner, Hartley et al. 2005) and is involved in the prevention of the spread of silent heterochromatin (Meneghini, Wu et al. 2003) and H2A.X which binds at DNA double-stranded breaks marking regions undergoing repair (Paull, Rogakou et al. 2000).

Figure 1.9: Histone 3 tail with the sites of modification highlighted, taken from (Sims and Reinberg 2008). Me = methylation; red = repression, green = activation, black = not associated with transcription, P = phosphorylation; orange, Ac = acetylation; blue.

Histone tails can be post-translationally modified by acetylation or methylation of lysine (K) and arginine (R), phosphorylation of serine (S) and threonine (T), ubiquitylation and symoylation of lysine and ribosylation (by example, the N-terminal tale of histone 3 is shown in Figure 1.9). These modifications alter the interaction between histones and DNA. Suppressive modifications increase binding affinity and result in DNA being more tightly 'packed' into condensed chromatin whereas modifications that decrease affinity allow for open DNA structures that can be accessed by transcription factors. The study of histone modifications in various cell types has generated an overall picture of their regulatory effect on gene expression (Table 1.3) (Mikkelsen, Ku et al. 2007; Pan, Tian et al. 2007; Heintzman, Hon et al. 2009; Aiden, Rivera et al. 2010; Creyghton, Cheng et al. 2010; Ernst, Kheradpour et al. 2011; Martens, Stunnenberg et al. 2011; Rada-Iglesias, Bajpai et al. 2011).

Locus	Histone modification	Gene expression
Active promoter	H3K4me3 or H3K27ac	On
Active enhancer	H3K4me1 or H3K27ac	On
Actively transcribed genes	H3K36me3	On
Repressed DNA	H3K27me3 or H3K9me2/3	Off
Bivalent domain	H3K4me3 or H3K27me3	Off, but 'poised' for expression or repression

Table 1.3: Chromatin modifications and their effect on gene expression. H3 = histone 3, K = lysine, me2 = dimethylation, me3 = trimethylation, ac = acetylation.

Overlapping H3K4me3 (Histone 3, Lysine 4 trimethylation) and H3K27me3 mark bivalent domains in ESCs (Pan, Tian et al. 2007). These domains show both the oppressive and active mark and appear at developmental loci poised for expression upon differentiation; these regions have been shown to be specifically disrupted in cancer (Pan, Tian et al. 2007; Doi, Park et al. 2009).

Histone modifications are flexible and enzymes exist to add or remove marks depending on cell requirements; for example histone acetylation is controlled by histone acetyltransferase (HAT) which adds an acetyl group to histones uncoiling the chromatin and facilitating gene transcription and is reversed by histone deacetylase (HDAC) which removes acetyl groups from histones coiling the chromatin and inhibiting gene transcription. Histone K methylation is also enzymatically reversible and is 'written' by K methyltransferases (KMTs) and 'erased' by K demethylases (KDMs). Aberrant modifications may lead to a chromatin conformational change

and therefore interfere with transcription of genes, including tumour suppressor genes and oncogenes.

This plasticity is essential during development where histone modifications are known to instruct transcriptional networks required for cellular differentiation. In particular, histone plasticity and not DNA methylation is essential for TGF-β-induced epithelial to mesenchymal transition (EMT) where repressive H3K9Me2 domains were shown to reduce upon EMT while transcription-associated bulk H3K36me3 and H3K4me3 domains were increased (McDonald, Wu et al. 2011). However, the mechanistic links that drive early-fate decisions as progenitor cells differentiate and its disruption that results in tumour formation are still unclear.

1.10.2 miRNA silencing

As previously described, miRNAs are small noncoding RNAs that regulate gene expression by binding to the 3' untranslated region of messenger RNA (mRNA) causing its degradation and preventing protein translation (Bartel 2004). Eukaryotes express hundreds of miRNAs that regulate thousands of mRNAs and their expression is cell and tissue specific and essential for differentiation.

1.10.3 DNA methylation

DNA methylation refers to the addition of a methyl group to carbon-5 position of cytosine residues. This most frequently occurs at CpG residues however non-CpG methylation has also been observed, commonly at TACAG motifs, with

high levels in pluripotent cells and a decrease upon differentiation with almost no non-CpG methylation in fully differentiated cell types (Chen, Feng et al. 2011; Ziller, Muller et al. 2011).

DNA methylation at CpG residues in healthy cells has a multitude of roles including silencing of transposable elements, aiding defence against viral DNA and determining gene expression, including the expression of imprinted genes so that only one allele is expressed in a parent-of-origin specific manner. Aberrant CpG methylation can be a marker of a particular disease state or can contribute to a disease by altering gene expression as methylation can cause steric hindrance for transcription factor binding.

1.10.3.1 Addition and removal of DNA methylation

DNA methylation is controlled by a family of DNA methyl-transferase (DNMT) enzymes including DNMT1, DNMT3A, DNMT3B and DNMT3L. These enzymes transfer a methyl group from S-adenosyl-L-methionine (AdoMet) to the C5 position of cytosine. DNMT1 is a maintenance enzyme that functions to ensure replication of methylation marks to daughter cells. It targets newly replicated hemi-methylated DNA and replicates the methylation pattern to daughter DNA strands. DNMT3B is specifically required for methylation of centromeric minor satellite repeats (Okano, Bell et al. 1999). DNMT3A has *de novo* methyl-transferase activity during embryogenesis working with its cofactor DNMT3B. Non-CpG methylation in embryonic stem cells is also thought to be mediated by DNMT3A (Ramsahoye, Biniszkiewicz et al. 2000; Ziller, Müller et al. 2011). Furthermore, DNMT3A

cooperates with DNMT3L during embryogenesis to establish maternal imprints at imprinting centres. (Bourc'his, Xu et al. 2001; Chédin, Lieber et al. 2002). DNMT3L itself has no *de novo* DNA methyltransferase activity and binds to DNA but not AdoMet however it stimulates *de novo* activity of other DNA methyltransferases by binding them directly via its C-terminal domain (Gowher, Liebert et al. 2005). This binding enhances catalytic activity of DNMT3A and DNMT3B ~15 fold and increases DNMT3A affinity for AdoMet binding and affinity for DNA binding 20-fold. This interaction is transient and DNMT3L acts as a substrate exchange factor in this process (Gowher, Liebert et al. 2005). *DNMT* overexpression is associated with DNA hypermethylation and loss of imprinting (Biniszkiewicz, Gribnau et al. 2002) in several cancers (Arai, Kanai et al. 2006; Lin, Hsu et al. 2007; Roll, Rivenbark et al. 2008).

Polycomb group proteins are a group of proteins that act to silence genes involved in development. Certain Polycomb group proteins, for example EZH2, directly affect DNA methylation by recruiting DNMT enzymes to target sites (Vire, Brenner et al. 2006).

The reciprocal process, loss of DNA methylation, is mediated by ten-eleven translocase enzymes (TET1-3) which target 5-methylcytosine and oxidise it into 5-hydroxymethylcytosine (Wang, Williamson et al. 2007; Tahiliani, Koh et al. 2009). TET enzymes can also perform iterative oxidation generating 5-formylcytosine and 5-carboxylcytosine (Ito, Shen et al. 2011). Unmodified cytosines are regenerated by passive dilution (replication of the DNA strand without copying the modification), by direct removal of the oxidised 5-position or by DNA repair-mediated excision by the enzyme thymine DNA glygosylase (TDG) (Kohli and Zhang 2013). Mutations in

TET genes are frequently found in myeloid cancers. One study found that patient bone marrow had significantly lower levels of 5-hydroxymethylcytosine compared to healthy controls and the authors suggested that disruption in TET2 activity favoured myeloid tumourigenesis (Abdel-Wahab, Mullally et al. 2009; Ko, Huang et al. 2010).

1.10.3.2 DNA methylation distribution

The distribution of CpGs throughout the genome is not uniform and is associated with gene expression. CpG islands (CGI) are calculated *in silico* as regions exceeding a set threshold of CG content (>55%) (Takai and Jones 2002) and are generally not methylated except for when located within a gene promoter. In general, gain of DNA methylation near a gene promoter silences gene expression by preventing transcription factor binding. It is now thought that CGI shores (located up to 2kb from CGI) hold much more significance in terms of tissue-specific gene expression due to their variable nature in both embryonic stem cells, normal tissue and in cancer (Doi, Park et al. 2009; Irizarry, Ladd-Acosta et al. 2009; Hansen, Timp et al. 2011). Several other features related to the distribution of genomic CpGs have recently been described (Table 1.4). In particular, global loss of DNA methylation in comparison to healthy tissue is a general property of cancer and is associated with loss of genomic stability as loss of tight epigenetic regulation has occurred. Conversely, aberrant gain of methylation, which usually occurs at CpG islands, has been seen in many cancers and is termed CpG island methylator phenotype, or CIMP (reviewed in (Teodoridis, Hardie et al. 2008)). CIMP is defined by methylation of multiple CpG islands, often near the promoter regions of tumour suppressor genes.

Contiguous regions of DNA that show differential methylation between tissues, cells or individuals are referred to as differentially methylated regions (DMRs); the methylation status of which may have a direct effect on a particular transcript. A small subset of DMRs are imprinting control regions (ICRs) which determine the expression of imprinted genes (discussed later). These are set in the germline and show allele-specific differential methylation whereas other DMRs are established after fertilisation to define gene expression in normal tissues and show biallelic methylation patterns that differ between tissues.

In general, DNA methylation prevents gene expression by preventing transcription factor binding, and therefore key tissue-specific transcription factors can be identified by studying the genomic regions that are not methylated in comparison to other tissues.

Feature	Significance
CpG island	CpG rich region, generally not methylated, show cancer-specific hypermethylation
CpG shore	2kb from CpG island, inversely correlated with gene expression (Doi, Park et al. 2009; Irizarry, Ladd-Acosta et al. 2009)
CpG shelf	2kb from CpG shore
Methylation within gene promoter region	In general, inversely correlated with gene expression
Methylation within gene body	In general, positively correlated with gene expression (Jjingo, Conley et al. 2012; Kulis, Heath et al. 2012)
DNA methylation valley	Large unmethylated fomains where early developmental regulatory genes are located. When these specific early developmental genes are silenced during cell differentiation, H3K27me3 is acquired instead of methylation (Xie, Schultz et al. 2013).
DNA methylation canyon	Regions of low methylation protected by DNMT3A that span conserved domains frequently containing transcription factors and are distinct from CpG islands and shores (Jeong, Sun et al. 2014).
Low methylated region (LMR)	Average methylation of 30%; poor CpG dense distal regulatory regions that bind transcription factors. This binding is sufficient to create LMRs. Cell-type LMRs are occupied by cell-type specific TFs.
Differentially methylated region (DMR)	Regions of contiguous methylation that have a different methylation status from the comparative tissue. These regions are defined by different algorithms depending on the dataset but usually contain at least 3 CpGs

Table 1.4: Features associated with genomic distribution of CpGs

1.10.3.3 Genomic imprinting

In a subset of genes termed imprinted genes, an epigenetic mark set in the germline restricts expression in a parent of origin dependent manner resulting in

hemizygous expression. In effect, one copy of an imprinted gene is actively silenced by DNA methylation depending on whether it was inherited from a paternal or maternal chromosome. Approximately 100 imprinted genes have been identified but only a small proportion of these experimentally validated (Morison, Paton et al. 2001). These imprinted genes, including *IGF2* and *H19*, are more susceptible to deregulation as only one copy is active.

The allele specific DNA methylation is controlled by DMRs which, as previously mentioned, are regions of differential methylation that effect transcript expression. A small selection of DMRs are imprinting control regions (ICRs) which dictate the methylation status of the surrounding DMRs in up to 1Mb of DNA and hence regulate the expression of the imprinted genes they control (Weaver, Susiarjo et al. 2009). Deletion of an ICR causes loss of imprinting (LOI) in adjacent genes resulting in biallelic expression (Fitzpatrick, Soloway et al. 2002).

Imprinting marks must be reset every generation as, for example in a female's oocytes, all paternally inherited chromosomes must be remodelled to be passed to her offspring as maternal imprints. For this to occur, all marks are erased within the germ cell. In the maternal gamete, methylation is lost passively by loss of *DNMT1* expression whereas in the paternal gamete, methylation remains until fertilisation (Smith, Chan et al. 2012). Upon fertilisation, before the two genomes combine, the paternal genome is actively demethylated by TET enzymes (Mayer, Niveleau et al. 2000). Primary imprinting marks at ICRs are then set in the germline (Mann 2001).

1.10.3.4 Analysis of DNA methylation and confounders

DNA methylation is commonly assessed by bisulfite conversion of DNA, which causes conversion of all unmethylated cytosine residues to thymine resides and leaves all methylated cytosines unchanged (see Figure 1.10). This allows distinction between methylated and non-methylated CpGs. After this treatment, DNA can be analysed by either array-based or sequencing-based methods.

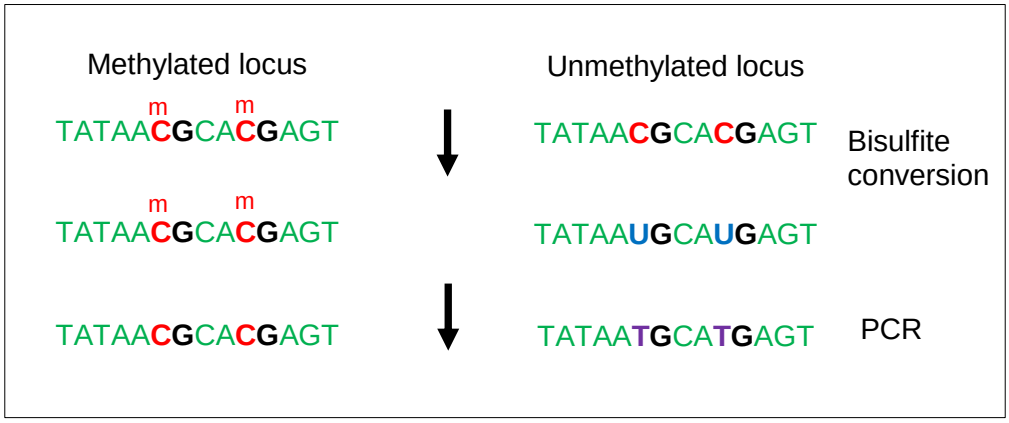


Figure 1.10: Bisulfite conversion of DNA to differentiate methylated from non-methylated cytosine residues Bisulfite treatment targets only unmethylated cytosines and deaminates them to uracil. Upon one round of PCR, these are substituted for thymine therefore permanently changing the unmethylated CpG so that it can be distinguished from a methylated CpGs.

With comparative DNA methylation analyses, there are many confounding factors including incomplete bisulfite conversion, batch/chip/bead effects (when using microarrays), age and ethnicity of the patient, SNPs and cell type composition. Cell composition effects have a marked effect on results, although until recently were difficult to account for. As each cell type has a unique epigenome, having different proportions of cell types in comparative tissue samples may result in false positive

results, or hide true positives. Likewise, in studies that assess methylation markers in blood, the many cell types present significantly confound results. This can be limited by cell sorting to single populations prior to analysis however this is not always possible as a definitive cell-specific marker may be unknown. Alternatively, novel computational algorithms have been designed to adjust comparative analyses for cell type composition effects.

One such algorithm was designed by profiling sorted subpopulations of blood and discerns the contribution of mixed cell type effects in blood case:control studies (Adalsteinsson, Gudnason et al. 2012), however this is the only method that uses such a rigourously defined reference (Jaffe and Irizarry 2014). Other current methodologies for correcting for cell-type composition include EWASher (Zou, Lippert et al. 2014) and RefFreeEWAS (Houseman, Molitor et al. 2014), both of which require no reference dataset. The RefFreeEWAS algorithm uses single value decomposition to estimate the number of cell types contributing to overall histology, then deconvolutes the β-values based on the estimated number of cell types and a design matrix specifying analysis covariates, and generates bootstrap-derived CpG-specific p-values and covariates that correspond to a 'true' methylation signal with no cell mixture effects.

Although cell mixture effects can now be taken into account, it is not entirely essential for certain study designs as a mixed cell composition, in some cases, is indeed a marker of the disease itself.

1.11 Biological epigenetics

As discussed, there are many different aspects of epigenetics, which work in unison during cellular differentiation in normal development and in disease.

1.11.1 Epigenetics in development

Waddington's theory of gene expression regulation during cellular differentiation in development used the analogy of a ball rolling down a hill. In this metaphor, the ball at the top of the hill represents the pluripotent stem cell which follows decisive pathways regulated by gene expression and hence travels downwards (through valleys) to become a fully differentiated cell (see Figure 1.11). As cells acquire cell-type specific features during differentiation, they cannot travel "up the hill". This early model, described prior to the understanding of epigenetics, has been adapted and refined since then, however today, the same hierarchical concept exists.

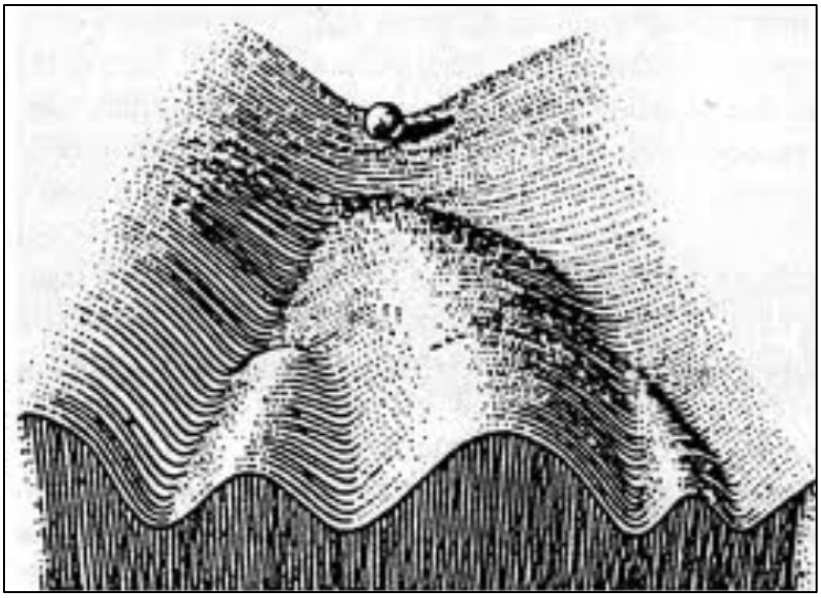


Figure 1.11: Waddington's hill theory. A pluripotent stem cell has the potential to differentiate into any cell in the organism. This cell (depicted as a ball) is shown at the top of a hill. As the ball travels down the hill, it runs into a valley. Each valley represents a change in the epigenetic architecture of the cell which corresponds to acquiring more differentiated properties, such as lineage specification. Once a decision has been made, the ball cannot roll upwards, or across to a different valley, but can only travel downwards taking one of many routes towards terminal differentiation.

This hierarchical concept, explained further by the understanding of epigenetics, centres around the fact that during development, the genome remains constant while the epigenome undergoes programmed remodelling to alter gene expression which is essential to define embryo patterning and for organ and cell-type specification. Cells acquire cell type-defining permanent epigenetic changes during differentiation, which specifies a lineage and then cell type. Consequently cells at an early stage of differentiation have a more variable and more easily reshaped epigenome than those nearing terminal differentiation, and once a cell fate decision has been made, cells cannot reverse the decision to alter, for example, the specified lineage. Upon

terminal differentiation, maintenance of the epigenome is essential to maintain cellular identity and the epigenome shows little plasticity.

1.11.2 Epigenetics in cancer

It is commonly observed that loci involved in regulation of normal development or differentiation are targeted by aberrant DNA methylation or chromatin structure in cancer compared to normal cells. As previously mentioned, during development the epigenome shows a great deal of plasticity, particularly at early stages. Cancer cells show the same properties either by proliferation of less well differentiated cells or by mutation of epigenetic regulators. Having the ability to alter the epigenome, gives a cell the ability to adapt to the local microenvironment and further insults (e.g. chemotherapy). Mutations in epigenetic regulators have recently been identified with the sequencing of many human cancers (van Haaften, Dalgliesh et al. 2009; Dalgliesh, Furge et al. 2010). In fact, mutations in a given epigenetic regulator can be activating or inactivating (Ernst, Chase et al. 2010; Nikoloski, Langemeijer et al. 2010; Sneeringer, Scott et al. 2010), suggesting tumour promoting or tumour suppressor functions for the same gene product depending on cellular context. Re-acquisition of epigenetic plasticity gives cancer cells self-renewal and proliferation capacity, similar to normal progenitor populations. Indeed, epigenetic events have been considered as the earliest events in tumourigenesis whereby Cancer has been described as originating from a pool of progenitor cells with unstable epigenomes that later acquire gene-specific genetic or epigenetic initiating mutations that result in tumour formation (Feinberg, Ohlsson et al. 2006). These tumours then acquire epigenetic and genetic plasticity as a later event which is

proposed to lead to tumour heterogeneity (Feinberg and Tycko 2004; Gerlinger, Rowan et al. 2012).

Despite the advantages of its plasticity during cellular differentiation, aberrations in the epigenome (epimutations) can impact on gene expression analogous to genetic mutations and can therefore confer undesirable effects. Differential gene expression has been attributed to epimutation in several cancer predisposition and congenital predisposition syndromes by overexpression of oncogenes or silencing of tumour suppressor genes. In cancer, aberrant hypomethylation causes expression of oncogenes, for example *WNT5B* in prostate cancer (Wang, Williamson et al. 2007), *R-RAS* in gastric cancer (Nishigaki, Aoyagi et al. 2005) and *MAGE1* in melanoma (Janssen, van de Locht et al. 1999) as well as global genomic instability in cancer whereas aberrant hypermethylation silences tumour suppressor genes for example *RB1* in retinoblastoma (Sakai, Toguchida et al. 1991), *p16* in melanoma (Gonzalez-Zulueta, Bender et al. 1995), *SLC5A8* in lung cancer (Park, Kim et al. 2013), *P16* in breast cancer (Lee, Ko et al. 2012) and *RASSF2* in cervical cancer (Guerrero-Setas, Perez-Janices et al. 2013). Hypermethylated tumour suppressor genes represent one of the most consistent hallmarks of human cancers (Merlo, Herman et al. 1995; Esteller, Corn et al. 2001; Esteller 2002).

Previous work has shown that bivalent domains are often targets of aberrant hypermethylation in cancer due to the vulnerability gained by going through transient chromatin-induced silencing during development (Ohm, McGarvey et al. 2007; Widschwendter, Fiegl et al. 2007). It is thought that the transient silencing is "locked in" by recruitment of DNA methyl transferases by the Polycomb proteins in cancer which deposit irreversible DNA methylation, which is the mechanism used in

normal development to silence these loci during cellular differentiation (Schlesinger, Straussman et al. 2007).

As well as common changes in methylation occurring at specific regions, an increase in stochastic variation in DNA methylation has been associated with tumourigenesis (Hansen, Timp et al. 2011). This variability occurs within common regions in different cancers; sites that can distinguish each respective normal tissue (Hansen, Timp et al. 2011) and is associated with giving cancer cells selective advantage through cellular Darwinian selection mechanisms (Issa 2011). In other words, cancer cells that display a mixture of epigenetic signatures are more likely to contain cells that have different properties such as chemotherapy resistance or metastatic or invasive potential, than a single clonal population. Supporting this hypothesis of methylation variability increase giving a selective advantage, an increase in methylation variability has been observed in colon adenocarcinoma precursor lesions and was seen to increase during tumourigenesis (Timp and Feinberg 2013). However, a more recent study in cervical cancer found that the peak of variability occurred just prior to malignant transformation as, at this stage, there is an absence of a dominant subclone typically seen in cancer (Teschendorff, Liu et al. 2014). Both studies agree that methylation variability is however cancer-associated. Taking this concept further, two studies were able to show that early changes in DNA methylation could be predictive of cancer generation. The first showed that DNA methylation signatures were more variable when taken from women who later developed cervical neoplasia in comparison to those that remained disease-free (Teschendorff, Jones et al. 2012), and the second found that white blood cell DNA methylation levels at an intragenic ATM locus could predict breast cancer risk (Brennan, Garcia-Closas et al. 2012).

Finally, the significance of the tight regulation of imprinted genes by DNA methylation is shown by the abundance of LOI seen in cancer, including LOI at *IGF/H19* in WT and colorectal cancer (Sakatani, Kaneda et al. 2005), *PEG1/MEST* in lung cancer (Nakanishi, Suda et al. 2004), *CDKN1* in pancreatic cancer (Sato, Matsubayashi et al. 2005), *DIRAS3* in breast cancer (Yu, Xu et al. 1999), and *TP73* in gastric cancer (Kang, Park et al. 2000). Problems with the establishment of coherent imprinted marks during development can cause germline syndromes such as BWS which increases susceptibility to WT (Reik, Brown et al. 1995).

1.11.3 Epigenetic cancer biomarkers: circulating cell-free DNA

As previously stated, epigenetic modifications can be markers of a particular disease state without necessarily contributing biologically to disease onset or maintenance. In particular, methylated DNA can act as a marker of a cancer which can be measured in the blood as tumours release DNA upon cell death. In other cancers, methylated circulating cell-free DNA has been extracted from blood serum or plasma and used to quantify tumour burden and response to chemotherapy (Mori, O'Day et al. 2005; Hauser, Kogej et al. 2013; Hauser, Zahalka et al. 2013; Ponomaryova, Rykova et al. 2013; Tian, Yip et al. 2013; Balgkouranidou, Chimonidou et al. 2014).

1.11.4 Epigenetics in Wilms tumour

Analysis of several components of the epigenome has been undertaken in WT including analysis of global chromatin modification, miRNA analysis and targeted DNA methylation analysis. These studies revealed epigenomic architecture associated with renal development and with cancer giving novel insight into WT biology.

1.11.4.1 Chromatin

Recently, chromatin immunoprecipitation was performed on three WT, EK and normal kidney (NK) tissues using antibodies against H3K4me3 (marking transcriptional initiation when present at promoters), H3K36me3 (marking active genes) and H3K27me3 (marking silent promoters) (Aiden, Rivera et al. 2010). The DNA captured from these chromatin selections was sequenced to identify active and suppressive loci and compared to gene expression in WTs and chromatin maps in ESCs.

Both ESCs and WTs showed similar patterns of H3K4me3, specifically at developmental loci where both cell types showed broad regions of H3K4me3 that, in ESCs, mark key master transcription factors such as *OCT4* and *SOX2*. One of these large domains overlapped with transcription factor *SIX2*, which is essential for renal progenitor maintenance, and has been shown to be overexpressed in both the MM and WTs. In total, 114 genes showed broad H3K4me3 domains present in WT but not NK that encompassed many genes essential for renal development, suggesting WTs are governed by a set of transcription factors essential for normal development.

The same genes also showed enrichment for H3K36me3 indicating positive expression.

Comparison of narrow H3K4me3 peaks, which mark active promoters, identified a set of genes enriched for epigenetic regulators that were positively expressed in ESCs and WT but not expressed in NK (Figure 1.12). As terminally differentiated NK cells will not need to remodel their epigenome, but ESCs need to for cellular differentiation, this evidence suggests that WTs have the potential to actively modify their epigenome for further differentiation, hence why a variety of cell types is observed in WTs.

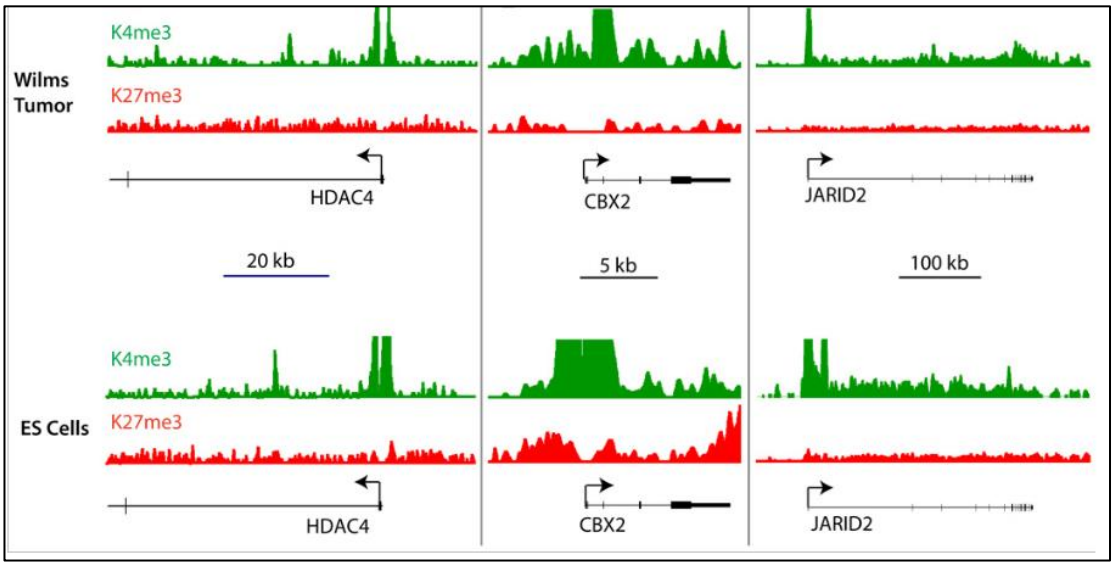


Figure 1.12: Epigenome remodellers are active in WT and ESCs but not normal kidney, taken from (Aiden, Rivera et al. 2010). Peaks identified by chromatin immunoprecipitation using an antibody for H3K4me3 and H3K27me3 are shown in green and red respectively. Three epigenetic regulators are shown, *HDAC4*, *CBX2* and *JARID2*. In both WT and ESCs, H3K4me3 peaks can be seen but not H3K27me3 suggesting that these genes are positively expressed in both tissues. Normal kidneys did not show H3K4me3 peaks at these genes (not shown).

Next, analysis of combined H3K4me3 and H3K27me3, or 'bivalent domains' in WTs identified a subset of genes with low expression in WTs that were also bivalent and not expressed in ESCs but were active in normal kidney. In ESCs, bivalent domains mark developmental loci that are considered essential for normal development and are resolved during cellular differentiation by losing either chromatin mark therefore causing the underlying gene to become either expressed or repressed. Unsurprisingly, a subset of these genes are strongly expressed in mouse kidney epithelial structures indicating that cells are prevented from expressing key epithelial genes and therefore differentiating properly. As other domains that are bivalent in ESCs showed H3K27me3 and no expression in WTs, the authors suggest that WTs show partial lineage commitment and concluded that developmental arrest occurs in undifferentiated MM as genes required for differentiation (*FOXD1* and *LHX1* which are involved in stromal and epithelial differentiation respectively) are both among the subset of genes with bivalent chromatin that are not expressed.

1.11.4.2 miRNA

Apart from the evidence described in chapter 1.9.4, the expression of miRNAs has been analysed in WT resulting in several associations between aberrant expression of global or specific RNAs, their processors or regulators and WT. These include detection of a miRNA signature in the blood of WT patients that distinguishes it from healthy control blood and may be used to diagnose WT (Schmitt, Backes et al. 2012), a miRNA signature in WT blastema that differentiates intermediate and high risk WTs and could therefore be used to predict response to chemotherapy (Watson, Bryan et al. 2013), expression of both transcription factor

E2F3 and Oncomir-1 (an oncogenic cluster of miRNAs on chr13 regulated by E2F3) which were increased in WT compared to normal tissue with increase correlating with high stage and metastatic disease (Kort, Farber et al. 2008).

1.11.4.3 DNA methylation

DNA methylation profiling in WT has shown both specific and global methylation changes. Epimutations specific to WT are listed in Table 1.5, however the most well-characterised of these, likely due to its syndromic association (Reik, Brown et al. 1995), is abnormal methylation at 11p15 either due to loss of imprinting (LOI) by gain of methylation at the imprinting control region (H19-ICR) or by paternal uniparental disomy, observed in ~70% WT (Scott, Murray et al. 2012). Interestingly, mosaic gain of methylation at *H19* and reciprocal overexpression of IGF2 has been observed within the adjacent 'normal' kidney tissue of WT patients and is commonly found in PLNR (Ohlsson, Cui et al. 1999; Vuononvirta, Sebire et al. 2008).

Aberration	Finding within Wilms tumour
CASP8 hypermethylation	Frequency of 19% and associated with *RASSF1A* methylation (Harada, Toyooka et al. 2002)
GLIPR1/RTVP hypomethylation	Frequency of 87.5%, results in overexpression (Chilukamarri, Hancock et al. 2007)
Global hypomethylation	Results in genome instability in tumour cells (Ehrlich, Hopkins et al. 2003; Ludgate, Le Mee et al. 2012).
HACE1 hypermethylation	Frequency of 73% (Zhang, Anglesio et al. 2007)
Hypermethylation of a CTCF binding site downstream of *WT1*	Correlated with high WT1 expression (Zitzmann, Mayr et al. 2014)
Hypermethylation of protocadherin cluster at 5q31	Results in expression loss of these proteins at the cell surface (Dallosso, Hancock et al. 2009)
LOI 11p15	Frequency of 69%, results in overexpression of *IGF2* and down-regulation of *H19 (Haruta, Arai et al. 2008; Gadd, Huff et al. 2012; Scott, Murray et al. 2012)*
P16 hypermethylation	Frequency of 23% (Arcellana-Panlilio, Egeler et al. 2000)
RASSF1 hypermethylation	Frequency of 54% (Wagner, Cooper et al. 2002)
WT1-antisense transcript hypomethylation	Results in biallelic expression (Malik, Salpekar et al. 2000)

Table 1.5: Epimutations in Wilms tumour, taken from Charlton et al (Submitted to Oncotarget)

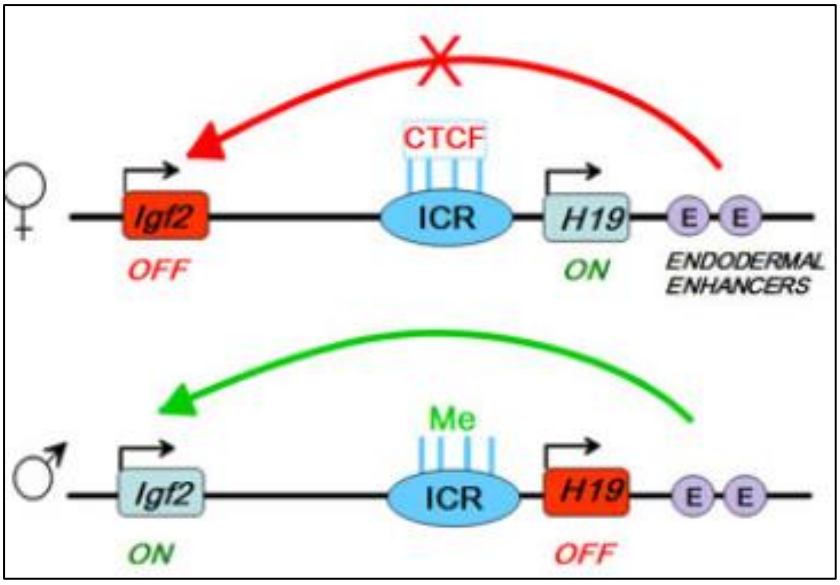


Figure 1.13: H19 imprinting control region. At the maternal locus (top) CTCF binds to the imprinting control region (ICR; blue) which prevents enhancer proteins binding which therefore cannot express IGF2. At this locus, H19 is actively expressed. Conversely, at the paternal allele (bottom), the ICR is methylated causing loss of H19 expression however enhancer proteins can now bind and IGF2 expression occurs.

The H19-ICR controls expression levels of *IGF2* and *H19*, and contains differentially methylated domains that show parent-of-origin dependent methylation, resulting in hemizygous expression of both genes in healthy tissue (Steenman, Rainier et al. 1994; Li, Franklin et al. 1998; Honda, Arai et al. 2008). The H19-ICR contains binding sites for CTCF (CCCTC-binding factor, zinc finger protein) and, upon binding, protects the maternal allele from *de novo* methylation (Fedoriw, Stein et al. 2004). Conversely, at the paternal allele, methylation prevents CTCF from binding and IGF2 is actively transcribed (see Figure 1.13). In WT, both alleles are methylated by the two mechanisms previously stated which results in biallelic IGF2 expression.

Besides the site-specific methylation changes in WT, global methylation events are also reported; large blocks of hypomethylation were observed in WT which corroborates the finding of genome-wide centromeric satellite DNA hypomethylation (Ehrlich, Hopkins et al. 2003; Ludgate, Le Mee et al. 2012).

1.12 Hypotheses to be tested

Using a well-characterised sample set from UK patients registered in two contiguous clinical trials in childhood WT that used the same diagnostic and treatment approach, the following experimental aims were proposed:

i) *To determine genome-wide methylation profiles for normal kidney, nephrogenic rest and Wilms tumour*

ii) *To identify epigenetic changes associated with the persistence of NR in post-natal kidney as well as evolution of WT from its presumed precursor in normal embryonic kidney*

iii) *To identify epigenetic markers that will be useful in distinguishing NR from WT – a particularly challenging differential diagnosis for histopathology, especially in the context of chemotherapy-treated WT*

iv) *To identify changes in methylation associated with progression towards malignancy rather than mixed cell populations by performing both non-corrected and corrected analyses*

Additional aims were subsequently added, once the initial work had established that methylation changes could be used to distinguish tissue types:

v) *To identify tumour-specific sites of methylation that can be detected in blood serum for use as a biomarker*

vi) *To confirm that the blood methylation biomarkers are discriminative in an independent cohort and have clinical utility*

Chapter 2: Materials and Methods

2.1 General laboratory equipment

A table of general laboratory equipment used in this project can be found in Appendix 3: Table A2.

2.2 Nephrectomy tissue samples

2.2.1 Source of samples and processing

All nephrectomy tissue (Wilms tumour (WT), nephrogenic rest (NR) and normal kidney (NK)) was derived from samples collected through either the Society of Paediatric Oncology (SIOP) 2001 study, the Improving Outcomes for Children with Renal Tumours (IMPORT) 2012 study both with signed parental consent or from overseas or occasional patients treated at Great Ormond Street Hospital (GOSH), not enrolled on either trial but with appropriate consent for biological studies. Sample collection was carried out with ethical approval granted by the NHS London Bridge Research Ethics Committee (ref: 12/LO/0101). Each centre contributing to these clinical trials collected tissue for research purposes which was either frozen fresh in liquid nitrogen or fixed in formalin overnight before being embedded in paraffin the next day. Formalin fixed paraffin embedded (FFPE) blocks can be stored long-term at room temperature while fresh frozen tissue was stored in cryogenic tubes in liquid nitrogen or at -80ºC, according to local policies across the CCLG principal treatment centre network.

2.2.2 Haematoxylin and eosin staining of tumour sections

To investigate the composition of an FFPE block, 3μm sections were cut using a microtome and mounted onto a slide by heating to 60ºC in an incubator. These sections were then stained with haematoxylin and eosin using a Leica Autostainer XL (protocol in Table 2.1).

Step	Reagent	Time
1	Xylene	2 min
2	Xylene	2 min
3	Abs alcohol	1 min
4	Abs alcohol	1 min
5	Abs alcohol	1 min
6	Tap water	2 min
7	Harris haematoxylin (Leica, Ref 3801560E)	6 min 30 sec
8	Tap water	2 min
9	1% acid alcohol	30 sec
10	Tap water	1 min
11	1% eosin (Leica, Ref. 3801590E)	4 min 30 sec
12	Tap water	1 min
13	Abs alcohol	45 sec
14	Abs alcohol	30 sec
15	Xylene	30 sec
16	Xylene	1 min
17	Mount with DPX (VWR Ref. 360294H)	

Table 2.1: Protocol for haematoxylin and eosin staining of FFPE sections using the Leica Autostainer XL

2.2.3 Pathological review and case selection

For both the SIOP trial and the IMPORT study, centralised pathological review for clinical quality assurance of the overall risk group was carried out by Professor Gordan Vujanic in Cardiff. He received the stained sections and analysed cell composition under the microscope to determine the predominant cell type (if present), the extent of post-chemotherapy changes, the presence of NRs and whether anaplasia or further differentiation was present. From these reports, cases were selected that indicated the presence of NRs and the block numbers recorded. Following this, as GOSH the designated biological study coordination hub for both studies and therefore the collection point for research tissue, the list of blocks with potential NRs was cross-referenced with those in storage and the available blocks (including many collected nationally with a higher number of GOSH cases) could easily be retrieved and fresh sections cut and stained for local pathology review.

Local pathology assessment was next carried out by three independent paediatric pathologists (Professor Neil Sebire, Dr. Sergey Popov and Dr. William Misfud) who assessed the H&E sections under a microscope and marked out regions of NK, NR and WT (see Figure 2.1).

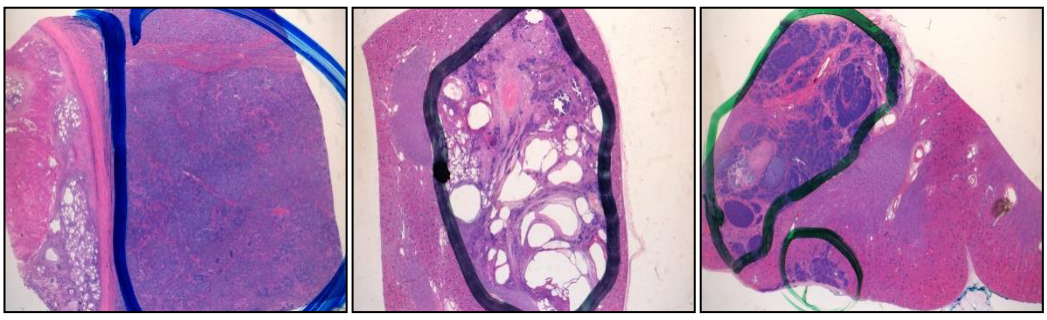


Figure 2.1: Haematoxylin and eosin stained 3µm sections of FFPE tissue mounted on glass slides with NRs circled by a paediatric pathologist.

Cases selected for DNA extraction had large areas of each tissue type identifiable with minimal necrosis or chemotherapy induced changes. From review of cases with pathology reports indicating the presence of NRs, we found many did not have suitable (large enough) NRs, several had very compressed kidney which was therefore considered "not normal" and many had only necrotic tumour tissue. Therefore, a total of 145 samples were deemed suitable for DNA extraction.

2.3 DNA extraction from clinical tumour samples

2.3.1 Preparation of samples for DNA extraction

After cases were marked for DNA extraction, a decision was made whether to cut multiple sections or to perform a core puncture into the tissue. This decision was based on the surrounding tissue composition, for example if a section showed the presence of both NR and NK, it was not possible to tell whether the full tissue block (up to 7mm in depth) was NR all the way to the base or whether contaminating NK was underneath. Conversely, in a block taken from the periphery of the lesion that showed a section composed entirely of NK, we could be confident that the whole piece was NK and therefore a core piece was taken.

For core punctures, a 2μm diameter puncture was taken and the piece finely chopped using a scalpel. For tissue requiring multiple sections, up to 30 5μm sections were cut (depending on area of target region) and the target region was scraped from the slides using a scalpel. In both cases, tissue was put into an Eppendorf.

2.3.2 Optimisation of protocol for DNA extraction

Prior to extraction of DNA from patient tissue, an FFPE block composed entirely of NK tissue was used to identify the method that yielded maximum DNA. Three Qiagen protocols were compared: DNeasy, QIAamp DNA mini kit and the QIAamp FFPE kit. The DNeasy kit had the most basic protocol while the QIAamp DNA mini kit added a PBS wash and incubation at 70ºC with buffer AL while the QIAamp FFPE kit added an extra incubation at 90ºC prior to buffer AL. All combinations of the DNeasy protocol with either one, two or all three additional steps were performed on duplicate tissue pieces and the final DNA yield was compared. The greatest yield was seen when both the 70ºC and 90ºC heating steps were included irrespective of the additional PBS wash, therefore the DNeasy protocol was adapted including these steps (see chapter 2.3.3 for protocol).

2.3.3 DNA extraction from FFPE

The final optimised protocol was as follows: 1ml xylene was added to the tissue samples which were vortexed for 15 seconds and centrifuged for 5 minutes. Then, the supernatant removed and two further washes in 100% ethanol were carried out before excess ethanol was evaporated on a 37˚C heat block. 180µl ATL buffer (DNA Mini Kit, Qiagen) with 20µl proteinase K was added for overnight digestion at 56˚C and following this, samples were heated to 90˚C for 1 hour. Samples were then incubated at 70˚C for 10 minutes with 200µl AL buffer (DNA Mini Kit, Qiagen) before 200µl 100% ethanol was added. DNA purification was carried out using Qiagen columns (Qiagen DNA Mini Kit). First, the mixture was added to a

column and spun at 8,000 rpm for 1 minute. Then, 500µl Buffer AW1 was added and samples were centrifuged at 8,000 rpm for 1 minute before the procedure was repeated with Buffer AW2. A full-speed centrifugation step for 2 minutes ensured the membranes were dry, and then the columns were transferred to new collection tubes for incubation with 70µl water for 5 minutes to elute the DNA followed by a final spin at full speed for 1 minute. The eluate containing the DNA was transferred to a clean Eppendorf and stored at -20ºC.

2.3.4 Quality control of samples

Concentration of DNA was calculated using a NanoDrop spectrophotometer (Thermo Scientific) which also calculates the purity of the sample by giving the 260/230 ratio. A more stringent calculation of only viable double stranded DNA was obtained using the Qubit 2.0 Fluorometer (Life Technologies) according to manufacturer's instructions.

2.4 Blood serum samples

2.4.1 Source of samples

Blood was collected on the IMPORT study at 5 time points: (1) at diagnosis, (2) mid-way during the course of pre-operative chemotherapy, (3) pre-surgery after finishing the course of chemotherapy, (4) post-surgery prior to starting post-surgery chemotherapy and (5) at end of treatment all covered by the appropriate consent given for participation in the study. Blood serum samples from age-matched controls

without cancer were obtained from Great Ormond Street Hospital Department of Chemical Pathology with appropriate signed parental consent for research use.

2.4.2 Initial handling and processing

Blood was collected into a serum tube and left to clot for at least 30 minutes in the respective centre that the patient was treated. Each centre then centrifuged the tube for 1 minute at 1000rpm at room temperature before the clear serum at the top of the tube was removed and decanted into a 1.5 ml Eppendorf before being frozen and stored at -80ºC.

2.4.3 Extraction of circulating free DNA from serum

The QIAamp circulating nucleic acid kit (Qiagen) was used to extract DNA from blood serum which was carried out according to manufacturer's instructions as follows. All serum samples were left to thaw on ice then brought up to 2ml by adding sterile phosphate-buffered saline (PBS). To begin, 200µl Proteinase K was added to the serum sample in a 50ml tube. To this, 1.6ml Buffer ACL was added and the mixture was vortexed prior to incubation at 60ºC for 30 minutes. Next, 3.6ml Buffer ACB was added and samples were incubated on ice for 5 minutes. To filter the samples, a QIAamp mini column was attached to a vacuum pump and the solution containing serum was filtered through. The filter was washed with 600µl Buffer ACW1, 750µl of ACW2 and 750µl 100% ethanol before the column was removed from the vacuum pump and centrifuged at full speed for 3 minutes. The

filter membrane was then dried completely by incubation at 56ºC for 10 minutes. Finally, 20µl Buffer AVE was added to the centre of the column filter membrane followed by a 3 minute incubation and a 1 minute centrifugation to elute the DNA.

2.4.4 Quality control of cfDNA

As concentrations of DNA extracted from blood serum were so low, the Qubit reading showed undeterminable levels of DNA. Therefore, a PCR was performed using primers for DMR2 to confirm the presence of DNA along-side a control genomic DNA sample that always amplified (Figure 2.2).

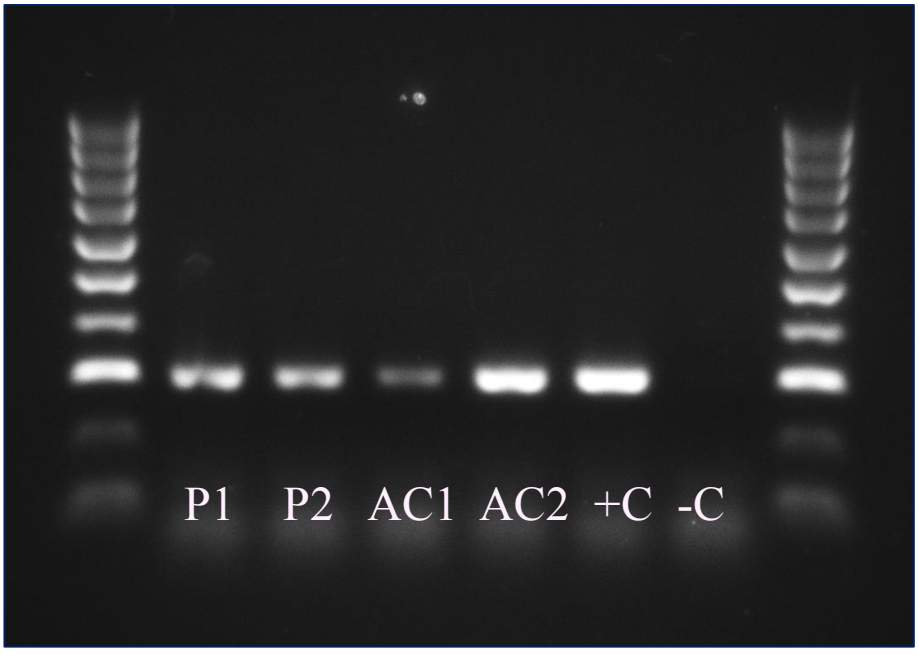


Figure 2.2: 2% agarose gel image from PCR amplification of DMR 2 from cfDNA extracted from blood serum. The gel was labelled with ethidium bromide. +C = positive control: genomic DNA that always amplifies, -C = negative control: water, P = cell-free DNA from a patient (child), AC = cell-free DNA from an adult.

2.5 Embryonic kidney samples

2.5.1 Source of samples

Human embryonic kidneys were provided by the Joint MRC (grant #G0700089)/Wellcome Trust (grant #GR082557) Human Developmental Biology Resource (www.hdbr.org) at the UCL Institute of Child Health through collaboration with Professor Paul Winyard. Ethical approval was covered by the HDBR HTA tissue bank licence and ethical approval. Details of approval terms can be found at www.hdbr.org.

2.5.2 DNA extraction from embryonic kidney

DNA was extracted by a colleague from four human embryonic kidneys (from gestational age 8 weeks and 12 weeks) using the AllPrep DNA/RNA Micro Kit (Invitrogen cat #80284) according to manufacturer's instructions.

2.6 Bisulfite conversion of extracted DNA

2.6.1 Bisulfite conversion treatment

DNA was bisulfite converted using the EZ DNA methylation kit (Zymo Research). For this procedure, 5µl M-dilution buffer was added to 45µl of DNA diluted in distilled water. These samples were incubated at 37ºC for 15 minutes before 100µl CT conversion reagent was added and samples were mixed by pipetting

up and down. Samples were then incubated for 16 hours at 50ºC with a 30 second heating step to 95ºC after each hour. Samples were then incubated on ice for 10 minutes before mixing with 400µl M-binding buffer and passing samples through a Zymo-spin column by centrifugation. Columns were then washed with 100µl wash buffer, then 200µl M-desulphonation buffer was added and samples were incubated at 15-20 minutes at room temperature. Following this, two 200µl washes with M-wash buffer were performed before DNA was eluted with 10µl M-Elution buffer.

2.6.2 Quantitative PCR to check for conversion success

To confirm bisulfite conversion, a quantitative PCR (qPCR) reaction was performed which selectively amplifies a post-conversion site within the gene *MLH1* (with many unmethylated cytosine residues that will be converted to thymine) with primers specific for bisulfite-converted DNA and a pre-conversion site within the gene *GAPDH* with primers specific for normal DNA (primers listed in Table 2.2, PCR conditions in Table 2.3).

qPCRs were performed using MESA blue qPCR Master Mix Plus (cat# RT-SY2X-03+WOUB 600rs) using the 7900KT Fast Real-Time PCR System (Applied Biosystems) with detection for SYBR green. The Applied Biosystems SDS programme Version 2.2.2 calculated the respective Ct values for every sample. In most cases, the *GAPDH* loci amplified after a large number of cycles as the DNA was bisulfite-converted and therefore the primers were no longer complementary. Conversely, fully converted DNA amplified the *MLH1* site in very few cycles. The

ΔCt was calculated as the difference between *MLH1* and *GAPDH* Ct values for every sample, and the % conversion for each sample was calculated by:

% unconversion = 100/(1 + (2^ΔCt))

Only samples with >98% conversion were included.

Primer	**Sequence**
MLH1 F (5'>3')	GGAGTGAAGGAGGTTACGGGTAAGT
MLH1 R (5'>3')	AAAAACGATAAAACCCTATACCTAATCTATC
GAPDH F (5'>3')	CGCCCCCGGTTTCTATAAAT
GAPDH R (5'>3')	CAAAAGAAGATGCGGCTGAC

Table 2.2: Primers used for qPCR validation of bisulfite conversion

Step	**Temp**	**Time**
1	50ºC	2 min
2	95ºC	10 min
3	95ºC	15 sec
4	60ºC	1 min (cycle steps 3 & 4 40 times)
5	95ºC	15 sec
6	60ºC	30 sec
7	95ºC	15 sec

Table 2.3: PCR conditions used for qPCR validation of bisulfite conversion

2.7 Methylation profiling with 450k microarrays

2.7.1 Ligation reaction

The quality of DNA extracted from FFPE tissue is poor due to the fixing and extraction method which causes fragmentation of the DNA to a range of 300-600bp. This is problematic for the Illumina Infinium assay as a downstream whole-genome amplification procedure requires templates to be >1kb in length. Therefore, following an optimised FFPE protocol (Thirlwell, Eymard et al. 2010) 1-2µg DNA (depending on available yield) was treated using the REPLIg FFPE kit (Qiagen) which induces random ligation of DNA. Once ligated, the DNA strands are lengthened and suitable for whole-genome amplification.

This protocol involved adding 8µl FFPE buffer, 1µl ligation enzyme and 1µl FFPE enzyme to each sample of bisulfite-converted DNA and incubating at 24ºC for 30 minutes, 95ºC for 5 minutes and then cooling down to 4ºC.

2.7.2 450k beadchip arrays

These ligated samples were analysed by an in-house service provider (UCL Genomics) who ran the DNA on the Illumina 450k HumanMethylation BeadChips using HumanMethylation450 DNA Analysis BeadChip Kits (Cat# WG-314-1001, Illumina) according to manufacturer's recommendations. Raw data were extracted from Illumia Genome Studio without any processing.

2.8 450k data analysis

2.8.1 Data pre-processing

For every CpG, a methylated and non-methylated measurement is taken, with the intensity of the scanned signal correlating with the quantity of methylated or non-methylated DNA present in the sample respectively. The raw IDAT files hold this intensity information and were extracted from the Illumina HiScan machine. These files also contain a detection p-value for every probe, which indicates how well the signal intensity has been interpreted (with small p-values indicating high significance and good performance). All data analyses were performed using R (Version 3.0.2).

Raw .idat files were loaded into R using the Bioconductor package ChAMP (Version 1.2.7) (http://www.bioconductor.org/packages/release/bioc/html/ChAMP.html). The algorithm

```
> champ.load(directory_of_raw_IDAT_files, p=0.01)
```

performed a pre-processing step removing probes with high detection p-values. The level of significance used as a cut-off can be altered however for this study it was consistently 0.01. The output of this algorithm was loaded methylated and unmethylated intensity values (an 'RGset' or 'MethylSet') with the total number of probes filtered for good detection p-values.

2.8.2 Data normalisation

After pre-processing, the data needed to be normalised as there are two probes types on the array. Type I probes originated from the 450k predecessor, the 27K array, and use two probes in the same colour channel to measure methylated and unmethylated intensities whereas type II probes were designed differently and use one probe and two different colour channels to make the same measurements. Due to these two different methods of signal detection, there are slightly different signals generated which need to be corrected for to avoid probe type-specific false positives. Therefore, intra-array normalisation was essential and was performed using either the Dasen method implemented through the Bioconductor wateRmelon package (Version 1.4.0) (http://www.bioconductor.org/packages/release/bioc/html/wateRmelon.html):

> Dasen(MethylSet)

or the SWAN algorithm implemented through Bioconductor package Minfi (Version 1.10.2) (http://www.bioconductor.org/packages/devel/bioc/html/minfi.html).

> preprocessSWAN(RGSet)

Both methods can use an 'RGset' or a 'MethylSet' as an input which is a measure of the raw methylated and unmethylated intensities for each CpG for each sample and both perform quantile normalisation, i.e. they adjust intensity signals to give both probe distributions identical statistical properties. Quantile normalisation shifts the type II probes to fit the type I probe distribution (Figure 2.3). The additional benefit of using Dasen is that between-array normalisation is also performed. The background difference between type I and type II probes for both methylated and

unmethylated intensities is adjusted to prior to quantile normalisation so that chips run at different time points are comparable. From the adjusted methylated and unmethylated signal intensities, a matrix of β-values was computed by the equation:

β = methylated probe intensity / methylated probe intensity + unmethylated probe intensity

The resulting data_matrix for downstream analysis had β-values for all samples with .idat files included in the pre-processing step and all CpGs included after the pre-processing step.

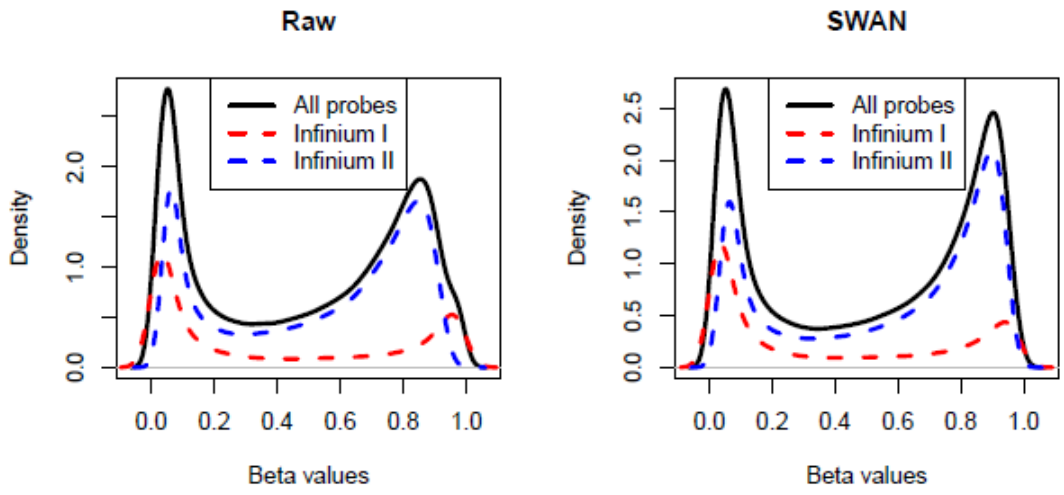


Figure 2.3: Quantile normalisation of 450 probes, taken from the Bioconductor Minfi Users Guide. The panel titled "raw" shows the distribution of the non-normalised probes. The panel labelled "SWAN" shows the probe densities after normalisation. The method of quantile normalisation fits the type II probe distribution onto the type I probes so that they both acquire identical statistical properties.

2.8.3 Cell type composition correction

Recently, a lot of attention has been given to the generation of algorithms to account for the mixed cell type composition effects that arise when comparing two tissue types. In this study, the algorithm RefFreeEWAS was used (Houseman, Molitor et al. 2014).

This method began with the command

```
> EstDimRMT(data_matrix)
```

which uses random matrix theory to estimate the significant components of variation (d) within the pre-processed and normalised data_matrix. In this model, d = the number of cell types present. As this approach uses vast amounts of computational memory, the command was run with 10 randomly selected smaller subsets of the full dataset (of 20 samples by 100,000 CpGs). In each instance, d was estimated as 3, which biologically makes sense as three predominant cell types were present: stroma, epithelia and blastema. There was also a very small number of blood cells, cells with rhabdomyoblastic features and fat cells; however these were very rare and should not confound the overall analysis.

Next, the command

```
> RefFreeEWASModel(data_matrix, unadjusted_linear_model, d)
```

was run. This command takes the pre-processed and normalised data_matrix of β-values together with a defined linear model (designed using the Bioconductor package Limma, see chapter 2.8.4 below) and the estimate for d, and performs the linear modelling while deconvoluting the DNA methylation arising as mixtures of

cell types. This algorithm returns mixture-adjusted covariate values according to the defined linear model, i.e. if a linear model comparing NK to matched NR in five patients was designed, the output would be the difference in β-value between NK and NR for each patient at every CpG. These covariates would be adjusted for cell type composition effects. Although the most substantial changes in methylation can be identified, the significance of these needs to be interpreted which is done using the algorithm:

> BootRefFreeEwasModel(output, n)

Where 'output' is the whole model generated by the RefFreeEWASModel algorithm (which includes many other variables) and n = the number of bootstraps to run. Bootstrapping is a method of generating p-values without generating more data. The algorithm defines a new dataset (d_new_1) by taking a smaller sample subset of the original data (with all CpGs included) in a random order and then computes the covariate values for these samples according to the specified linear model. This process is repeated up to d_new_n and the covariates are then compared to generate standard deviations and corresponding p-values for every CpG. In this study, n was set to 100. Ideal studies would run n>500 however with the computation systems available, n=100 still gives reliable p-values for good interpretation.

2.8.4 Linear modelling

Linear modelling was performed using the Bioconductor package Limma (Version 3.20.8) (http://www.bioconductor.org/packages/release/bioc/html/limma.html). This package

involves generating a design matrix specifying the number of samples and the comparisons to make, i.e. a model to compare levels of methylation within one patient between histology types (such as NK vs NR). The design matrix can also factor in multiple group analyses, for example an ANOVA analysis. Both of these formats were used in this study. Wherever a matched design was used, SNPs and sex chromosomes could be included as the matched design avoids patient-specific genomic events from confounding the analysis. In grouped comparisons, these CpGs were excluded.

Once the design matrix had been specified, the linear model was fitted to the data followed by a Bayesian framework fit which moderates standard errors across genes. Finally, the topTable command generated an output table of the full list of CpGs with covariates, p-values and p-values adjusted for multiple sampling (by a user-specified method). In this study, the Benjamini-Hochberg model (Benjamini and Hochberg 1995) was used.

2.8.5 TREAT

The TREAT algorithm (McCarthy and Smyth 2009) allows users to identify changes in methylation that are both mathematically and biologically significant. The TREAT function is applied to a linear model prior to the Basesian model and adjusts p-values based on the detected Δβ (covariates) giving more weight to those with a small Δβ (which would be not interesting biologically) but increased weight to those with a large Δβ (i.e. the p-value is decreased). The end result is a list of all

CpGs sorted by a p-value with the CpGs that show the largest and most significant differences in methylation ranked top.

2.8.6 DMR finder

The algorithm used to identify differentially methylated regions in this study:

> champ.lasso

was implemented through the Bioconductor package ChAMP. This algorithm takes all the background CpGs and calculates the distance to the nearest neighbouring CpG. According to the Illumina annotation, the probes are defined by whether they are located in a CpG island, shore, shelf or other and then by location in a 3'UTR, 5'UTR, TSS200, TSS1500, gene body, IGR or 1st exon (where TSS = transcription start site), giving 28 unique categories. The distance to the nearest neighbouring probe for every CpG was identified using the Illumina MAPINFO annotation to calculate the average CpG density for probes within each category (Figure 2.4). This data demonstrates the unusual distribution of probes on the array with large between-probe distances seen at intergenic regions and small distances seen at TSS200 regions. The data for each category was plotted linearly and the minimum lasso size was set to 10bp. From reading off the quantile required to reach a minimum lasso size of 10bp, this quantile was applied to every category to get a distinct lasso size for each feature that best fits the distribution of probes on the array (Figure 2.5). Using this information, the category-specific lasso was sent out centred on each probe and only those probes with adjusted p-value <0.01 were called significant. DMRs were then defined if the lasso connected three or more significant probes.

Any non-significant probes within the lasso region were also included in the DMR and those DMRs within 1kb of each other were encompassed into one region.

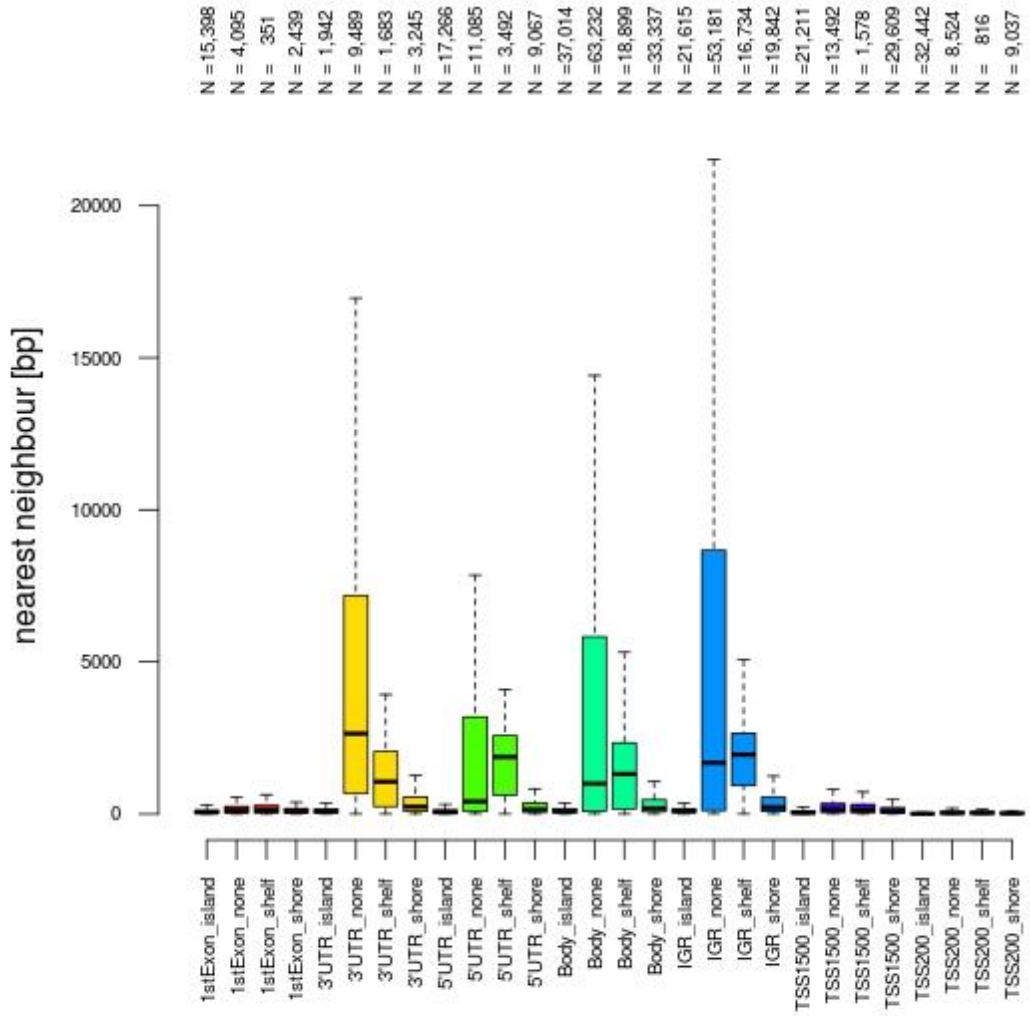


Figure 2.4: The distance to the nearest neighbouring probes for all CpGs on the 450k array separated into categories. The number of CpGs included in each category (N) is shown with the distribution of the nearest probe. The IGR_none category shows the greatest range of distributions while the 3'UTR_none category shows the greatest mean distance. Probes within the 1st exon and TSS200 regions are generally located very close together.

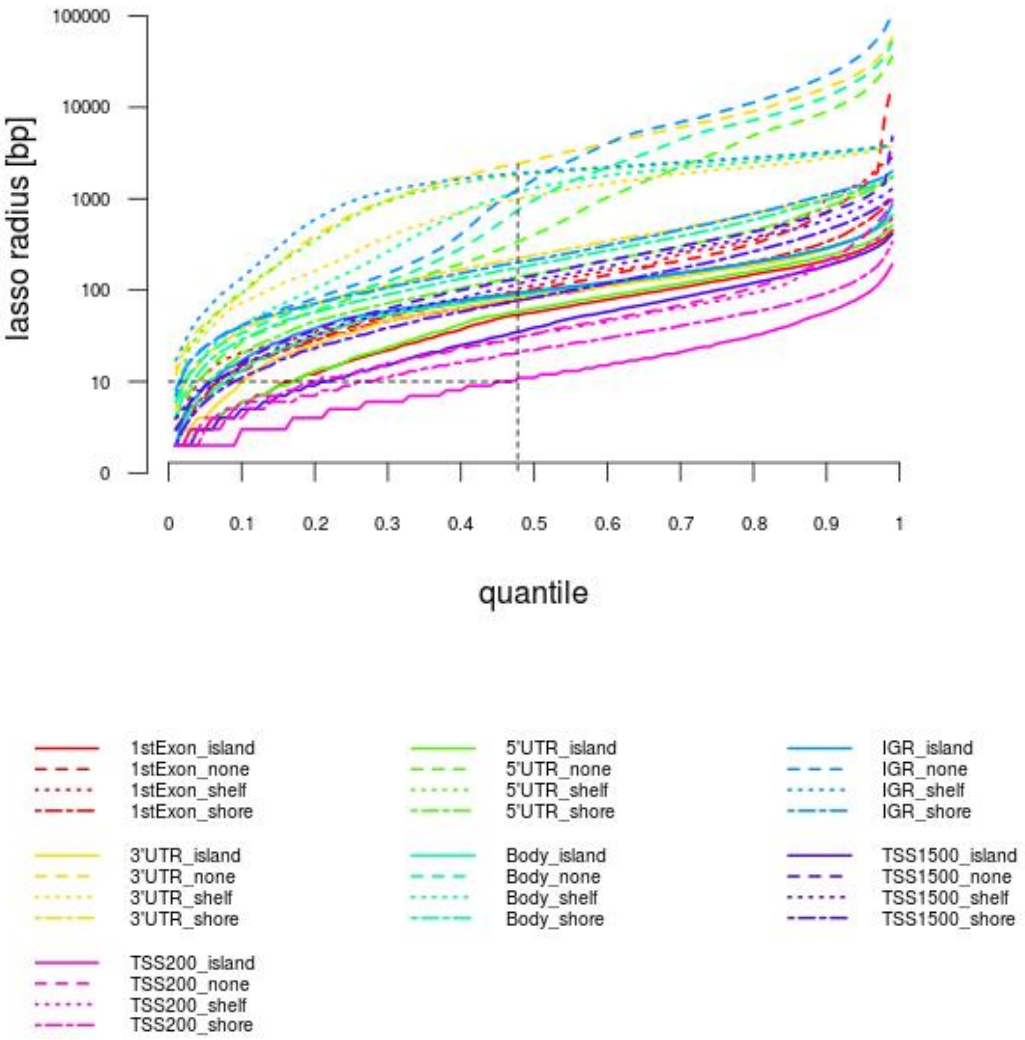


Figure 2.5: Defining category-specific lasso radii for 450k probes. A horizontal line is seen at 10bp, which intersects the TSS200_island category which shows the greatest CpG density. The quantile required for this to be set as a minimum parameter is ~0.48, which is then applied to all categories and a readout of the lasso radius is taken from the y-axis.

2.8.7 Defining the most variable data points

For several analyses, the top most variable data points were inputted. These were identified using the "varFilter" command from R package genefilter

(http://www.bioconductor.org/packages/release/bioc/html/genefilter.html) which requires an input of a data matrix of β-values (with columns as samples and rows as CpGs) with a user-specified method of discerning those variable positions. The parameters chosen were: based on interquartile range, select the top 1% most variable data points.

2.8.8 Multidimensional scaling

Multidimensional scaling (MDS) plots were generated using Minfi command "mdsPlot" in R. The input requires a data matrix of β-values and the number of positions used to generate the plot, i.e. if number of positions = 1000, the top most variable 1000 CpGs were used to perform the multidimensional scaling. The algorithm then calculates the distance between samples to generate a distance matrix in two dimensions and then represents this as a scatterplot (with the first dimension on the x-axis and the second on the y-axis) so that those samples that are more similar are closer together.

2.8.9 Hierarchical clustering

To perform hierarchical clustering, the open source package Heatmap.3 (Version 3.1.1) (http://www.inside-r.org/packages/cran/GMD/docs/heatmap.3) was used. This script uses a matrix of β-values and evaluates the Euclidian average distance between samples before generating a hierarchical cluster. The script allows

for an input table of colours to generate a user-specified sample-associated colour grid at the top of the cluster.

2.8.10 Overlapping regions

BedTools Version 2.18 (http://bedtools.readthedocs.org/en/latest/) was installed and the intersectBed command was used via the command line to compare two bed files. The output was a list of all regions that directly overlapped.

2.8.11 Support Vector Machine algorithm

A support vector machine algorithm was applied using R CRAN package e1071 (http://cran.r-project.org/web/packages/e1071/index.html). This algorithm used a training set with two groups (NK and WT) specified to identify a group-specific methylation pattern. Next, the algorithm compared this known pattern to a test set and outputted a sample-specific evaluation of whether the sample more closely resembled NK or WT.

2.9 Pathway analysis

Pathway analysis on significant CpGs, (either MVPs or DMRs) was performed using Genomic Regions Enrichment of Annotations Tool (GREAT) Version 2.0.2 (http://bejerano.stanford.edu/great/public/html/). This freely available software takes input Bed files of target sites (i.e. MVPs) as well as a background file

of all CpGs that could have been considered significant. This software outputs a series of informative tables (listed in Table 2.4) by combining annotations for over 20 ontologies (McLean, Bristor et al. 2010).

Ontology	
GO molecular function	MSigDB pathway
GO biological process	MGI Expression: detected
GO cellular component	MSigDB Pertubation
Mouse phenotype	Transcription factor targets
Human phenotype	MSigDB predicted promoter motifs
Disease ontology	MSigDB miRNA motifs
MSigDB Cancer Neighbourhood	miRNA targets
Placenta disorders	InterPro
Panther pathway	TreeFam
Pathway commons	HGNC gene families
BioCyc pathway	

Table 2.4: Output of software GREAT

2.10 Comparison to publically available datasets

As genomic data must now be uploaded with submission of a related paper, there is a large repository of 450k methylation datasets online. These can be accessed through Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) or The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/). Multiple downloads

of target regions across many samples can be done using the Marmal-aid package (http://marmal-aid.org/).

In this study, data from four embryonic kidneys (GSM868048, GSM868049, GSM848050 and GSM868051), six embryonic stem cell lines (GSM867936, GSM867937, GSM867938, GSM867939, GSM867940 and GSM867942) and six induced pluripotent stem cell lines (GSM867968, GSM867969, GSM867970, GSM867972, GSM867973 and GSM867980) were extracted from GEO (GEO series GSE31848). Analysis using these data is described in chapter 4.

Marmal-aid Version 1.2.1 was used to extract publically available methylation data either annotated as disease = 'Healthy' and tissue = 'blood' (n = 411) or disease = 'Cancer' (n = 5,238) for comparative analyses discussed in chapter 6.

2.11 DNA sequencing using Illumina Miseq

2.11.1 Primer design

MethPrimer (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi) was used to design primers specific for bisulfite converted DNA. All primers were checked using an ePCR tool to confirm amplification of a single product only (http://bisearch.enzim.hu/) and ordered from Sigma Aldrich (Table 2.5).

Region	Forward primer	Reverse primer	Tm
DMR1.1 bisulfite-converted	TTTTTATTTTTATAAATAGTAAGGATTG	AAAAACTTCAAACAACTATTCAACC	51
DMR1.2 bislufite converted	GTTAAAATTTTGGGTTAAGGAGAT	AAAATAAAAATTCAAAAAATAACCC	51
DMR2.1 bisulfite-converted	GGTTTTTGAGGAATTGGATTT	TCCCAAACTCTTTCTACAATCATAC	51
DMR2.2 bisulfite-converted	ATTGTAGAAAGAGTTTGGGAGAT	TTCTAAAAACCATCTTAAAAAAAA	51
HISTH1A bisulfite-converted	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGAGGTATTGTTTTAGATATGGTG	GAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCCTAAAAACCCCCAACTACCTTCTA	59
HOXD12 bisulfite-converted	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGAAGAATTTGTAGGTAAAGTTT	GAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCCCTCAAATTAAAAAAATAAAAAAAATCTAA	56
ZFP42 bisulfite-converted	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAGGGTAAATGTGATTATA	GAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTAATCAAACTACAACCACCCA	57
PRKCZ bisulfite-converted	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGGGTTGTTTTTGTGATTT	GAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTAAAAATTCTACCCACTTCC	58
AGAP1 bisulfite-converted	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTAGTTTTTTTTTTGGTTAGGG	GAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCCCTCCTATAACCTCTTTATTAAAA	60

Table 2.5: List of primers used. DMR1 and DMR2 primers (described in chapter 6) were split into two regions to cover the entire region.

Once received, primers were briefly centrifuged at 13,200 rpm to ensure the primer pellet was in the base of the tube. For all primers ordered from Sigma Aldrich, a varying volume of nuclease-free water was added, as per the accompanying primer information sheet, to make a stock concentration of 100 μM. This was then further diluted 1:10 to generate 10μM concentrations as required for the PCR reactions.

2.11.2 Library preparation

2.11.2.1 Library preparation: PCRs

Library preparation began with PCR amplification using region-specific primers. For bisulfite-converted DNA, 10ng genomic DNA was diluted in PCR-grade water to make a total of 25µl. To this, 1µl of forward primer and 1µl reverse primer were added with 25 µl 2X KAPA HiFi HotStart Uracil+ ReadyMix (Kapa Biosystems). PCR involved initial denaturation at 95ºC for 5 minutes followed by 35 cycles of (98ºC for 20 seconds, specific annealing temp for 15 seconds, 72ºC for 60 seconds) before a final extension step at 72ºC for 1 minute.

2.11.2.2 Library preparation: clean-up

Next, samples were cleaned up using AMPure XP magnetic beads (Beckman Coulter). To the PCR sample, 45µl beads were added, mixed by pipetting and incubated at room temperature for 5 minutes. Samples were then put on a magnetic stand and left for 2 minutes. During this time a pellet formed on the magnet which included the target DNA. The supernatant was removed and the beads were washed twice with 200µl 100% ethanol to remove the PCR contaminants. Samples were air dried to remove residual ethanol and then were removed from the magnetic stand. 20µl elution buffer was then added and the beads were resuspended by pipetting. After incubation at room temperature for 2 minutes, 45µl binding buffer was added and samples were incubated for a further 5 minutes at room temperature. The beads were subsequently pelleted on the magnetic stand and the supernatant was removed. Samples were washed twice with 200µl 100% ethanol then air dried before

resuspension in 25µl elution buffer. The beads were pelleted on the magnet one last time before the eluate containing the DNA was transferred to a fresh tube.

2.11.2.3 Library preparation: Quantification

After the clean-up procedure, samples were quantified using the Quant-it Picogreen dsDNA assay kit (Life Technologies, cat#P11496). To begin, 1µl Picogreen dye was diluted in 199µl TE buffer. Serial dilutions of a known concentration of DNA were made up in a plate along-side test concentrations of 8µl of H_2O with 2µl DNA. To every control and test sample, 90µl Picrogreen mix was added. The plate was immediately read on a qPCR machine with detectors for SYBR and VIC. A linear regression was then performed using the control samples to determine the final concentrations for all test samples. With these concentrations, PCR products for each patient were pooled into one plate and samples were quantified using a 2100 Bioanalyser (Agilent) to check the range of PCR products and the absence of contamination.

2.11.2.4 Library preparation: Final processing

Next, a PCR was performed to add patient-specific tags to each product. Another clean up with beads (as above) was performed prior to a final quantification with a bioanalyser chip and a final pooling step of all samples. This pool was then run on the Illumina Miseq by UCL Genomics.

2.12 Miseq data analysis

2.12.1 Data alignment

Raw fast-q files were extracted from the Illumina sequencer. These were then processed from the command line using Perl scripts. Data was first mapped using Bismark Version 0.9.0 (http://www.bioinformatics.babraham.ac.uk/projects/bismark/), which uses Bowtie2 Version 2.2.2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) and a reference genome (hg19) to perform the alignments. The reference genome gets converted to a 'bisulfite-converted' genome so that the sequences reads can be aligned. 'Methylation extractor' was then run using Bismark which extracts a methylation call for every C analysed and generates a BedGraph file which is used for visualisation in Integrative Genomics Viewer (IGV; http://www.broadinstitute.org/igv/home).

Finally, SAMTools (http://samtools.sourceforge.net/) was used to sort and index the Bam files so that regions of interest could be assessed in IGV (i.e. the number of C and T reads for every CpG of interest were quantified).

2.12.2 Interpretation of reads

For every CpG of interest, the number of reads that remained C (i.e. were originally methylated) and those that had been converted to T during bisulfite treatment (i.e. were originally not methylated) were quantified. The proportion of methylated reads was calculated as the number of 'C' reads divided by the total

number of reads for each CpG. This gave a value between 0 and 1 which was comparable to the methylation β-values derived from the 450k array which also used the same scale.

Comparison between platforms involved generating a scatterplot with a linear line of best fit and then deriving the R^2 value as a measure of correlation.

2.12.3 Getting base-specific allele counts

To generate allele counts for the entire sequence of a specific region (instead of selecting target CpGs), the ANGSD package (http://popgen.dk/wiki/index.php/ANGSD) was used. This package reads in bed files and exports a text file of the entire sequence and the number of reads that were A, C, G or T.

2.13 Genome-wide hydroxymethylation quantification

Quantification of hydroxymethylation (H-Me) was carried out using the Global DNA Hydroxymethylation ELISA Kit (Cell Biolabs, Inc, STA-381) following manufacturer's instructions. Samples were initially incubated in the DNA High-Binding plate overnight at 4˚C to improve DNA binding. All samples were assayed in duplicate.

Optical density (OD) at 450nm was recorded including 6 H-Me DNA standards which were used to determine a standard curve following manufacturer's instructions

with formula y = 0.0186x – 0.0371 (R^2 = 0.9351). From the standard curve, H-Me quantification was carried out on unknown samples using their OD reading at 450nm.

2.14 RNA-sequencing

2.14.1 RNA extraction

RNA was extracted from 12 samples (4 trios of NK, NR and WT) by cutting multiple 5µm sections of FFPE tissue and scraping the target region using a scalpel. Tissue was put into an Eppendorf then the RNeasy FFPE kit (Qiagen) was followed according to manufacturer's instructions as follows.

1ml xylene was added to the Eppendorf which was vortexed for 10 seconds before centrifugation for two minutes. The supernatant was then removed and 1ml 100% ethanol was added. The Eppendorf was then vortexed and centrifuged for two minutes. The supernatant was again removed before the eppendorfs were left to dry so that residual ethanol could evaporate. Next, 150µl buffer PKD was added and the samples were vortexed and centrifuged for 1 minute, then 10µl proteinase K was added and samples were mixed by pipetting up and down. Samples were then incubated for 15 minutes at 56ºC then at 80ºC for 15 minutes. Next, samples were incubated on ice for 3 minutes then centrifuged for 15 minutes at 13,500 rpm. The supernatant was transferred to a new tube leaving behind tissue debris and 16µl DNase Booster Buffer was added with 10µl DNase I stock solution. The tube was mixed by inverting and incubated at room temperature for 15 minutes. Next, 320µl Buffer RBC was added and the lysate was mixed thoroughly. Then 720µl 100%

ethanol was added which was mixed well and transferred (in two stages) to an RNAeasy MinElute spin column and centrifuged at 10,000 rpm for 15 seconds. The flow through was discarded and 500µl Buffer RPE was added and passed through by centrifugation at 10,000 rpm for 15 seconds. Next, the column was placed into a new collection tube and spun at full speed for 5 minutes with the lid open to fully dry the membrane. Finally, the columns were put into a clean 1.5ml Eppendorf and 17µl RNase-free water was added directly to the membrane to release the DNA. After centrifugation at full speed for 1 minute, the eluate containing the RNA was stored.

2.14.2 RNA library preparation and analysis

Library preparation, sequencing, alignment and analysis were performed by UCL Genomics. Library preparation for the 12 samples was performed using the TruSeq RNA access kit (Illumina) and run on the Illumina NextSeq 500. Reads were aligned using TopHat2 (Kim, Pertea et al. 2013) and counted using HTseq (https://pypi.python.org/pypi/HTSeq) in Python. Two samples were excluded from the analysis due to poor read coverage and aberrant clustering in unsupervised analysis. The Bioconductor package DESeq (http://bioconductor.org/packages/release/bioc/html/DESeq.html) was used in R to make group-wise comparisons between NK and NR, then NR and WT, run with default parameters.

Chapter 3: Genome-wide methylation analysis of normal kidney, nephrogenic rest and Wilms tumour

3.1 Introduction

To gain an overview of how DNA methylation differs between tissues, DNA methylation analysis was performed on normal kidney (NK), nephrogenic rest (NR) and Wilms tumour (WT). Chapter three describes how this data was obtained and the pre-processing and normalisation methods used that will be referred to again in later chapters. Unsupervised analysis of DNA methylation in these tissues was performed and will be discussed and finally, an analysis of the composition of the tissues included in this study was carried out to understand the importance of cell type composition contributing to overall methylation signal.

3.2: Generation and quality control of methylation data

In total, 120 samples (40 matched trios of normal kidney (NK), nephrogenic rest (NR) and Wilms tumour (WT)) were run on the Illumina 450k array, which interrogates ~480,000 CpG sites genome-wide at single base pair resolution. Clinical details of the patients included in this study can be found in Table 3.1. Although a large number of samples were originally included, several sampling problems, identified only after the arrays had been run, as well as samples failing the 450k quality checks resulted in a final dataset including 36 NK, 22 NR and 36 WT.

Quality control was carried out using R Bioconductor package Minfi with all algorithms run using standard parameters. Algorithms within this package take the raw data (.idat files that are output from the Illumina BeadScan software) and generate density plots and density bean plots so that the distribution of β-values between samples can be compared. By studying these plots, it was easy to identify samples with poor profiles that failed the array procedure (Figure 3.1). Overall, 5 samples failed and were removed.

After the arrays were run, the sections from which DNA was extracted were re-analysed and, after discussion between three paediatric pathologists, it was decided that 13 NR samples were to be excluded from the study even though they had produced good quality, usable data. Reasons for exclusion were extremely stringent and included that the area removed was contaminated with either NK or WT or that the pathologists could not be certain, from such a small piece of tissue, that the section was definitely NR and not WT showing post-chemotherapy changes. On top of this, five samples were extracted from tissue showing vast chemotherapy induced changes or large areas of necrosis, and therefore were also removed from the study. Finally, after checking the centrally reviewed histopathology reports (reviewed by Professor Gordan Vujanic in Cardiff), it was found that one patient's final review diagnosis was diffuse nephroblastomatosis that is morphologically very similar to WT but is a distinct entity; therefore this trio was also removed. A summary of the samples excluded from the study (of the 40 trios originally included) with the reasons for exclusion can be found in Table 3.2.

Patient number	Sex	Age at diagnosis	Kidney	Bilateral disease	Nephrogenic rest type	Overall tumour histology	Tumour stage
0	M	unknown	L	Bilateral	PLNR	mixed	5
119	F	39	R	Bilateral	PLNR	stromal	2
152	F	33	R	Bilateral	PLNR	mixed	3
209	U	58	U	Unilateral	PLNR	mixed	2
230	F	12	L	Unilateral	PLNR	mixed	1
250	M	10	R	Unilateral	PLNR	mixed	1
253	M	41	R	Unilateral	PLNR	mixed	3
257	M	21	U	Unilateral	PLNR	regressive	4
264	M	44	R	Unilateral	ILNR	regressive	4
266	M	10	L	Bilateral	ILNR	mixed	1
272	M	47	R	Unilateral	PLNR	blastemal	3
297	F	29	R	Unilateral	PLNR	mixed	3
314	M	10	L	Unilateral	PLNR	epithelial	1
323	F	11	L	Unilateral	PLNR	epithelial	1
324	M	10	R	Unilateral	PLNR	nephroblastomatosis	NA
332	M	44	L	Unilateral	ILNR	mixed	2
335	F	71	R	Unilateral	PLNR	regressive	3
340	F	10	R	Unilateral	PLNR	blastemal	2
350	M	42	L	Bilateral	PLNR	regressive	1
350	M	42	R	Bilateral	PLNR	anaplastic	1
381	F	54	R	Unilateral	PLNR	regressive	3
403	F	23	R	Unilateral	PLNR	mixed	1
404	F	66	L	Unilateral	PLNR	regressive	2
407	F	38	L	Unilateral	PLNR	regressive	3
450	F	29	L	Bilateral	PLNR	diffuse anaplasia	1
460	F	16	R	Bilateral	ILNR	epithelial	1
464	M	62	R	Unilateral	PLNR	mixed	1
467	M	47	L	Unilateral	PLNR	regressive	4
480	M	20	L	Unilateral	ILNR	stromal	1
483	F	28	R	Unilateral	PLNR	mixed	3
517	F	21	R	Bilateral	ILNR	stromal	2
580	M	39	L	Unilateral	PLNR	stromal	1
593	M	49	R	Unilateral	PLNR	necrotic	4
609	F	42	L	Bilateral	PLNR	regressive	3
651	F	31	L	Unilateral	PLNR	regressive	3
693	F	12	R	Bilateral	ILNR	stromal	3
722	F	144	L	Bilateral	ILNR	blastemal	3
754	F	20	L	Bilateral	ILNR	mixed	1
754	F	20	R	Bilateral	ILNR	regressive	NA
798	F	49	R	Unilateral	PLNR	blastemal	2

Table 3.1: Clinical and pathological information on the patients included in this study

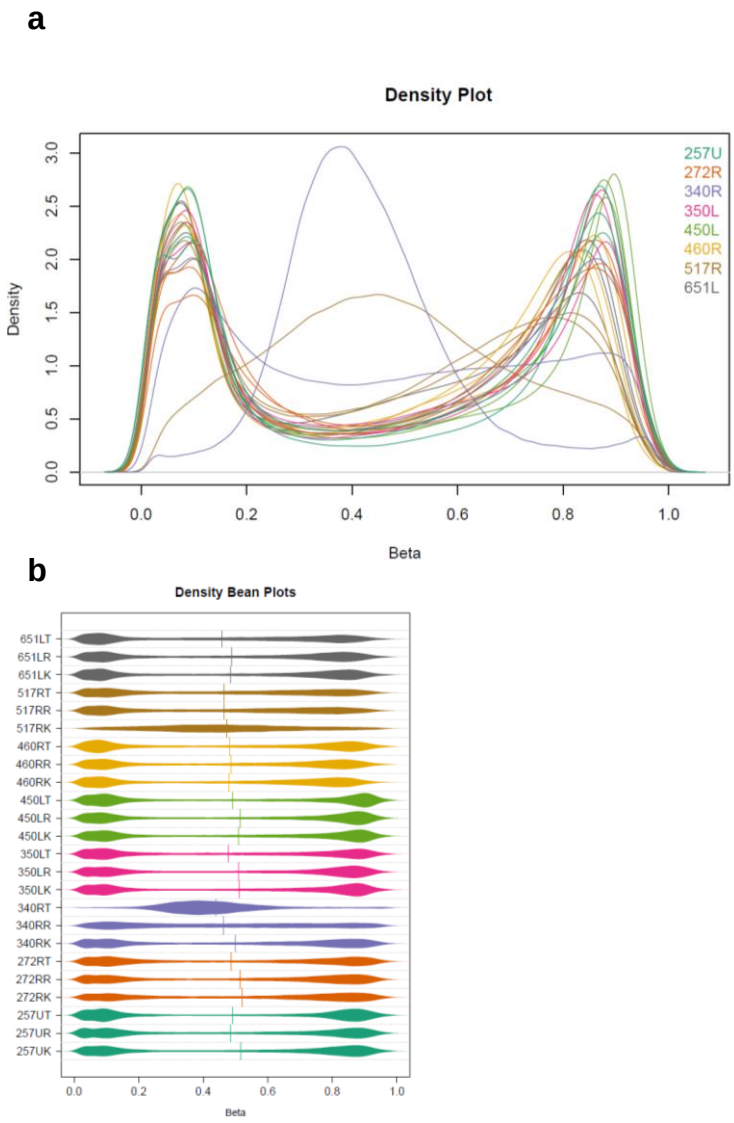


Figure 3.1: 450k quality control: β-value density plots. Both images were plotted using Bioconductor package Minfi. (**a**) Density plots of 8 trios coloured by patient shows that two samples have an aberrant distribution with extreme peaks at ~β=0.5. These do not fit the usual bimodal distribution which shows peaks at around β=0.1 and β=0.9. (**b**) Density bean plots show the same data in a different view allowing the exact samples that failed to be identified. In this case, 517RK (brown) and 340RT (blue) show abnormal density profiles and were excluded from further analysis.

Patient ID and side of tumour	NK	NR	WT
0L			
119R		Bad NR	
152R			
209R		Fail	Fail
230L			
250R		Bad NR	
253R		Bad NR	
257U		Bad NR	
264R		Bad NR	
266L		Bad NR	
272R			
297R			
314L			
323R		Bad NR	
324R	NB	NB	NB
332L		Bad NR	
335R			
340R		Fail	Fail
350L			
350R			
381R		Bad NR	
403R			
404L	Bad section		Bad section
407L		Bad NR	
450L	Bad section	Bad section	
460R			
464R			
467L			
480L		Bad NR	
483R		Bad NR	
517R	Fail		
580L			
593R		Bad section	
609L			
651L			
693R			
722L		Bad NR	
754L			
754R			
798R			
Total number of usable samples	36	22	36

Matched trio
Matched pair
Good sample
Not included

Fail	Failed 450k array QC
Bad NR	Nephrogenic rest considered not "true" after second analysis
NB	Nephroblastomatosis with no evidence of Wilms tumour
Bad section	Chemotherapy induced change or necrosis

Patient number (sample ID) followed by L=left and R=right indicating the origin of the nephrectomy specimen, U=unknown

Table 3.2: Description of the cohort of samples run on the 450k array

3.3 Initial data pre-processing and normalisation

To begin with, raw data for the 36 NK, 22 NR and 36 WT were loaded into R using the Bioconductor package ChAMP (Morris, Butcher et al. 2013) and with the same package, probes that showed poor detection (with detection p-value >0.01) in any sample were removed. Next, data was normalised using Bioconductor wateRmelon algorithm Dasen (Pidsley, CC et al. 2013), which corrected for the two probe types included on the 450k array. For comparative analysis of methylation values, as non-matched samples were included, the dataset was stringently curated to remove "confounding" CpGs that have a SNP either at the target C or G (according to the database of SNPs v131; http://www.ncbi.nlm.nih.gov/projects/SNP/), those that are located on the sex chromosomes and those that bind repetitive regions or to more than one loci, as defined by in silico analyses (Price, Cotton et al. 2013), which, for this dataset, left β-values (ranging from 0=unmethylated to 1=methylated) for 327,628 CpGs and 94 samples.

3.4 Validation of the 450K platform

To validate the methylation levels derived from the 450k array platform, as the starting material was FFPE and therefore very degraded, bisulfite sequencing was performed. A total of five regions covering 18 CpGs were selected (Table 3.3; columns 1-4), chosen to cover a wide range of 450k β-values ranging from 0.05 to 0.94 including intermediate methylation levels. DNA from four trios (12 samples) was bisulfite-converted and sequenced using the Illumina MiSeq. Raw fast-q files were aligned using custom Perl scripts and visualised using Integrative Genomics

Viewer (https://www.broadinstitute.org/igv/home). A text file giving the number of C and T reads for every loci was also generated for easy comparison.

The number of sequencing reads per bisulfite-converted PCR ranged from 9 to 57,335 (average = 13,649, median = 5,615), showing that the sequencing worked well. To enable comparison between 450k β-values and the sequencing data, the number of reads with a "C" residue at the target CpG site was calculated then this value was divided by the total number of reads for that CpG. As bisulfite treatment converts any unmethylated CpG residues to TpGs, the proportion of CpGs directly related to the level of methylation of a given site. For example if 2,000 reads were generated in total, and 40 were C while 1,960 were T, the level of methylation corresponds to 40/2,000 = 0.02. This put the sequencing data on the same scale as the 450k β-value which is measured from 0 to 1, as a sample with 2,000/2,000 C reads generated a methylation score of 1. Table 3.3 shows the difference in respective β-value for each sample at each site. After studying the number of reads per sample in more detail, along with the corresponding Δβ between platforms, the data was filtered to remove sites with <100 reads. After this, the β-values generated by each platform correlated well (R = 0.8365; Figure 3.2) with a median difference in β-value of 0.09, (min = 0.0001, max = 0.70) and the 450k methylation data was therefore considered reliable and analysed further.

Probe	Gene	MAPINFO	CHR	230LK	230LR	230LT	250RK	250RR	250RT	350RK	350RR	350RT	460RK	460RR	460RT
				Absolute change in methylation level between 450k array and bisulfite sequencing											
cg08439149	PRKCZ	2004163	1	0.05	0.62	0.06	0.1	0.11	0.13	0.01	0.02	0.08	0.69	0.05	0.06
cg13077150	PRKCZ	2004222	1	0.26	0.64	0.02	0.1	0.16	0.18	0.08	0.08	0.16	0.75	0.17	0.1
cg03874199	HOXD12	176964456	2	0.52	0.15	0.37	0.16	0.02	0.25	0.16	0.14	0.14	0.19	0.25	0.33
cg12268637	HOXD12	176964502	2	0.31	0.15	0.19	0.19	0.01	0.07	0.19	0.09	0	0.37	0.19	0.16
cg11754318	HOXD12	176964506	2	0.46	0.21	0.14	0.15	0.06	0.04	0.19	0.04	0.06	0.42	0.19	0.19
cg02694427	HOXD12	176964512	2	0.31	0.01	0.27	0.02	0.02	0.04	0.04	0.03	0.07	0.14	0.16	0.17
cg22926824	AGAP1	236761421	2	0.07	0.06	0	0.14	0.07	0.04	0.04	0.03	0.23	0.7	0.16	0.06
cg19579344	AGAP1	236761550	2	0.11	0.01	0.03	0.12	0.09	0.03	0.04	0.06	0.27	0.84	0.2	0.23
cg04134958	AGAP1	236761574	2	0	0.03	0.06	0.03	0.06	0.04	0.02	0.07	0.29	0.76	0.28	0.32
cg15718581	ZFP42	188916581	4	0.26	0.7	0.18	0.06	0.05	0.07	0.35	0.37	0.19	0.08	0.31	0.06
cg17214381	ZFP42	188916668	4	0.08	0.2	0.16	0.06	0.08	0.49	0.19	0.07	0.3	0.09	0.42	0.16
cg00663077	ZFP42	188916709	4	0.06	0.03	0.09	0.63	0.05	0.22	0.1	0.05	0.35	0.05	0.44	0.08
cg14817655	ZFP42	188916724	4	0.05	0	0.48	0.04	0.17	0.2	0.13	0.02	0.27	0.04	0.42	0.1
cg24896649	ZFP42	188916726	4	0.02	0.03	0.1	0.04	0.01	0.24	0.09	0.18	0.24	0.07	0.43	0.11
cg16762735	ZFP42	188916814	4	0.21	0.09	0.48	0.07	0.23	0.42	0.03	0.28	0.19	0.16	0.43	0.19
cg23082339	PCDHGA4	140810051	5	0.08	0.02	0.09	0.37	0.11	0.05	0.01	0.07	0.08	0.11	0.46	0.07
cg21627409	PCDHGA4	140810106	5	0.13	0	0.04	0.04	0.09	0.08	0.09	0.06	0.06	0.12	0.26	0.02
cg02452944	PCDHGA4	140810109	5	0.1	0.19	0.04	0.04	0.06	0.09	0.06	0.1	0.01	0.14	0.15	0.1

Table 3.3: Comparison of methylation levels detected by 450k array and bisulfite sequencing

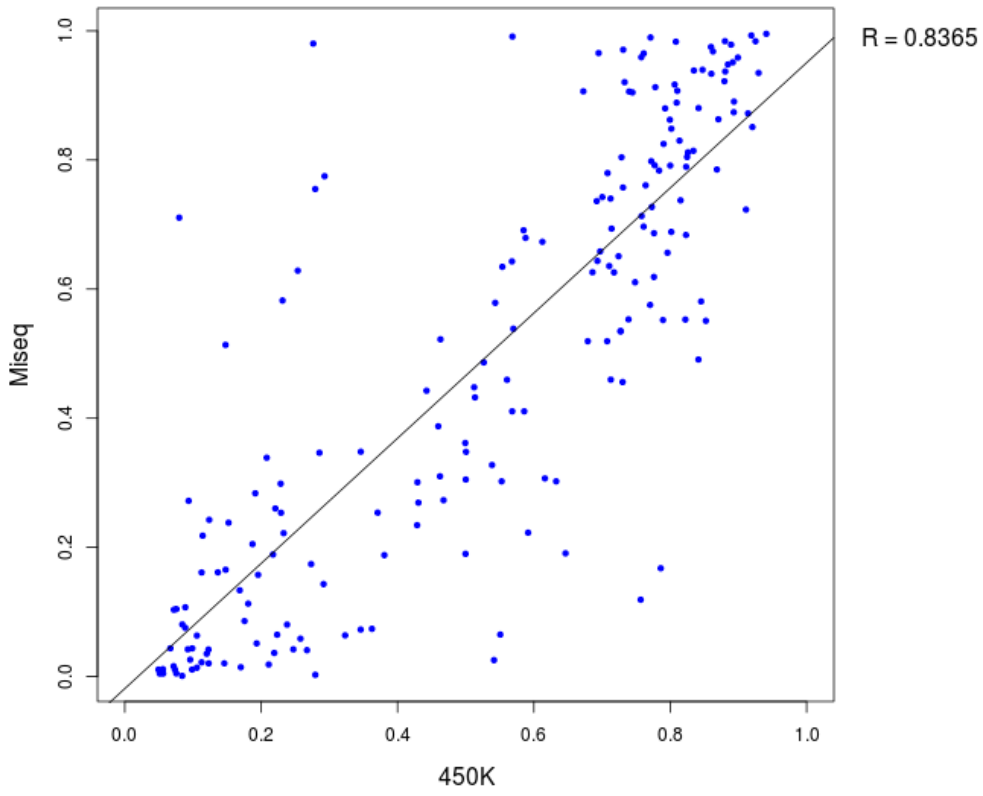


Figure 3.2: validation of array β-values by bisulfite sequencing. The β-values discerned by the 450k array (x-axis) compared to level of methylation detected using bisulfite-sequencing (y-axis) show good correlation (Pearson correlation coefficient = 0.8365).

3.5 Unsupervised analysis of methylation data

Next, the 94 normalised samples were subjected to a variation filter in R, using Bioconductor package genefilter (http://www.bioconductor.org/packages/release/bioc/html/genefilter.html), which pulls out a defined number of CpGs (f) that show the most variability across samples. Arbitrarily, f=3,277 was set to capture the 1% most variable CpGs across the dataset and multidimensional scaling (MDS: Figure 3.3) and unsupervised hierarchical clustering using heatmap.3 (http://www.inside-r.org/packages/cran/GMD/docs/heatmap.3; Figure 3.4) were performed to assess the relationship between samples.

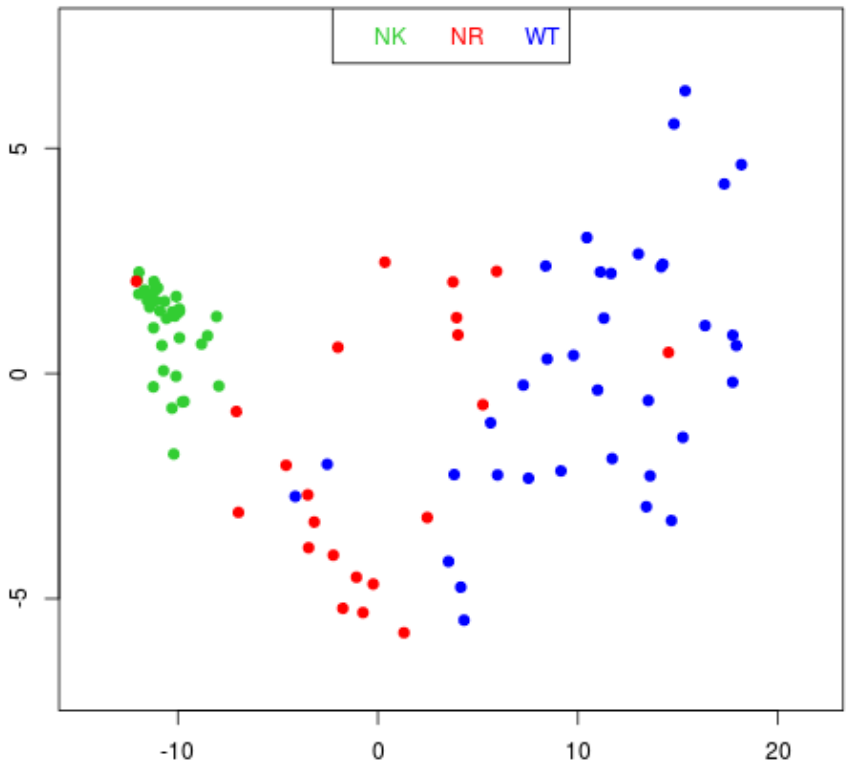


Figure 3.3: Multidimensional scaling of the top 1% most variable positions across the full dataset. Multidimensional scaling of the top 1% most variable positions in normal kidney (NK; green), nephrogenic rest (NR; red) and Wilms tumour (WT; blue)

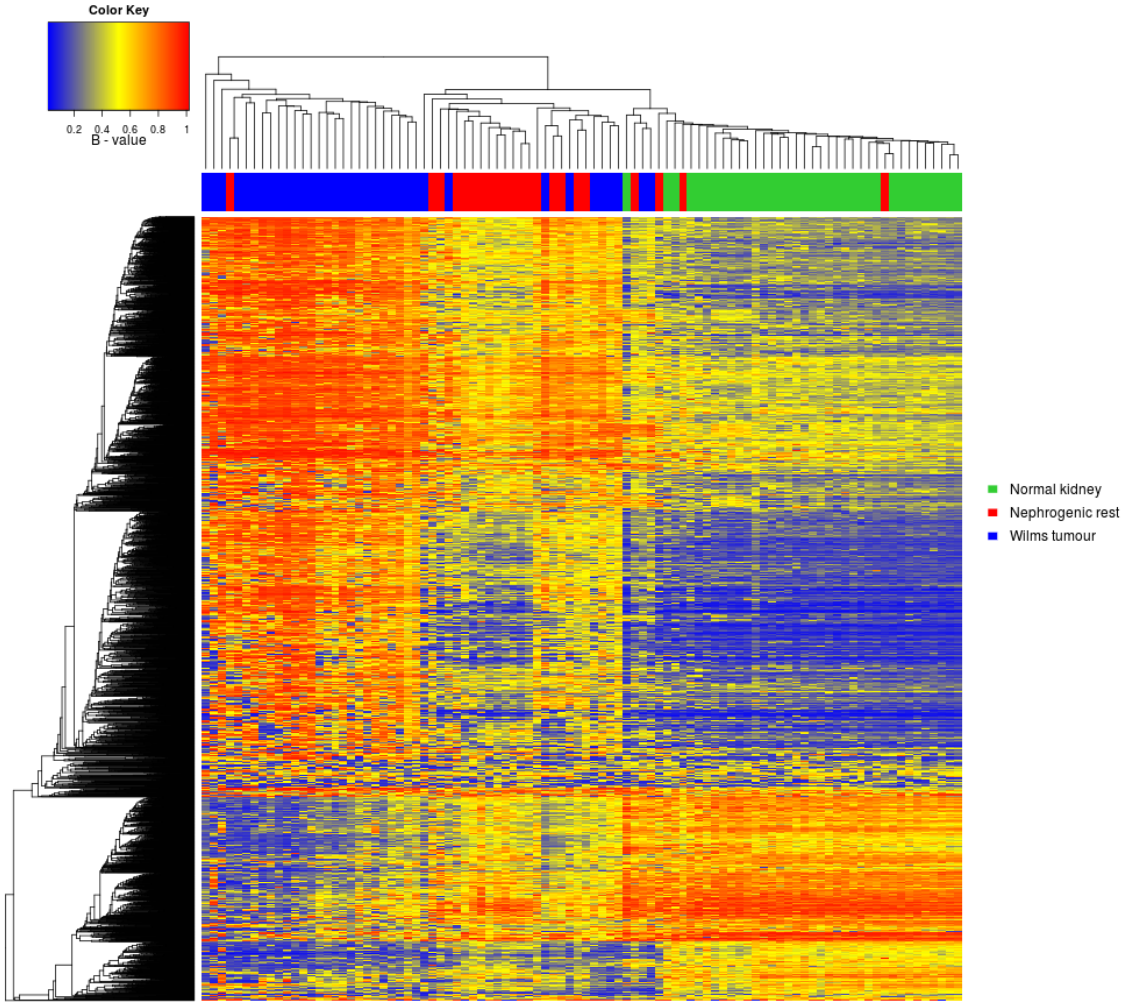


Figure 3.4: Hierarchical clustering of the top 1% most variable positions across the full dataset. Unsupervised consensus clustering and heatmap of the top 1% most variable positions across the full dataset (36 NK, 22 NR and 36 WT) as determined based on interquartile range where blue = low levels of methylation and red = high levels of methylation. Three clusters were formed which predominantly separated tissue types.

MDS generally separated samples into distinct groups although a wider distribution of samples was seen in both NR and WT groups in comparison to NK. This may be associated in an increase in epigenetic plasticity seen in samples as they progress towards malignancy or may be due to the increase in cell types seen which adds variability.

Hierarchical clustering revealed clear separation of samples into tissue-related groups, confirming the significant association between tissue type and methylation. Of the two clusters formed, cluster 1 included 26 WTs and 1 NR. The original section from which this NR tissue was taken is no longer available, however was previously confirmed by three pathologists as being composed of definite NR tissue.

The second cluster separated all normal kidneys into one cluster with 4 NRs and 2 WTs, then the remaining 16 NRs and WTs into another. One NR clustered very closely with NKs (0LR). The FFPE section from this case was re-analysed to confirm the presence of NR (Figure 3.5A). Although each sample contained differing proportions of cell types (an issue that will be addressed later), this preliminary evidence suggests that these three tissue types – NK, NR; the benign proliferating tissue and WT; the fully malignant tissue – have distinct patterns of DNA methylation.

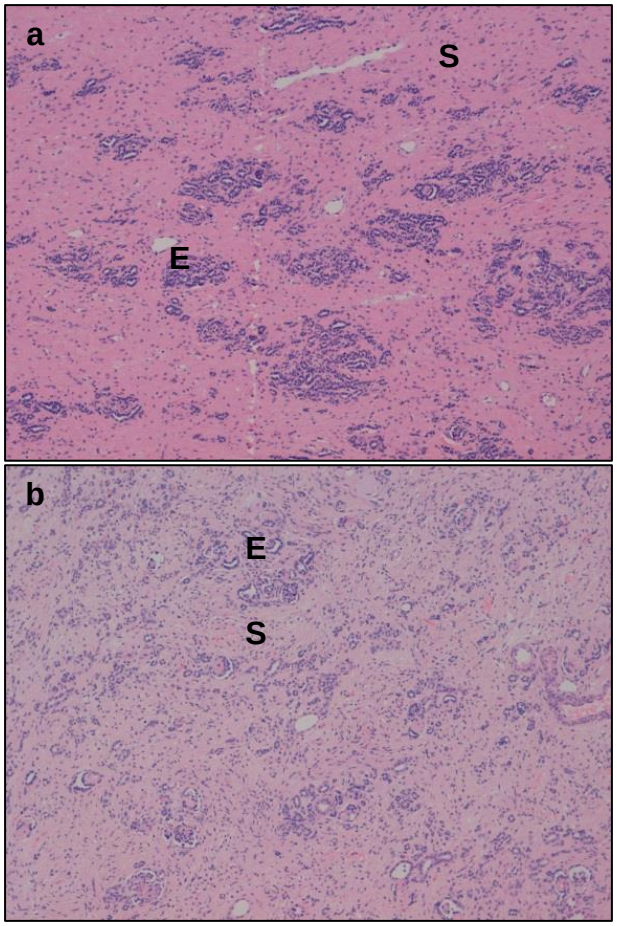


Figure 3.5: Review of nephrogenic rest FFPE sections. Two NR samples are shown to X10 magnification: (**a**) The NR that clustered with normal kidney and (**b**) a NR with similar composition that clustered with the other NRs. Both images show regions of stroma (S) and immature epithelial structures (E) without the presence of blastema. There is no contaminating NK present in (**a**), indicating that the epigenome, not cellular composition, is responsible for aberrant clustering.

3.6 Unsupervised analysis of 20 matched trios and embryonic kidney shows that EK is most similar to NK

After demonstrating that DNA methylation can distinguish tissue types, the 20 matched trios were separated to perform comparative analyses. As embryonic

kidney (EK) is the predicted cell-of-origin of both NR and WT, and both show morphological features that closely resemble EK, DNA methylation data from four embryonic kidney (EK) samples were included (samples were extracted and run by a colleague who supplied the raw .idat files).

Each time the number of samples changes, a new pre-processing step must be performed as the number of probes that fail the detection p-value is dependent on the samples present. Furthermore, normalisation had to be repeated to include the EK samples. The same procedure as above was used: data was imported using ChAMP then normalised using Dasen, which uses a between-array normalisation based on the control probe intensities on the array. This is necessary to prevent confounding artefacts as the EKs were run in a separate batch. After pre-processing, normalisation and removal of "confounding" CpGs, a data matrix of 330,731 CpGs was obtained for 64 samples. MDS (Figure 3.6) and hierarchical clustering (Figure 3.7) were performed with f=3,307.

MDS again separated samples into tissue-related groups and, of interest, clustered EK samples together far away from NK, EK and WT. Several WT and NR samples seemed to tend towards the EK samples in terms of separation along the Y-axis indicating that some samples showed features more similar to EK than others.

Hierarchical clustering separated samples into two predominant clusters (Figure 3.7). Cluster 1 was WT-predominant (14 WTs and 1 NR), cluster 2a contained predominantly "healthy" tissues (4 EK, 20 NK and 3 NR) and cluster 2b the majority of the NRs (16 NRs and 6 WT).

These unsupervised analyses show that EK is more similar to NK than to either NR or WT and that the majority of NRs are more similar to NK than WT. It was of interest that EK and WT did not cluster together, which may be due to the lower levels of blastemal elements in EK, or that WT is epigenetically not very similar to EK. Furthermore, one has to remember that these are just the most variable CpGs and not representative of genome-wide similarities or differences. This clustering also showed samples that seemed to be outliers, including 3 NRs in the NK cluster and 1 NR in the WT cluster. Re-analysis of FFPE sections showed no remarkable features with all 4 of these NRs containing 12-60% stroma and the rest epithelia with no blastema or chemotherapy induced changes present. It can be hypothesised that the 3 NRs that cluster with NK show more differentiated epithelia, not visible by histopathological review, but present epigenetically. Likewise, the NR that clustered with the WTs may show neoplastic features not visible by histopathologial review, but present in the epigenome.

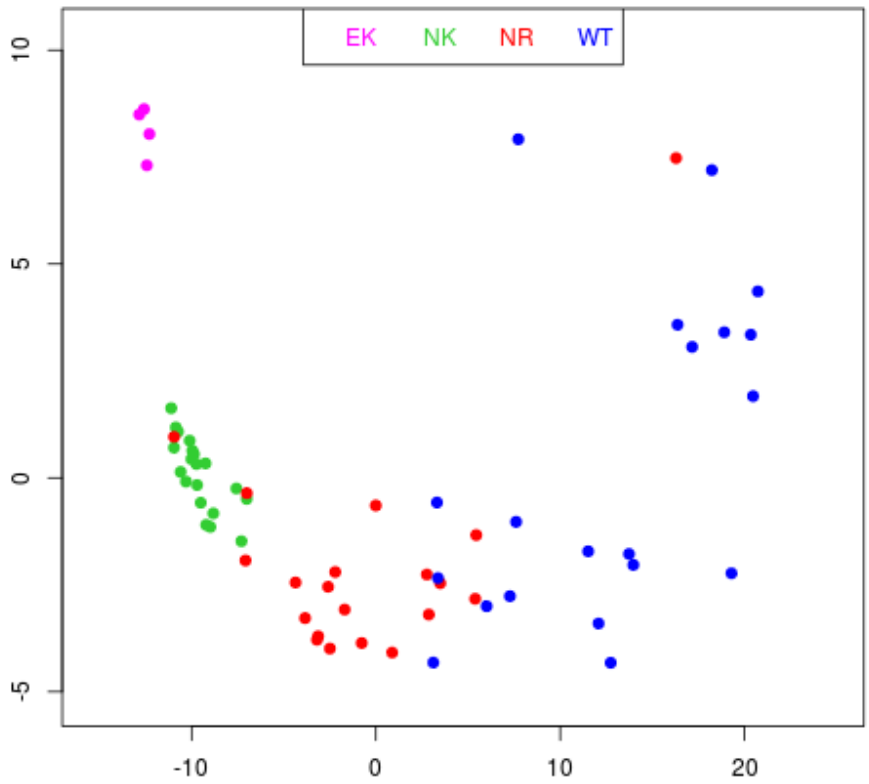


Figure 3.6: Multidimensional scaling of the top 1% most variable CpGs between trios of NK, NR, WT (n=20) and EK (n=4). Multidimensional scaling of the top 1% most variable positions in embryonic kidney (EK; pink), normal kidney (NK; green), nephrogenic rest (NR; red) and Wilms tumour (WT; blue). It can be predicted that the component separating samples on the X-axis is tumourigenicity as normal samples are on the left (NK and EK), then the cancer precursor lesions are in the centre with WT samples on the far right. Alternatively, the component separating samples on the Y-axis may be blastema, as both NK and NR lack blastema while WTs and EK do have blastema cells.

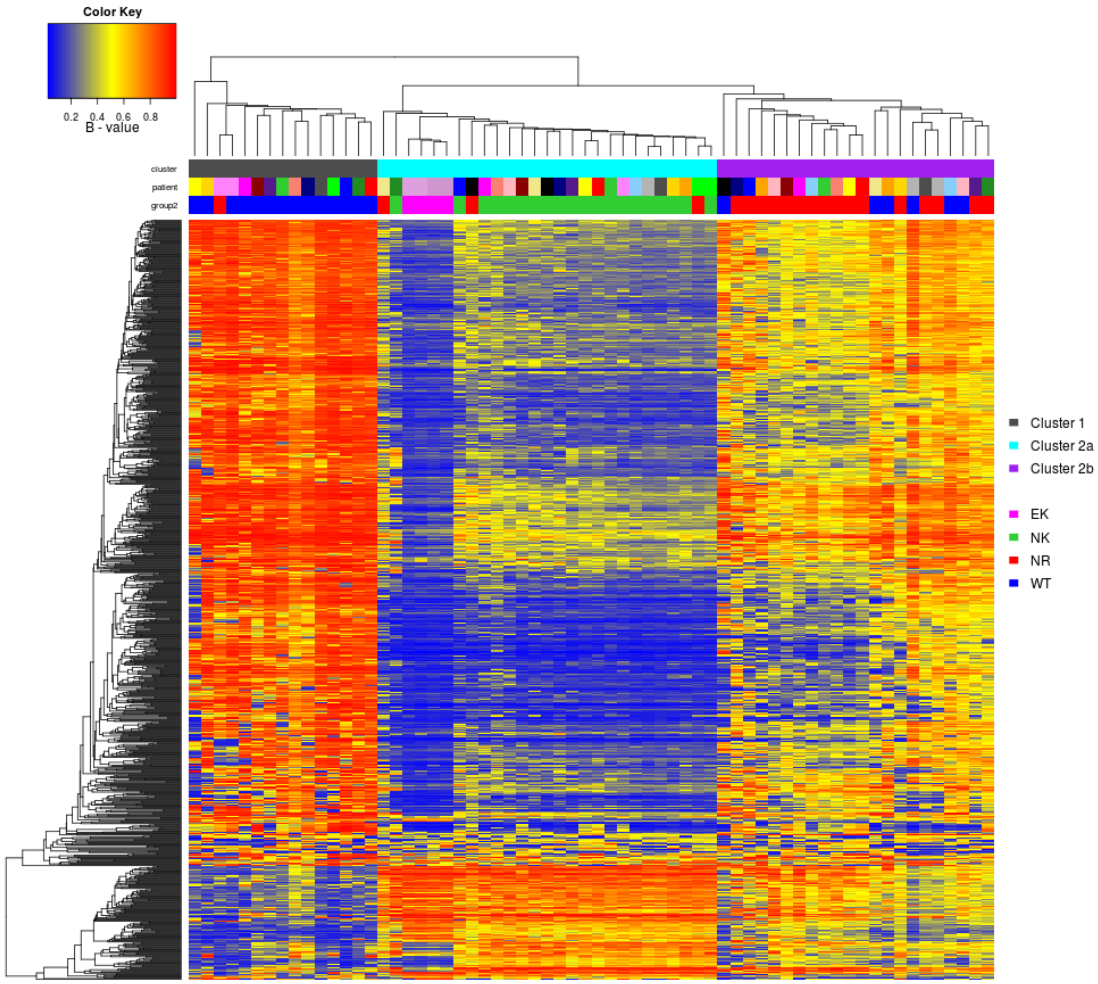


Figure 3.7: Hierarchical clustering of the top 1% most variable CpGs across trios of NK, NR, WT (n=20) and EK (n=4). Unsupervised consensus clustering of the top 1% most variable positions across the full dataset as determined based on interquartile range. Three clusters were formed which predominantly separated tissue types. Each sample is labelled by the cluster they fell into (cluster 1; grey, cluster 2a; cyan, cluster 2b; purple), by the patient from which the tissue originated (there are 20 matched trios), and by the histology (EK; pink, NK; green, NR; red and WT; blue). Samples are not clustered by patient but by histology as cluster 1 is WT-predominant, 2a is NK and EK-predominant and 2b is NR-predominant.

3.7 Analysis of tissue composition

Although clear separation of tissues was seen by unsupervised analysis of DNA methylation, when performing comparative methylation analysis, it is important to consider that the composition of the tissue will contribute to the overall methylation signal. Each cell type has a cell-type specific methylation pattern, which regulates the expression of cell-specific genes. The overall methylation signal obtained will consist of the cell-specific methylation signatures in the relative proportions of the cells present. In terms of WT, the methylation signal, although generated from mixed cell proportions, is a biomarker for the tissue, which is inherently heterogeneous.

To understand the contribution that the mixed cell types may have on methylation in this dataset, each microdissected section for the 20 matched trios was analysed to identify the proportion of stroma, epithelia and blastema cells present (Figure 3.8). Although not included in the figure, NK sections, which are far more homogenous, were also studied and contained roughly 5% stroma and 95% epithelia in every case.

This analysis showed that very few NR samples had regions of blastema (6/20) whereas many more WTs did (11/20). WTs also tended to show more necrosis and chemotherapy-induced changes. The fact that WTs have more blastema may confound between-tissue analyses as CpGs associated with blastema cells will contribute to the overall methylation signal in many more WT samples than NR samples. Furthermore, as NKs have no blastema, analyses comparing to this tissue will also be confounded by blastema-specific CpG methylation. The same confounding would be seen by comparing samples with high levels of epithelia (for example) against samples with low levels, such as NK comparison to WT.

Consequently, comparative analyses between tissues will identify true biomarkers of tissue composition, which will be described in chapter 4, but analysis with correction for cell type composition will allow identification of underlying changes in methylation associated with progression towards malignancy instead of cell type. These will be identified and discussed in chapter 5.

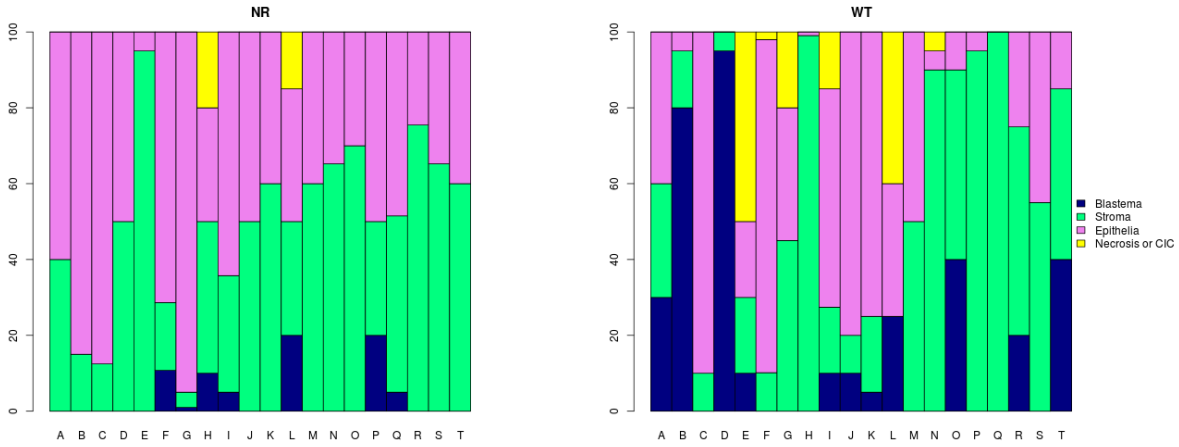


Figure 3.8: Cellular composition of each microdissected tissue section. Each of the 20 NRs (left) and WTs (right) were analysed for the proportions of each cell type within the section that was microdissected. Here, blastema is shown in navy, stroma in green, epithelia in pink and necrotic or chemotherapy-induced changes in yellow.

3.8 Summary and conclusions

To summarise, this chapter has described the generation of methylation β-values, the quality control, pre-processing and normalisation steps required prior to data analysis as well as unsupervised analysis of NK, NR, WT and EK samples. The stringency of sample selection was discussed which is essential for biological interpretation of methylation data.

These preliminary analyses showed that DNA methylation can distinguish between NK, EK, NR and WT tissues. The β-values derived by the 450k platform were validated by bisulfite sequencing showing good concordance. MDS of the 1% most variable CpGs in the dataset showed an increase in methylation variability between NR and WT samples compared to NK samples and hierarchical clustering showed that EKs were more similar to NK than to NR or WT. This was surprising but may be explained by the origin of the tissue. The EKs included in this analysis were derived from aborted specimens of whole kidney which, although they contain embryonic blastema and immature epitheial structures, also contain differentiated tubules, nephrons and glomeruli. Therefore, the methylation profile derived from the whole sample may be more similar to NK due to the signal from the mature kidney elements over the immature.

Finally, analysis of sample-specific cell type compositions showed that NRs and WTs varied greatly between samples. This cell type composition heterogeneity will be a recurring theme throughout this thesis, particularly addressed in chapter 5.

Chapter 4: Comparative analysis of DNA methylation in Wilms tumours, nephrogenic rest and normal kidney

4.1 Introduction

As shown in chapter three, DNA methylation status clearly distinguished between tissue types (NK, NR and WT). To understand the biological significance of between-tissue differences in DNA methylation, the 20 matched trios were subjected to several supervised analyses to identify CpGs that differ in methylation between WT and NR (transformation-associated) and between NR and NK (developmental arrest-associated). The biological significance of these changes was interpreted using bioinformatics pathway analysis tools and validated using RNA-sequencing.

4.2 Sample variability

To begin with, pre-processing and normalisation steps were carried out as described in chapter 3.2 using SWAN normalisation instead of Dasen, for the 20 matched trios. This dataset encompassed 453,385 CpGs and 60 samples. As observed in chapter 3, and within this dataset of only 20 trios, multidimensional scaling of the top 1% most variable data points (of a dataset with the SNPs and sex chromosomes removed to avoid confounding) showed wide dispersion of NR and WT samples in comparison to NK (Figure 4.1). To quantify this degree of variability, a Bartlett test was performed in R which quantified probe-specific

variances between NK and NR and then NK and WT. In each comparison, those probes that showed significant non-equal variance (N; $p<0.01$) were predominantly due to increased variance in the NR and WT groups compared to NK ($N_{NK>NR}$ =9,334; $N_{NR>NK}$ =94,546 and $N_{NK>WT}$ =14,933; $N_{WT>NK}$ =158,189; see Figure 4.2).

An increase in methylation variability as normal tissues progress towards cancer from a premalignant stage has been previously observed in adult colon adenocarcinoma (Timp and Feinberg 2013). The study associated gain of methylation variability with cancer cell heterogeneity however in adenocarcinoma, the cell-of-origin to cancer progression via the pre-malignant precursor stage is linear whereas in WTs, NK does not represent the cell-of-origin and a much wider variety of cell types are seen at both the precursor and tumour stages. This large difference in the histological composition between NRs and WTs compared to NK which is generally very uniform with ~5% stroma and ~95% epithelia results in major cell type composition effects that will largely affect the overall level of variability. Therefore, although this increase observed is likely true and associated with a more hypervariable epigenome in the NR and WT, the extreme increase in methylation variability is also contributed to by tissue composition effects. This large confounding factor will be addressed in chapter 5 which described analyses performed with correction for cell mixture effects.

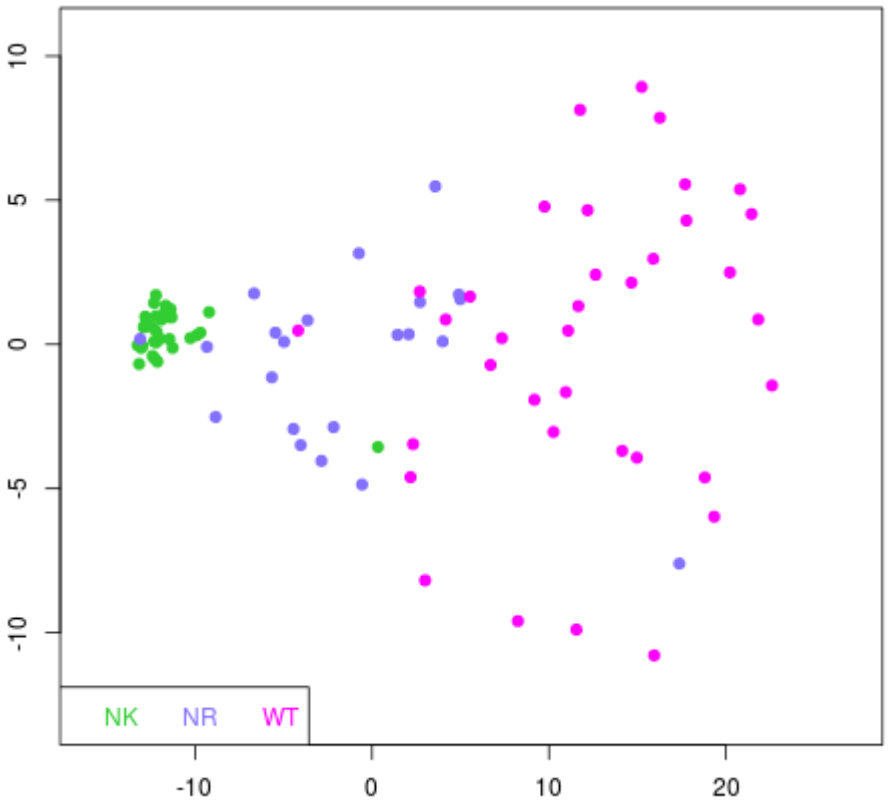


Figure 4.1: Nephrogenic rest and Wilms tumour groups show greater between-sample variability than normal kidney. Multidimensional scaling of the top 1% most variable positions (selected using a variance filter based on interquartile range) showed greater dispersion of points within both the NR and WT datasets compared to NK suggesting a greater degree of between-sample variability.

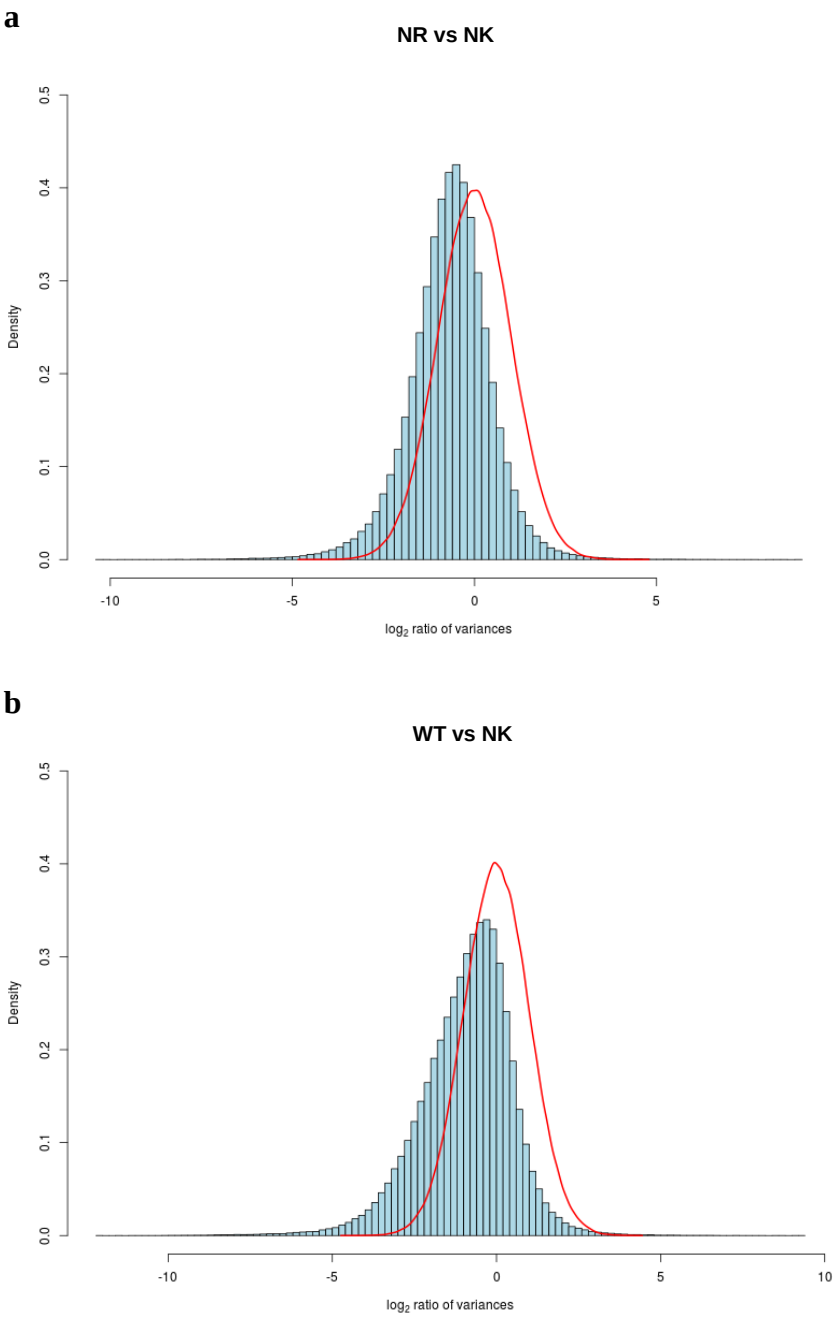


Figure 4.2: Probe-wise variance between groups shows NR and WT are much more variable than NK. The x-axis shows the log2 ratio of variance which is a measure of the probe-specific difference in variance between NK and either (a) NR or (b) WT. The density of probes with each level of variance is shown by the blue bars. If both sets of probes showed equal variance, the null distribution would be observed (shown in red), however in both figures, a shift of the observed data (blue bars) is seen indicating that the probe-specific level of variance is not equal for both the NR and WT groups in comparison to NK. In both cases, the NK group was substantially skewed toward decreased variability (a shift to negative x-axis values) compared with NR groups (skew = -87.89, p <

$2.2x10^{-16}$, D'Agostino skewness test) and WT groups (skew = -160.10, $p < 2.2x10^{-16}$, D'Agostino skewness test).

4.3 Supervised analysis identifies two Wilms tumour groups

To identify those CpGs that most significantly distinguish each tissue type, the full data matrix of 453,385 CpGs and 60 samples was subjected to ANOVA using the Bioconductor package Limma (Smyth 2005). In total, 7,921 CpGs with genome-wide significance ($P<5x10^{-8}$) were identified and were clustered using heatmap.3 (Figure 4.3) which revealed two predominant clusters: cluster 1 (13 WT, 1 NR) and cluster 2 which further separated into cluster 2a (20 NK, 1 NR) and cluster 2b (7 WT, 18 NR). These results were similar to the results of unsupervised clustering seen in chapter 3 indicating the strong association DNA methylation has with tissue identity.

All NK samples clustered into the same group (cluster 2a) however, the WT samples fell into two distinct groups. WTs in cluster 1 were distinctly separated from their NRs whereas WTs in cluster 2b were clustered with their respective NR. Due to this difference, the WTs were termed group-1 and group-2 WTs respectively. To understand the reason behind why these WTs may be behaving in this way, biological and clinical information was studied in more detail. There seemed to be no association between the separation of WTs and proportions of any cell type, sex, tumour stage, overall histology or course of pre-operative chemotherapy (data not shown). However, all patients with bilateral disease were found in group-1. By performing a chi-square test, the relationship between grouping and bilateral disease

was found to be significant (p=0.032). As a consequence of this finding, the MDS plot was re-evaluated and clearly showed that group-2 WTs appear closer to the NR samples than group-1 WTs (Figure 4.4). As this MDS plot showed a wider dispersion of group-1 WTs, a Bartlett test was performed to compare levels of probe-specific variance between group-1 and group-2 WTs. This test showed that group-1 WTs had 2.4 times as many probes with a significance increase in variance compared to group-2 WTs (31,638 compared to 13,124; $p<0.01$) suggesting that group-1 WTs have a more hyper-variable epigenome.

To validate the finding that group-2 tumours appear epigenetically similar to their associated NR while group-1 WTs appear distinct, groups 1 (n=13) and 2 (n=7) were separated and linear modelling was performed within each group to compare matched NR-WT pairs. This analysis was carried out using Bioconductor package Limma and identified 22,344 significant CpGs that were differentially methylated between group-1 WT and NR pairs (FDR<0.01). Strikingly, no significant sites of differential methylation were identified in the comparison of group-2 WT and NR pairs. This evidence confirmed that group-1 WTs were epigenetically distinct from their NR, whereas group-2 WTs were not.

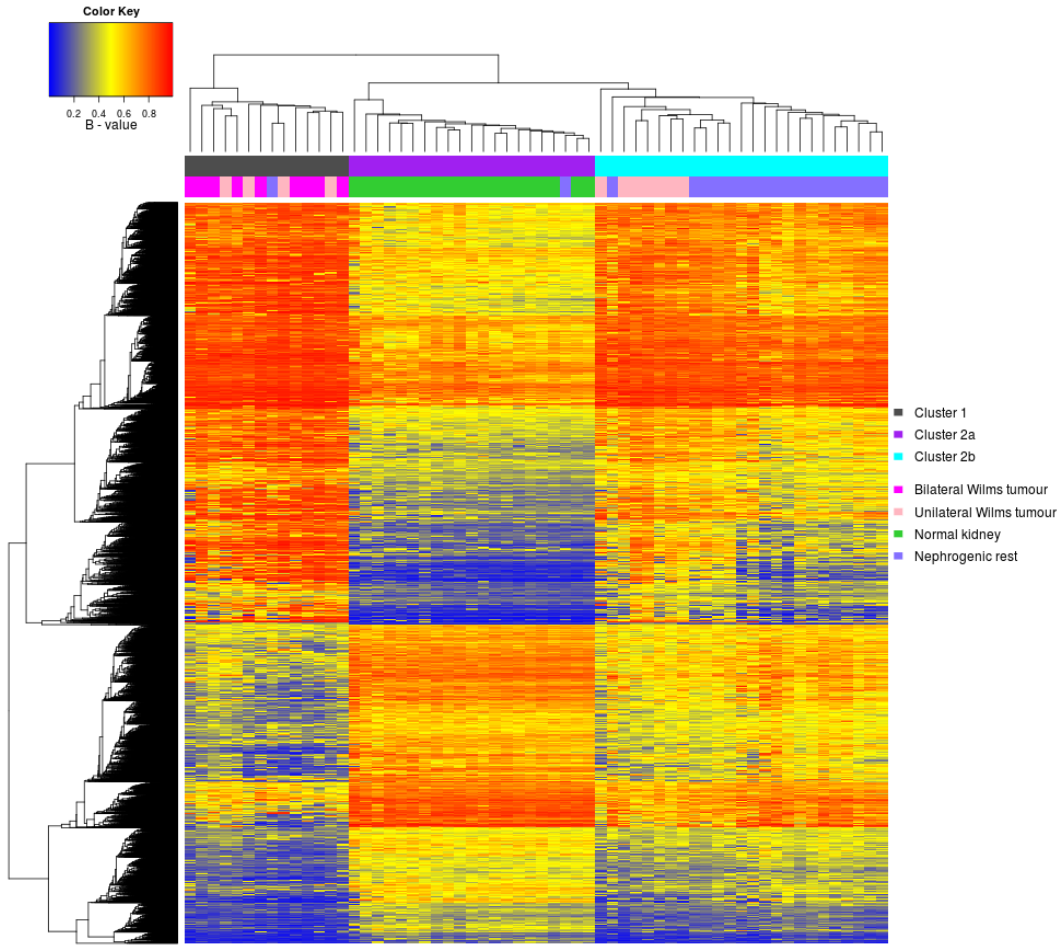


Figure 4.3: ANOVA analysis identifies two Wilms tumour groups. Consensus clustering of the significant CpGs (n=7,921; $P<5x10^{-8}$) from ANOVA analysis of 20 trios of normal kidney (green), nephrogenic rest (blue) and Wilms tumour (pink). Here, three clusters can be seen which show the presence of two distinct WT groups. Cluster 1 (grey) comprises group-1 WT (n=13) which includes all bilateral WT (dark pink) and 4 unilateral cases (light pink). Group-1 WT cluster separately from their associated NR. The second cluster further separates into two, with cluster 2a (purple) containing all NK samples and cluster 2b containing group-2 WT (n=7) which are all unilateral and cluster together with their associated NR.

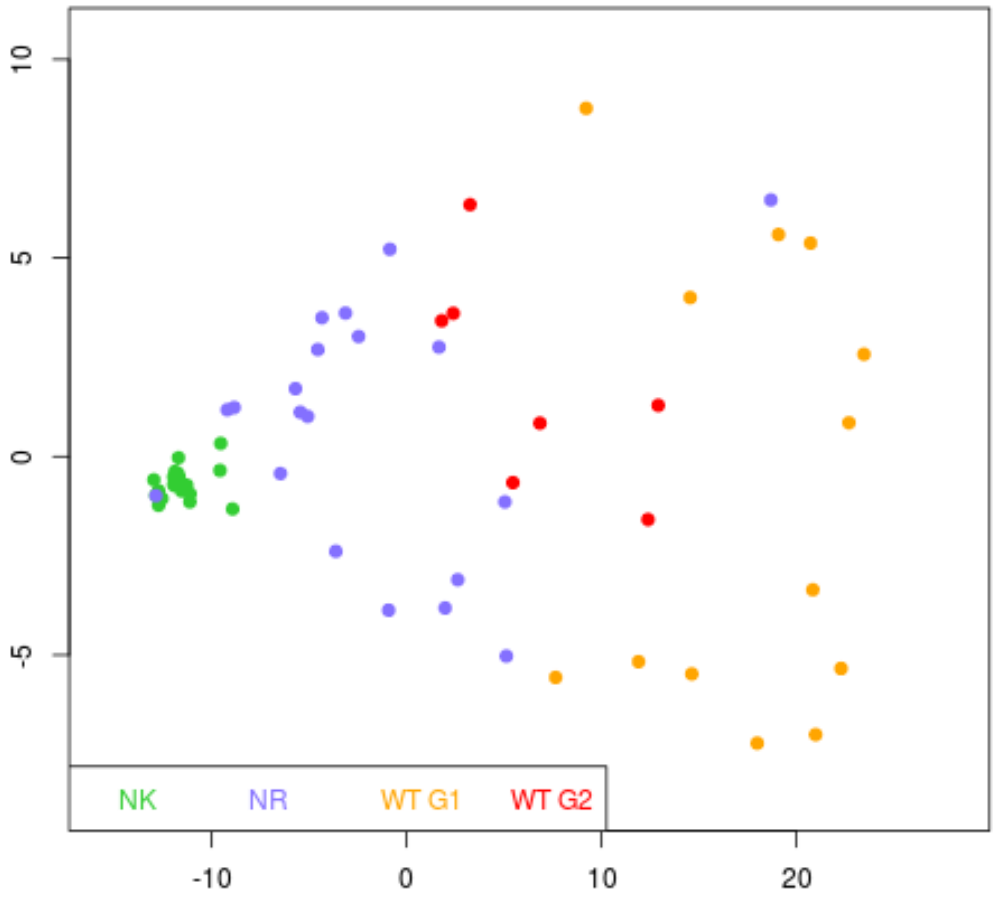


Figure 4.4: Unsupervised analysis shows two Wilms tumour groups. Re-evaluation of tumour separation by multidimensional scaling of the top 1% most variable positions showed separation of the two Wilms tumour groups 1 (orange) and 2 (red). Group-2 are more intermixed with the NR samples whereas group-1 are located further away and wide dispersion of samples is seen.

4.4 WTs silence genes involved in cell adhesion and regulation or transcription

As group-1 WTs appeared to show changes in methylation that render them distinct from their NRs, these significant CpGs were focussed on in more detail. To begin with, a novel algorithm was applied that groups significant CpGs into discrete clusters depending on the local CpG enrichment. The algorithm, champ.lasso, implemented through the Bioconductor ChAMP package (Morris, Butcher et al. 2013), identified 625 differentially methylated regions (DMRs) in total. The parameters used for DMR definition in this algorithm can be found in chapter 2.8.6, but briefly; a DMR consisted of at least three significant CpGs (FDR<0.01) in close proximity (which is assessed by CpG location as some are closely packed while others are widely separated).

Of these DMRs, 460 (73.6%) were hypomethylated and 165 (26.4%) were hypermethylated in WT with respect to NR, termed hypo- or hyper-WT-DMRs. As previously mentioned, global hypomethylation is a common feature of cancer suggesting that during the transformation step from NR to WT, the WTs acquire general cancer properties as evidenced by the higher proportion of hypo-WT-DMRs over hyper. The hypo-WT-DMRs were enriched in gene bodies and not associated with CpG islands, shores or shelves (summary in Table 4.1), suggesting a positive correlation between methylation and gene expression for these DMRs (Jones 2012). In comparison, the hyper-WT-DMRs were much smaller and were often located at gene transcription start sites or within CpG island shores which indicates an inverse relationship with regulation of gene expression (Doi, Park et al. 2009; Irizarry, Ladd-Acosta et al. 2009) (Table 4.1).

To infer biological significance from these DMRs, the location of the DMR CpGs were entered in bed format into Genomic Regions Enrichment of Annotations Tool (GREAT) version 2.0.2 (http://bejerano.stanford.edu/great/public/html/) which associates genomic positions with gene regulatory domains. The significant ($p<0.01$) gene ontology biological processes were studied to identify common processes with more than 2-fold enrichment. This analysis showed that hypo-WT-DMRs fell into regulatory regions affecting genes involved in developmental processes including metanephric nephron development and nephron development involving genes such as *GDNF*, *IRX2*, *PDGFB*, *POU3F3* and *SOX8* and in processes involved in stem cell maintenance, development and differentiation (Table 4.2). The same analysis for hyper-WT-DMRs found enrichment for genes involved in cell adhesion and regulation of transcription (Table 4.3).

These data suggests that genes involved in renal development and stem cell maintenance are repressed in group-1 WTs compared to NRs, suggesting that NRs are in a more primitive state than WTs. Hypermethylation (and reciprocal silencing) of genes involved in cell adhesion would allow cells to overcome the local microenrironment and proliferate as a tumour.

As a proof of concept that RNA extracted from these FFPE samples could give quality interpretable sequencing data, RNA sequencing was performed on four trios. Although a much larger study is required to rigorously test differential expression between tissue types, promising results was seen. After alignment and unsupervised analysis, one pair of matched NR and WT samples failed the QC and was excluded. The remaining NK was included in group-wise comparisons, which were performed using the Bioconductor DESeq package in R

(http://bioconductor.org/packages/release/bioc/html/DESeq.html). The RNA sequencing data showed 75 genes with significant differential expression between NR and WT (FDR<0.05) which were analysed with Panther (http://www.pantherdb.org/) to identify common pathways. Even within this small list of genes, the over-enriched gene sets identified through interrogation of hyper-WT-DMRs were validated nicely as six genes involved in cell adhesion or cell signalling (*CD200*, *GPR108*, *TSPAN2* and *ADAMTS8* decreased in expression while *MDK* and *NCAM1* increased in WT compared to NR) and three genes involved in regulation of transcription (*NFKB1*, *MYSM1*, *PREPL* decreased in expression) were found to be significantly differentially expressed validating the relationship between differential DNA methylation and gene expression. One of these genes, *NCAM1* has previously been associated with being a marker for cancer-propagating WT cells, supporting the association with the transformation step between NR and WT tissues (Pode-Shakked, Shukrun et al. 2013).

Regarding the hypo-WT-DMRs, where enrichment for stem cell processes was observed, the genes differentially methylated in particular, *GDNF*, *IRX2*, *PDGFB*, *POU3F3* and *SOX8* showed no significant difference in expression across samples. Reasons for non-concordance could be due to the very small number of samples included in the validation set. Likewise, no genes involved in renal development were found to be significantly differentially expressed. In this small sample set, the relationship between DNA hypomethylation and gene expression still requires further validation.

	Hyper-methylated DMRs	Hypo-methylated DMRs
Median size (bp)	527	1555.5
Mean size (bp)	1130.4	1749.8
Range (bp)	108-5,118	110-6,780
Total (bp)	186,507	804,907
% enrichment: CpG island	0.8	-22.3
% enrichment: CpG shore	5.5	-0.5
% enrichment: CpG shelf	-2.7	2.2
% enrichment: 'other'	-3.6	20.6
% enrichment: gene body	-3.5	14.6
% enrichment: TSS200	1.6	-5.6
% enrichment: TSS1500	-3.5	-10.4
% enrichment: 3' UTR	-2.7	-1.3
% enrichment: 5' UTR	-1.6	-1.6
% enrichment: 1st exon	-1.7	-3.0
% enrichment: intergenic region	11.5	7.2

Table 4.1: Description of hypermethylated and hypomethylated differentially methylated regions (DMRs) in group-1 Wilms tumour compared to matched nephrogenic rests. Genetic features and those relating to CpG islands, shores (within 2kb of islands) and shelves (2-4kb from CpG islands) are extracted from the Illumina 450k annotation.

Enriched Term in GO biological processes	Hypergeometric fold enrichment	Raw p-value
Axon ensheathment	2.5	1.87E-07
Body morphogenesis	3.04	1.56E-08
Cell chemotaxis	3.27	6.35E-13
Embryonic placenta development	2.74	6.61E-10
Fat cell differentiation	3.16	4.42E-14
Head development	2.77	1.44E-07
Head morphogenesis	3.54	3.44E-10
Intermediate filament-based process	5.44	4.39E-22
Labyrinthine layer development	4.07	9.57E-15
Metanephric nephron development	3.83	2.35E-11
Myelination	2.57	9.16E-08
Myelination in peripheral nervous system	4.65	1.18E-09
Negative regulation of myeloid leukocyte differentiation	6.43	1.48E-22
Nephron development	2.61	1.28E-07
Organ regeneration	3.86	1.14E-13
Palate development	2.8	2.45E-12
Peripheral nervous system development	2.05	6.43E-05
Placenta development	2.42	2.43E-09
Regeneration	2.39	2.63E-08
Regulation of action potential in neuron	2.19	4.73E-06
Regulation of GTP catabolic process	2	5.30E-14
Regulation of myeloid cell differentiation	2.69	3.62E-12
Regulation of myeloid leukocyte differentiation	3.64	2.07E-14
Regulation of purine nucleotide catabolic process	2.07	1.01E-16
Regulation of Ras GTPase activity	2.14	1.27E-10
Regulation of Ras protein signal transduction	2.02	7.60E-14
Schwann cell development	4.42	3.10E-09
Schwann cell differentiation	3.91	3.11E-08
Stem cell development	2.52	2.46E-09
Stem cell differentiation	2.03	2.01E-06
Stem cell maintenance	2.78	8.61E-11

Table 4.2: Significantly overrepresented biological processes identified by hypomethylated WT-DMRs

Enriched Term in GO biological processes	Hypergeometric fold enrichment	Raw p-value
Cell adhesion	2.17	3.45E-16
Cell-cell adhesion	4.83	3.00E-44
Homophilic cell adhesion	12.97	5.85E-81
Intracellular receptor mediated signalling pathway	8.07	2.23E-61
Negative regulation of biosyntheic process	2.35	5.46E-30
Negative regulation of cellular biosynthetic process	2.37	1.50E-30
Negative regulation of cellular macromolecule biosynthetic process	2.51	1.22E-33
Negative regulation of cellular metabolic process	2.13	3.76E-26
Negative regulation of gene expression	2.58	1.79E-36
Negative regulation of macromolecule biosynthetic process	2.44	3.08E-32
Negative regulation of macromolecule metabolic process	2.1	2.59E-25
Negative regulation of metabolic process	2.03	6.12E-24
Negative regulation of nitrogen compound metabolic process	2.5	1.70E-33
Negative regulation of nucleobase-containing compound metabolic process	2.51	1.04E-33
Negative regulation of protein metabolic process	4.19	1.92E-31
Negative regulation of RNA metabolic process	2.61	5.53E-36
Negative regulation of transcription from RNA polymerase II promoter	3.79	8.10E-54
Negative regulation of transcription, DNA-dependent	2.67	2.24E-37
Positive regulation of gene expression	2.06	4.97E-22
Positive regulation of RNA metabolic process	2.06	1.76E-21
Positive regulation of transcription from RNA polymerase II promoter	2.87	4.43E-35
Positive regulation of transcription, DNA-dependent	2.13	4.89E-23
Regulation of gene expression, epigenetic	5.98	2.39E-30
Regulation of organelle organisation	3.57	3.43E-34
Regulation of transcription from RNA polymerase II promoter	2.41	9.04E-40
Transcription from RNA polymerase II promoter	4.87	5.31E-49

Table 4.3: Significantly overrepresented biological processes identified by hypermethylated WT-DMRs

4.5 Assessing the pluripotency of WTs

As the methylation data strongly suggested that WTs have stem cell-like epigenomes but the RNA sequencing did not validate this, the potency of WTs was studied in comparison to NRs and NKs. Embryonic stem cells are pluripotent; they have the potential to differentiate into any of the three germ cell layers, endoderm, mesoderm or ectoderm and form teratomas when injected into mice. In WT, a mixed histology is frequently observed with occasional differentiation into skeletal, fat or bone tissue. These observations imply a cell of origin with multipotency and potentially pluripotency. Fully pluripotent cells express transcription factors *OCT4*, *NANOG*, *SOX2* and *KLF4* and can also be identified by the increase in methylation of non-CpG residues (CpA and CpT) (Ramsahoye, Biniszkiewicz et al. 2000) and increase in genome-wide hydroxymethylation (H-me). To assess the level of pluripotency in WT trios, comparison to publically available data, quantification of H-me and analysis of non-CpG methylation was carried out.

Publically available 450k methylation data were obtained online for embryonic kidney (EK; n = 4), induced pluripotent cell lines (IPS; n = 6) and embryonic stem cell lines (ESC; n = 6). These were put with the methylation data from the 20 matched trios with SNPs and sex chromosomes removed (total CpGs = 427,092) to compare differences in genome wide methylation. MDS (Figure 4.5) showed a progressive relationship between EK, NK, NR and WT but IPS and ESCs clustered far away from this trend. Unsupervised clustering analysis with the top 1% most variable probes also separated IPS and ESC tissues from EK, NK, NR and WT (data not shown) therefore WTs do not show global methylation profiles resembling pluripotent cells.

Although global methylation profiles did not show any association between WT and pluripotent tissues, the sites of methylation associated with pluripotency may have been masked by more variable global changes. Embryonic stem cells have significant cytosine methylation at CpA and some CpT sites (Ramsahoye, Biniszkiewicz et al. 2000) and an increase in methylation at these sites in WT similar to EK would imply stem cell likeness. The 450k array platform is designed to interrogate 3,091 non-CpG sites within the genome, however after pre-processing and normalisation, only 1,845 remained. Therefore, these probes were separated and MDS and hierarchical clustering was performed to assess the relationship between the trios, embryonic and pluripotent tissue. MDS clustered EK with WT (Figure 4.6), which had not occurred previously upon unsupervised clustering of the most variable CpGs globally (chapter 3.5). Hierarchical clustering formed two groups, one containing the pluripotent samples (ESC and IPS) and one containing all other samples (data not shown). Overall, average β-values for each tissue for all non-CpG probes were EK=0.07, ESC=0.42, IPS=0.46, NK=0.14, NR=0.14 and WT=0.13 showing that pluripotent tissue had higher levels of non-CpG methylation as expected. NK, NR and WT showed similar levels while EK showed lower levels. This data indicates that although WT and EK are more similar in terms of non-CpG methylation, these tissues are very distinct from pluripotent samples and do not show significant non-CpG methylation.

Next, genome-wide hydroxymethylation (H-me) was quantified in 6 trios using the Global DNA Hydroxymethylation ELISA Kit (Cell Biolabs). The levels of H-me did not correlate with histology and were not increased in WT compared to NK or NR (Figure 4.7; p=0.68 and p=0.47 for WT compared to NR and NK respectively, 2-tailed T-test). Unfortunately, as no ES or FK tissue was available for this analysis,

comparative levels of H-me methylation in these tissues could not be obtained. Therefore, it can only be concluded that levels of hydroxymethylation in WT are not significantly higher than in NK or NR.

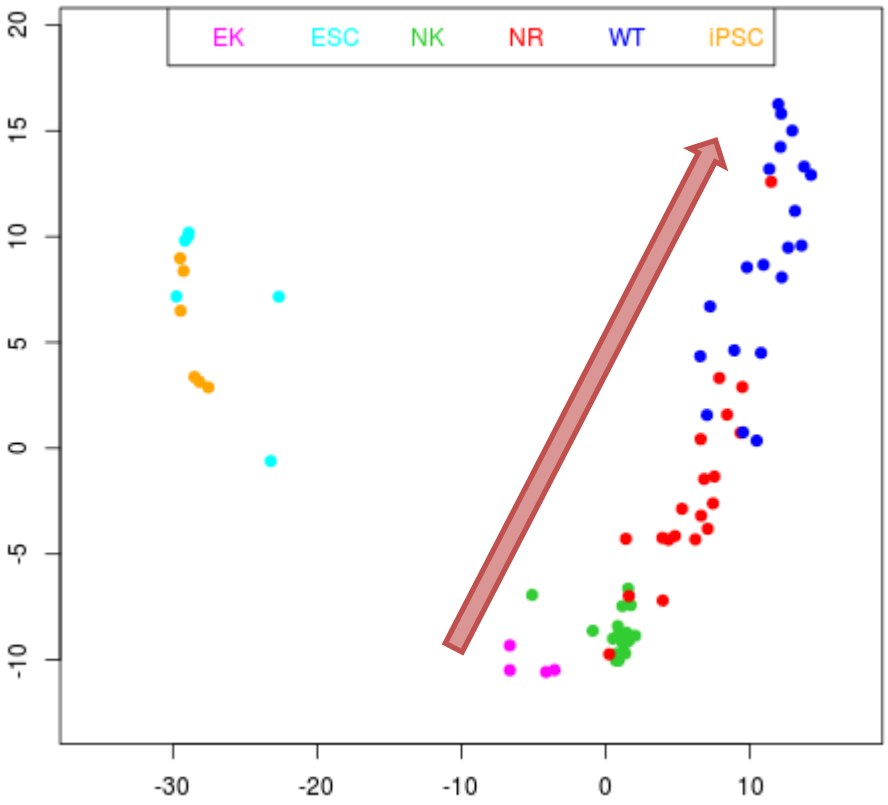


Figure 4.5: Multidimensional scaling to assess potency associated phenotypes. Multidimensional scaling of publically available datasets of embryonic kidney (EK; pink), induced pluripotent stem cells (IPS; orange), embryonic stem cells (ESC; cyan) with generated datasets of normal kidney (NK; green), nephrogenic rest (NR; red) and Wilms tumour (WT; blue). There seems to be a directional association between EK, NK, NR and WT, shown by the red arrow. Conversely, IPS and ESC samples cluster far away from these samples suggesting they are very different to the kidney associated populations.

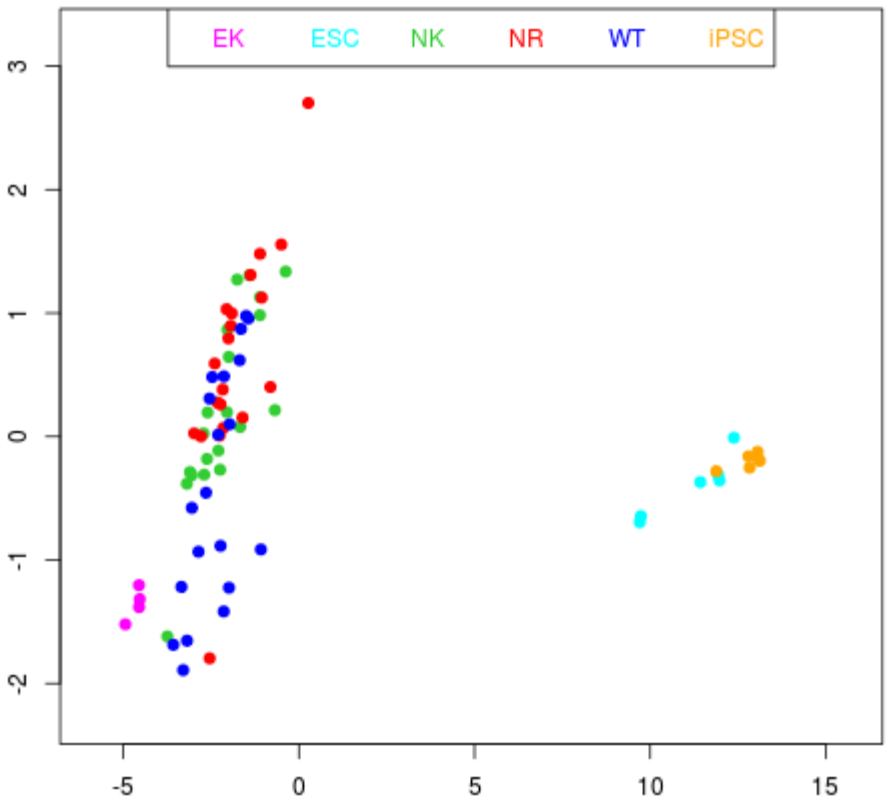


Figure 4.6: MDS of all non-CpG probes on the 450k retained after pre-processing (n=1,845). While embryonic stem cells (ESC; cyan) and induced pluripotent stem cells (IPS; orange) cluster far away from the more differentiated tissues, several Wilms tumours (WT; blue) cluster closely with embryonic kidney (EK; pink) while the majority of samples, including normal kidney (NK; green) and nephrogenic rests (NR; red), cluster close by.

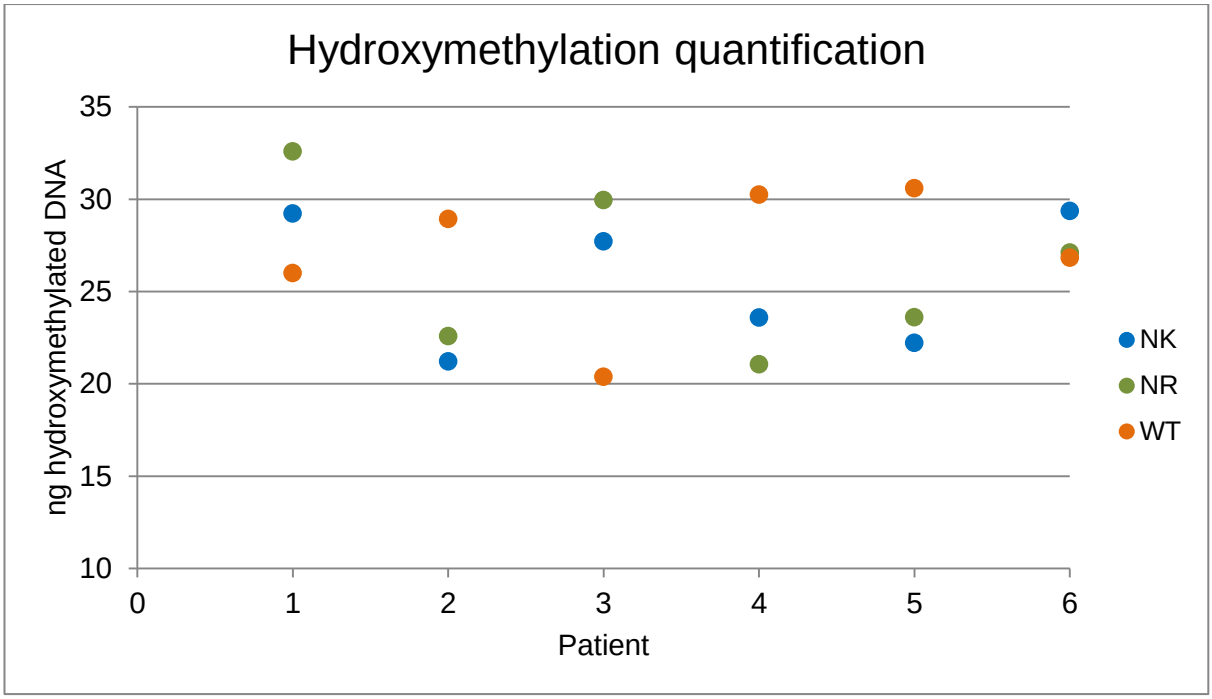


Figure 4.7: Quantification of hydroxymethylation in normal kidney, nephrogenic rest and Wilms tumours. The total nanograms (ng) of hydroxymethylated DNA was quantified using ELISA and showed varying levels across all samples of NK (blue), NR (green) and WT (orange).

The parameters tested here indicate that both NRs and WTs do not show characteristics of embryonic stem cells. Non-CpG methylation and H-me levels were low across all samples and CpG methylation did not cluster NRs or WTs with ES or IPS cell lines. Confounding factors in this analysis include the use of cell lines instead of primary tissue and batch effects from the experiments being run at different times in different locations/machines/chips.

4.6 WTs silence tumour suppressor genes by hypermethylation

After characterising general properties of group-1 WTs, the hyper-WT-DMRs were specifically studied to identify any methylation changes that could be readily associated with the transformation step from NRs, i.e. silencing of tumour suppressor genes. There were 123 genes associated with the 165 hyper-WT-DMRs which were cross-compared to the list of tumour suppressor genes held within TSgene, the Tumour Suppressor Gene Database (http://bioinfo.mc.vanderbilt.edu/TSGene/index.html). In total, 5 genes matched (Table 4.4) and due to the relationship between hypermethylation and gene silencing, these were considered likely targets for inactivation in group-1 WT. Indeed, RNA sequencing showed downregulation of *CASP8, RB1* and *TSPAN32* in WT compared to NR (Table 4.4). Due to low sample numbers, these differences did not reach statistical significance. As *MIR-195* and *H19* are a miRNA and ncRNA respectively, these were not detected by this assay.

Of these genes, *CASP8* and *H19* have been previously associated with WT (Okamoto, Morison et al. 1997; Morris, Hesson et al. 2003), whereas *RB1, MIR-195* and *TSPAN32* aberrations have not previously been identified in WT, although they have been detected in other cancers (Simpson, Hibberts et al. 2000; Nakamura, Yonekawa et al. 2001; Gonzalez-Gomez, Bello et al. 2003; Akino, Toyota et al. 2007; Kim, Chang et al. 2008; Kim, Lebron et al. 2008; Malekzadeh, Sobti et al. 2009; Lakomy, Sana et al. 2011; Li, Zhao et al. 2011).

Gene	DMR P-value	Hypermethylated WT	Average reads in NR	Average reads in WT
CASP8	0.0037	10 OF 13	104	20
H19	0.0045	11 OF 13	NA	NA
MIR195	0.0049	13 OF 13	NA	NA
RB1	0.0020	13 OF 13	55	11
TSPAN32	0.0091	10 OF 13	18	8

Table 4.4: tumour suppressor genes hypermethylated in group-1 Wilms tumours

4.7 Comparison of NRs to healthy tissue suggests that aberrant hypermethylation in NR tissue is associated with developmental arrest

After successfully identifying changes in methylation with biological significance in WTs, the same approach was applied to NRs. As there was no evidence that differences existed between NRs as 18/20 fell into the same cluster after the ANOVA analysis (Figure 4.3), despite there being two types of NR present (ILNR and PLNR), all NRs were considered equal.

To identify methylation changes associated with aberrant and incomplete renal development, comparative analysis was performed between the 20 NK-NR matched pairs as the NK represents the end-point of normal renal development whereas the NR represents aberrant disruption of this process. Linear modelling using Limma (Smyth 2005) identified 23,667 significant differentially methylated CpGs

(FDR<0.01) between the two groups. The champ.lasso algorithm was applied to these generating 629 DMRs of which 55% were hypermethylated and 45% were hypomethylated. These were termed kidney-rest DMRs (KR-DMRs with hypo-KR-DMR and hyper-KR-DMR referring to the methylated state in the NR with respect to NK), and are summarised in Table 4.5.

Again, these loci were interrogated using GREAT however analysis of hypo-KR-DMRs did not identify any biological processes that could be readily associated with developmental arrest (Table 4.6). Conversely, analysis of hyper-KR-DMRs showed that these DMRs fall into regulatory regions of genes involved in developmental or multicellular organismal processes including early embryonic patterning and embryonic organ morphogenesis (Table 4.7).

From this analysis, it appeared that gain of methylation in NRs was occurring at regions involved in development, which were not methylated in NK; hence it was postulated that this gain of methylation may be occurring at developmental loci required for nephrogenesis to complete properly. To test this hypothesis, the hyper-KR-DMRs were compared to regions known to be involved in development that show bivalent chromatin structure in embryonic stem cells (ESCs). Bivalent domains in ESCs are regions that show both the active chromatin mark H3K4me3 as well as the repressive chromatin mark H3K27me3 and are therefore held in a transient state. Depending on the signals received by the ESCs, these loci can be rapidly resolved to either silence or activate the underlying gene as required for cellular differentiation and are therefore critical for cellular development and specificity. Bivalent domains in ESCs were extracted from the literature (Bernstein, Mikkelsen et al. 2006; Pan, Tian et al. 2007) and multisampling analysis was performed to identify the

proportion of enrichment of hyper-KR-DMRs compared to what would be expected by chance. This analysis revealed a strong, significant enrichment of hyper-KR-DMRs within bivalent domains (10.8%, empirical p-value = 0.01, Figure 4.8; by comparison, hypomethylated KR-DMRs were negatively enriched -1.9%).

As bivalent domains mark key genes involved in development, this positive enrichment suggests that aberrant DNA hypermethylation, either deposited to these regions actively or passively through failure to resolve the bivalent structures, may contribute to the developmental arrest seen in NRs.

There were 35 genes that showed significant changes in expression between NK and NR by RNA sequencing of the four trios (FDR<0.05) and these did not show enrichment for developmental processes and none overlapped with the genes associated with hyper-KR-DMRs, however 2/35 of these genes (*AHSG* and *COBLL1*) have previously been implicated in mesoderm and ectoderm development. A larger study of gene expression in NK and NR is required to address this question.

	Hyper-methylated DMRs	Hypo-methylated DMRs
Median size (bp)	646	1513.5
Mean size (bp)	1418.5	1701.7
Range (bp)	104-8,088	154-5,807
Total (bp)	487,962	483,284
% enrichment: CpG island	-11.5	-21.9
% enrichment: CpG shore	9.9	2.5
% enrichment: CpG shelf	-1.0	4.1
% enrichment: 'other'	2.5	15.3
% enrichment: ESCs bivalent domain	10.8	-1.9
% enrichment: ESCs H3K4me3 only	-3.0	-15.1
% enrichment: ESCs H3K27me3 only	3.6	1.7
% enrichment: ESCs no H3K27 or H3K4-me3	-11.4	15.3
% enrichment: gene body	-1.3	16.5
% enrichment: TSS200	-2.1	-5.6
% enrichment: TSS1500	-3.3	-6.2
% enrichment: 3' UTR	1.0	-0.3
% enrichment: 5' UTR	0.0	-0.8
% enrichment: 1st exon	-2.0	-2.4
% enrichment: intergenic region	7.7	-1.2

Table 4.5: Description of hypermethylated and hypomethylated differentially methylated regions (DMRs) in nephrogenic rests compared to matched normal kidney Genetic features and those relating to CpG islands , shores (within 2kb of islands) and shelves (2-4kb from CpG islands) are extracted from the Illumina 450k annotation. Chromatin data (bivalent, H3K4me3 and H3K27me3 domains) refer to embryonic stem cells (ESCs) (Pan, Tian et al. 2007).

Enriched Term in GO biological processes	Hypergeometric fold enrichment	Raw p-value
Actin cytoskeleton organisation	2.43	1.31E-13
Actin filamentbased process	2.26	5.07E-12
Chromosome organisation involved in meiosis	10.76	1.33E-18
Collagen fibril organisation	6.17	3.30E-19
Extracellular matrix organisation	2.28	2.23E-07
Fat cell differentiation	4.94	4.72E-24
Hormone biosynthetic process	4.43	1.72E-13
Hormone metabolic process	2.52	7.31E-08
Induction of apoptosis	2.24	1.70E-13
Induction of programmed cell death	2.22	2.56E-13
Meiosis	2.75	3.61E-07
Meiosis I	4.43	6.91E-11
Meiotic cell cycle	2.71	4.72E-07
Negative regulation of transforming growth factor beta receptor signalling pathway	6.38	2.22E-23
Negative regulation of transmembrane receptor protein serine/threonine kinase signalling pathway	4.47	3.18E-17
Positive regulation of GTPase activity	2.48	2.46E-15
Positive regulation of Ras GTPase activity	3.31	1.89E-16
Regulation of GTP catabolic process	2.28	1.75E-14
Regulation of GTPase activity	2.25	1.39E-13
Regulation of myeloid cell differentiation	3.14	4.19E-12
Regulation of myeloid leukocyte differentiation	5.43	2.51E-21
Regulation of Ras GTPase activity	2.63	5.98E-13
Regulation of Ras protein signal transduction	2.14	1.12E-11
Regulation of transforming growth factor beta receptor signalling pathway	3.99	2.28E-15
Regulation of transmembrane receptor protein serine/threonine kinase signalling pathway	2.7	1.29E-11
Synapsis	11.04	7.12E-19
Synaptonemal complex assembly	16.25	5.36E-23
Synaptonemal complex organisation	13.37	6.78E-21
White fat cell differentiation	13.41	1.05E-42

Table 4.6: Significantly overrepresented biological processes identified by hypomethylated KR-DMRs

Enriched Term in GO biological processes	Hypergeometric fold enrichment	Raw p-value
Amine transport	2.99	9.88E-15
Anterior/posterior axis specification	2.5	1.16E-04
Anterior/posterior axis specification, embryo	5.39	4.04E-10
Anterior/posterior pattern specification	2.13	9.87E-11
Axis specification	2.24	2.78E-06
Blastoderm segmentation	4.98	1.36E-10
Cell-cell adhesion	2.08	2.92E-11
Cell-cell signalling involved in cell fate commitment	3.73	1.16E-08
Developmental induction	4.05	2.12E-09
Embryonic axis specification	3.27	1.28E-06
Embryonic organ morphogenesis	2.05	1.18E-10
Embryonic pattern specification	2.17	1.29E-04
Homophilic cell adhesion	4.91	1.09E-29
Intracellular receptor-mediated signalling pathway	3.57	7.02E-25
Negative regulation of kinase activity	3.39	7.73E-23
Negative regulation of organelle organisation	2.94	1.40E-15
Negative regulation of protein kinase activity	3.75	9.52E-26
Negative regulation of protein modification process	2.52	3.46E-11
Negative regulation of transferase activity	3.35	6.67E-24
Nitrogen compound transport	3.22	2.34E-19
Peptidyl-proline modification	4.17	5.81E-10
Protein peptidyl-prolyl isomerisation	4.6	6.36E-11
Regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	2.02	3.71E-03
Regulation of T-helper 1 type immune response	7.62	2.07E-11
Transition metal ion transport	3.55	9.90E-10

Table 4.7: Significantly overrepresented biological processes identified by hypermethylated KR-DMRs

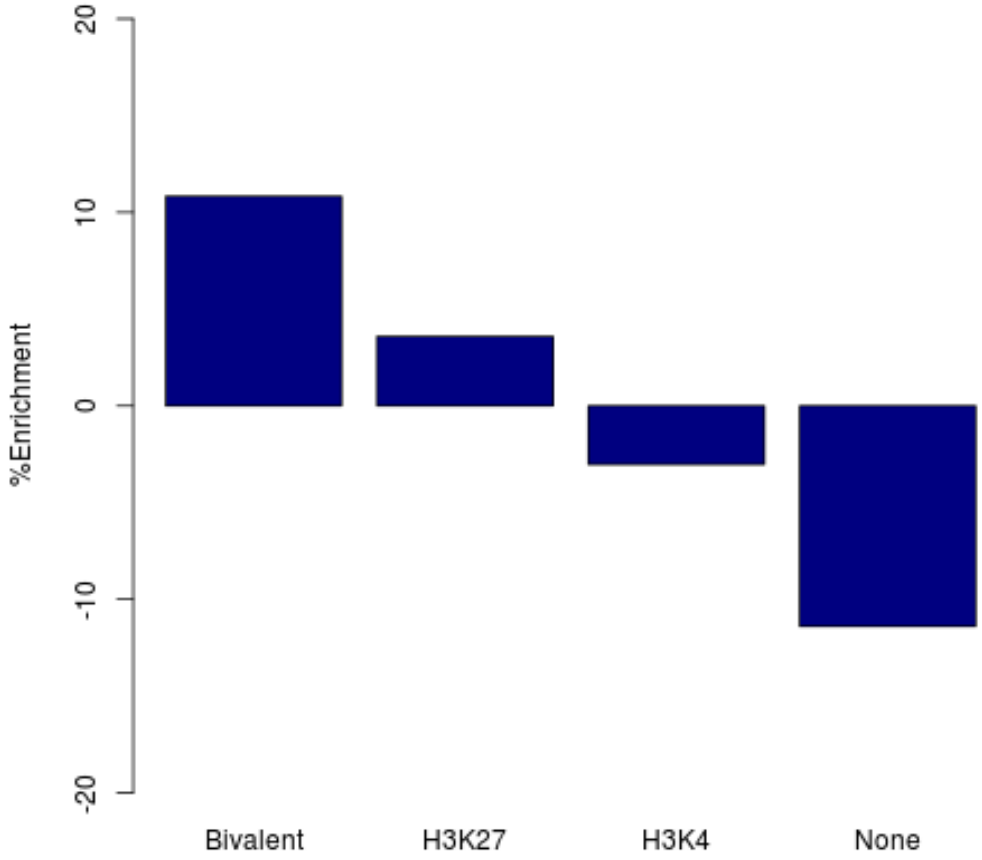


Figure 4.8: Hypermethylated KR-DMRs are enriched in developmental loci. Hypermethylated KR-DMRs showed 10.8% enrichment (empirical $p \leq 0.01$) for location within domains that are bivalent in embryonic stem cells (considered as developmental loci) compared to levels ascertained by repeated multiple sampling of all array CpGs.

4.8 Key genes in β-catenin signalling are hypermethylated in nephrogenic rests

Next, the Panther pathway analysis tool (http://www.pantherdb.org/) was used to highlight any pathways that were commonly targeted by the hyper-KR-DMRs. Three significant pathways were identified (Figure 4.9A) and of these, both the Wnt and Cadherin pathways had overlapping genes (14 in total, Table 4.8) which have previously been associated with β-catenin localisation and expression. As *CTNNB1* is a common mutation in WT (Li, Kim et al. 2004), these data suggest the pathway may be hit by other mechanisms including, but not limited to, long range epigenetic silencing involving the protocadherin cluster.

To understand the role of methylation of these genes in the progression of WT, the same regions that were significantly associated with the Wnt and Cadherin signalling pathways were studied across the full dataset. This analysis showed that WTs had an even higher average β-value for these genes than NR (which was higher than NK; Figure 4.9B). To validate this, Panther pathway analysis was also performed on hyper-WT-DMRs which indeed also identified the Wnt and Cadherin signalling pathways as significant (Wnt: $p=8.64x10^{-39}$; Cadherin: $p=5.20x10^{-69}$).

To associate these methylation changes with changes in gene expression, the RNA sequencing data was studied to assess levels of expression of these 18 genes in NK, NR and WT. Of the genes listed, none showed significant levels of differential expression between NK, NR and WT.

This evidence suggests sequential increase of dysregulation of Wnt and protocadherin signalling pathway genes in progression towards WT; however this was not validated in the small set of samples that underwent RNA sequencing. Previous evidence showed that genetic (*CTNNB1* mutation) and epigenetic (*PCDH* gene cluster hypermethylation) events resulted in increased activity of β-catenin in WT (Maiti, Alam et al. 2000; Fukuzawa, Heathcott et al. 2007; Major, Camp et al. 2007; Dallosso, Hancock et al. 2009) however this novel evidence indicates that β-catenin activation may occur as an early event in WT tumourigenesis. This activation may in fact be required for cell renewal capacity rather than cellular transformation and proliferation as it is in WTs (Miki, Yasuda et al. 2011; Merrill 2012).

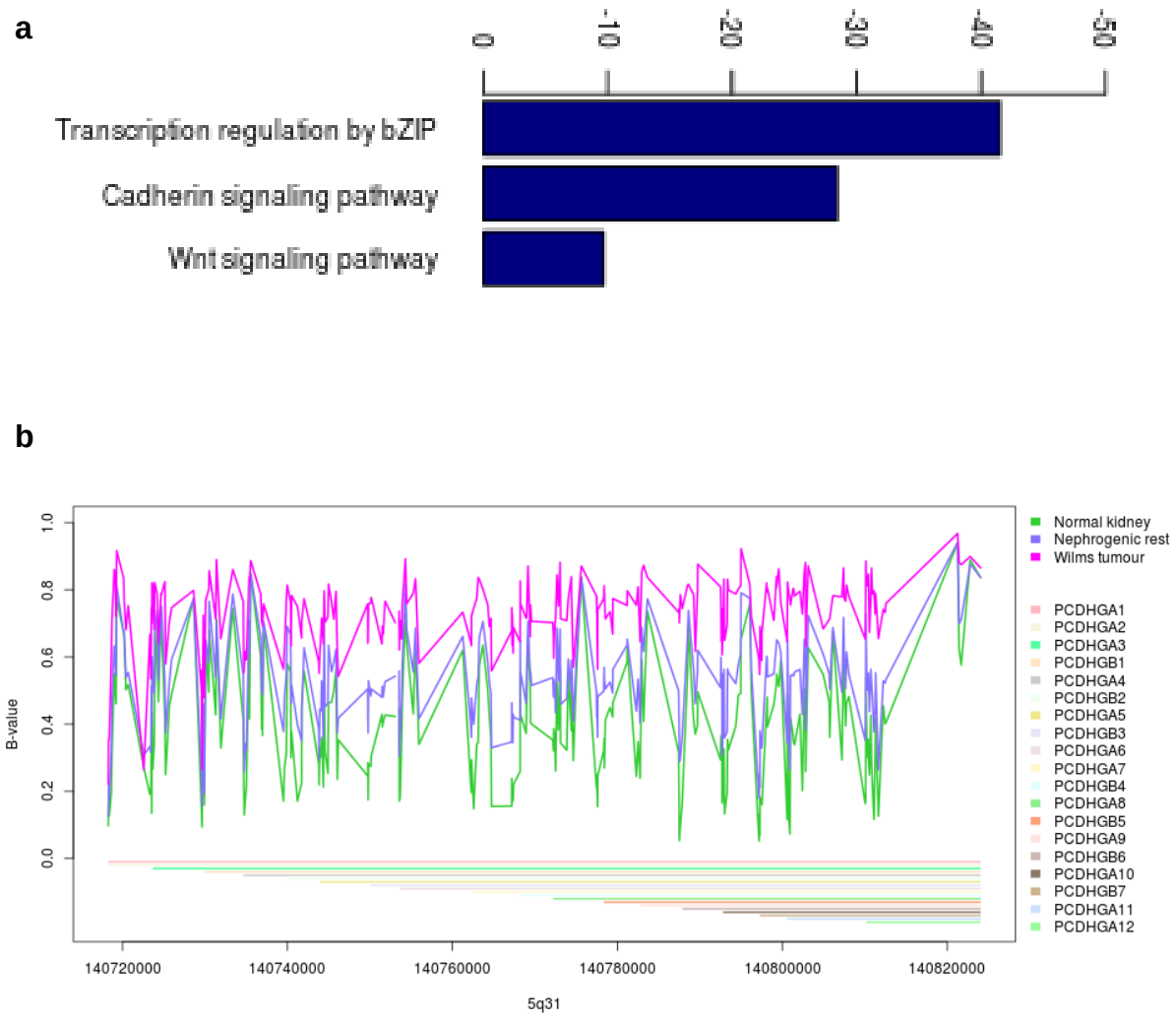


Figure 4.9: Hypermethylated KR-DMRs are enriched genes involved in β-catenin localisation. (**a**) Significant signalling pathways implicated by aberrant hypermethylation in NR compared to NK tissue identified using Panther with the log p-value for each pathway. (**b**) Protocadherin cluster methylation status in NK (green), NR (blue) and WT (pink) shows gain of methylation associating with tumourigenesis. Numbers refer to genomic coordinates from genomic built NCBI37.2 (hg19).

Gene	CpGs
ACTC1	4
CDHR2	4
EN1	3
FAT1	3
FZD10	3
FZD4	3
FZD5	6
PCDHAC2	7
PCDHB4	3
PCDHGC5	54
PLCB4	2
SRCAP	4
WNT1	1
WNT10B	4

Table 4.8: Genes with hypermethylated CpGs in nephrogenic rest tissue compared to normal kidney involved in Wnt and/or cadherin signalling pathways. The "CpGs" column lists the number of CpGs differentially methylated per gene.

4.9 Summary and conclusions

In this chapter, regional differences in DNA methylation were identified that were able to discriminate between NK, NR and WT. These changes showed biological significance with respect to WT progression including that NR formation is associated with gain of methylation at developmental loci and that two WT groups exist that differ in terms of their relationship with their respective NR.

No differences were seen in clustering that distinguished ILNRs from PLNRs however there were very few numbers of ILNR (n=4) within a total cohort of 20 rendering any comparative analysis challenging. Therefore, no additional analyses were carried out to identify rest-specific methylation changes as a much larger cohort would be required.

This chapter presented novel evidence that the relationship between a WT and its respective NR is not consistent. Group-1 WTs had very distinct methylomes from their respective NR whereas group-2 WTs had similar methylation profiles that showed no significant sites of differential methylation. Transformation in group-2 WTs, therefore, may be driven by somatic mutation leading to WTs with the same general epigenomic properties of the NR with expansion by proliferation. Further analysis of a panel of WT associated genes is planned to answer this question. Group-1 WTs significantly associated with bilateral disease. The insult that lead to tumour formation in patients with bilateral disease likely occurred at an earlier time point than patients with unilateral WT as both kidneys are affected. In these cases, the cell-of-origin is likely to be more primitive and therefore retain more epigenetic plasticity than a more differentiated cell which would allow for the widespread remodelling observed here. Conversely, the separation of WT groups may in fact

result as a confounder of cell type composition once again. Although, histologically, there was no relationship between cell type proportions and grouping, group-1 tumours may have an underlying shared composition that is not detectable by current histopathological review methods (i,e, may be composed of more immature elements than group-2 WTs).

To conclude, this chapter has presented a wide range of evidence of significant methylation changes in progression from NR to WT. These analyses were not corrected for cell type composition effects however they represent the true heterogeneity of the samples included in this study and therefore represent true biomarkers of the progression of this childhood malignancy. Several of the results were validated by RNA sequencing; however a much larger cohort is required to generate sufficient data for effective comparative analyses.

As these genome-wide and gene-specific assays work well on fresh frozen as well as FFPE tissue, there is potential for using these in a clinical setting to aid differentiation between NR and WT for patients with bilateral WT. This is essential to ensure that a resection margin is clear of tumour and also to aid treatment planning as NR can be left in a patient with no additional treatment however residual WT requires additional chemotherapy. Furthermore, as nephron-sparing surgery is critical for patients with bilateral disease, having the ability to differentiate between a NR and WT prior to surgery would enable efficient surgical planning so that less tissue is removed and the child can retain maximal renal function. Finally, as group-1 WTs showed increased epigenetic instability, these patients may be considered for epigenetic modifying therapy where few targeted therapies have been identified by

mutational analysis. The data described in this chapter has been submitted to Genome Medicine as an article.

Chapter 5: Analysis with cell type correction

5.1 Introduction

As mentioned throughout chapters three and four, cell type composition is a major confounder of methylation studies (also reviewed in (Jaffe and Irizarry 2014)). Recently, several algorithms have been designed to compensate for mixed cell type composition effects that may confound methylation studies. These include correcting for mixed blood leukocyte populations between cases and controls using known methylation profiles obtained from profiling cell-sorted blood fractions (Adalsteinsson, Gudnason et al. 2012; Houseman, Accomando et al. 2012), and two case-control comparison algorithms that require no reference dataset EWASher (Zou, Lippert et al. 2014) and RefFreeEWAS (Houseman, Molitor et al. 2014). EWASher is implemented through Python while RefFreeEWAS is implemented through R. Both reference-free algorithms use linear modelling for their comparisons and correct for cell type composition effects using estimated cell proportions generated by principal component analysis. The algorithms deconvolute the methylation signal generating p-values and covariates that represent direct epigenetic effects.

In this chapter, similar comparisons performed in chapter four are carried out using the RefFreeEWAS algorithm and the analyses are taken further by also comparing tissues to embryonic kidney (EK).

5.2 Between-tissue comparisons show that embryonic kidney is more similar to NK than to NR or WT.

To characterise tissue-specific methylation changes for EK (n = 4), NK, NR and WT (n = 20 matched trios), pre-processing and normalisation steps were repeated, as described in chapter 3.2 using the Dasen algorithm, and probes that bind to multiple loci were removed (Price, Cotton et al. 2013) leaving a matrix of 330,731 CpGs.

For between-tissue analyses, linear modelling was performed using Bioconductor package Limma with correction for cell type composition using the RefFreeEWAS algorithm (Houseman, Molitor et al. 2014). Using this method, a total of five comparisons were made between tissue groups and the significant methylation variable positions (MVPs; defined here as CpGs with $<|0.2|$ change in β-value between groups with $p<0.01$) were identified (Table 5.1). MVPs were focussed on throughout this analysis as the RefFreeEWAS package does not include a DMR-finder algorithm. Although between-array normalisation was carried out, comparisons between matched NK, NR and WT identified many less MVPs than comparison of each tissue to EK (Table 5.1), likely due to the patient specific covariate design used in matched analyses. Based on the number of significant CpGs from each analysis, we observed that NK and EK were more similar than NR and EK, and that WT was even more different with 8-fold more MVPs. These results mirror the unsupervised analyses discussed in chapter 3.5.

Group 1	Group 2	Group1-Hypo-MVPs	Group1-Hyper-MVPs
NK	NR	42 (35%)	77 (65%)
NR	WT	223 (82%)	48 (18%)
NK	EK	817 (49%)	851 (51%)
NR	EK	2108 (48%)	2349 (52%)
WT	EK	5814 (44%)	7538 (56%)

Table 5.1: Methylation variable positions identified by between group comparisons

5.3 NRs are hypermethylated at renal development loci compared to EK

To begin with, as NRs morphologically resemble EK, the MVPs that differentiate between NRs and EK were studied. Of the 4,457 MVPs, 2,108 (47%) were hypo-MVPs and 2,349 (53%) were hyper-MVPs in NR with respect to EK. Although hypo-MVPs did not allow immediate identification of genes that could be readily associated with nephrogenic disruption, many of the hyper-MVPs fell within key genes involved in renal development including *BMP4*, *BMP7*, *GATA3*, *GDNF*, *FOXD1*, *FOXC1*, *IRX3*, *LHX1*, *OSR1*, *PAX2*, *PAX8*, *SALL1*, *SIX1*, *SIX2*, *SOX4*, *SOX9*, *WNT4*, *WNT7B* and *WT1*. Predictably, interrogating hyper-MVPs using GREAT (http://bejerano.stanford.edu/great/public/html/), identified enrichment of processes involved in nephron morphogenesis, nephron epithelial morphogenesis, mesonephros and metanephros development as well as other kidney-related and general development-related gene ontology processes (Table 5.2).

GO biological processes	hypergeometric fold enrichment	p-value
Anterior/posterior pattern specification	2.78	9.26E-39
Cartilage development	2.84	1.17E-25
Ear development	2.93	7.17E-38
Ear morphogenesis	3.53	1.54E-37
Embryonic morphogenesis	2.35	5.30E-56
Embryonic organ development	2.33	3.56E-38
Embryonic organ morphogenesis	2.94	5.54E-50
Kidney epithelium development	4.03	9.95E-26
Kidney morphogenesis	4.74	8.66E-31
Mesonephros development	6.81	2.04E-30
Metanephros development	3.52	8.96E-30
Nephron epithelium morphogenesis	5.2	4.39E-30
Nephron morphogenesis	5.38	2.13E-33
Organ morphogenesis	2.03	1.20E-53
Pattern specification process	2.43	2.41E-56
Regionalisation	2.8	1.31E-59
Renal vesicle development	6.35	1.31E-25
Sensory organ development	2.23	2.69E-40
Skeletal system development	2.34	1.60E-40
Tube development	2	1.26E-28

Table 5.2: Biological processes associated with MVPs hypermethylated in NR compared to NK

5.4 The PRC2 complex is implicated in developmental arrest by aberrant hypermethylation in NRs

Analysis of hyper-MVPs with GREAT also identified a very significant association between CpGs hypermethylated in NRs and regions identified as Polycomb Repression Complex 2 (PRC2) targets in ESCs ($p=2.79x10^{-66}$) (Ben-Porath, Thomson et al. 2008). To assess whether this association was detected due to NR cells progressing down a normal differentiation lineage past the embryonic stages captured in this experiment, i.e. acquiring NK-like properties, the MVPs identified by comparison of EK tissue to NK were studied in further detail. In total, 851 hyper-MVPs were identified in NK with respect to EK however no association between hyper-MVPs and PRC2 regions was observed confirming that the aberrant hypermethylation seen is branched away from normal renal development. In general, PRC targets include developmental and lineage specific genes, and complexes have been shown to recruit DNA methyl-transferase enzymes resulting in aberrant cancer-associated deposition of methylation at genes required for cellular differentiation (Bracken, Dietrich et al. 2006; Vire, Brenner et al. 2006; Bracken and Helin 2009; Mohammad, Cai et al. 2009). The regions that were targeted by aberrant hypermethylation in NRs included a set of 189 genes and 480 CpGs (20% of hyper-MVPs). This evidence suggests a role for the PRC2 complex in nephrogenic developmental arrest and therefore NR formation.

5.5 WTs are expanded NRs with gain of methylation at general developmental loci associated with Polycomb targets

Next, MVPs identified from comparison of WT to EK were analysed, of which 5,814 (44%) were hypomethylated and 7,538 (56%) were hypermethylated in WT with respect to EK, a relatively even distribution. The primary focus was on WT hyper-MVPs which, by interrogation with GREAT, identified similar developmental processes as identified in the NR-EK comparison suggesting maintenance of the epigenetic landscape from the NR. The surprising difference was that the WT-EK comparison highlighted many more processes involved in general embryonic development instead of specifically renal development including, for example, 228 genes involved in embryonic pattern specification and 251 genes involved in embryonic morphogenesis (the most methylated included *FOXD1*, *GLI2*, *HOXA5*, *HOXD10*, *LBX1*, *PAX2*, *SIM2*, *SIX3*, *TBX3*, *UNCX*, *VAX2* and *WNT10A*). Furthermore, a significant enrichment was again seen for hyper-MVPs within regions of PRC2 binding ($p=3.92x10^{-217}$), however there was also a very significant enrichment for regions of H3K27me3 ($p=2.91x10^{-247}$), Polycomb EED targets ($p=1.08x10^{-241}$) and Suz12 targets ($p=8.65x10^{-207}$) all identified by ChIP on chip in human ESCs (Ben-Porath, Thomson et al. 2008). This evidence suggests a further dysregulation of methylation at Polycomb targets sites and developmental loci as cells progress towards malignancy.

Verifying the involvement of Polycomb family proteins in Wilms tumourigenesis, a previous chromatin profiling study of WTs identified a network of epigenetic regulators that had similar narrow H3K4me3 peaks and expected high expression in both WT and ESCs but not NK (Aiden, Rivera et al. 2010). This set of genes

included Polycomb group proteins *JARID2*, *FBXL10* and *CBX2* (Aiden, Rivera et al. 2010) associating their aberrant high expression with WT. Furthermore, a study of *in vivo* reprogramming in mice associated formation of WT-like lesions with failed reprogramming associated specifically with Polycomb group proteins failing to silence their targets (Ohnishi, Semi et al. 2014).

5.6 The Ras network is active in WTs

Next, hypo-MVPs in WT compared to EK were analysed using GREAT which identified over-enrichment for GO biological processes involved in RAC, RAS and RHO GTPase activity (Table 5.3). RAS, RHO and RAC are an interrelated family of proteins involved in cell growth, cell cycle, cytoskeleton reorganisation and activation of protein kinases. Their aberrant activation can lead to oncogenesis; therefore WTs are likely, in part, driven by an oncogenic network of Ras signalling molecules. These aberrantly hypomethylated genes included *ABR*, *ARHGEF16*, *NDEL1*, *SCRIB* and *RASA3*. In chapter 4, Ras signalling was associated with hypomethylated DMRs in both comparison between NK vs NR and NR vs WT. The finding that this relationship is not associated with cell type composition backs up previous work where a kidney-specific β-catenin activating mutation was shown to result in epithelial tumours in mice, however when introduced along-side K-RAS activating mutation, bilateral, metastatic, multifocal renal epithelial tumours were observed that resembled epithelial WT (Clark, Polosukhina et al. 2011). This suggests a synergistic role for these two proteins in WT tumourigenesis.

GO biological processes	Hypergeometric fold enrichment	p-value
Activation of Rac GTPase activity	6.78	1.09E-58
Activation of Ras GTPase activity	5.06	1.25E-49
Activation of Rho GTPase activity	5.77	1.40E-55
Brown fat cell differentiation	5.32	1.27E-51
Fat cell differentiation	2.71	7.32E-33
Negative regulation of granulocyte differentiation	8.04	1.48E-59
Negative regulation of myeloid leukocyte differentiation	4.04	3.45E-35
Negative regulation of transforming growth factor beta receptor signalling pathway	2.93	7.73E-26
Positive regulation of Cdc42 GTPase activity	7.16	3.37E-58
Positive regulation of establishment of protein localisation in plasma membrane	5.99	6.58E-50
Positive regulation of Rac GTPase activity	4.47	1.36E-44
Positive regulation of Ras GTPase activity	2.06	6.65E-22
Postitive regulation of Rho GTPase activity	2.94	2.30E-31
Regulation of Cdc42 GTPase activity	5.77	3.41E-49
Regulation of establishment of protein localisation in plasma membrane	5.43	8.34E-46
Regulation of generation of precursor metabolites and energy	2.83	1.35E-27
Regulation of granulocyte differentiation	6.5	1.55E-51
Regulation of Rho GTPase activity	2.47	3.83E-25
Tongue development	4.62	2.08E-45
White fat cell differentiation	5.93	2.06E-60

Table 5.3: Biological processes associated with MVPs hypomethylated in WT compared to EK

5.7 NR and WT are epigenetically remarkably similar

Considering the similarities observed in MVPs hypermethylated in NR and WT compared to EK, the MVPs identified in the NR vs WT comparison were studied next. Strikingly, there were very few MVPs: 223 (90%) hyper- and 24 (10%) hypo-MVPs in WT with respect to NR and when entered into GREAT to discern biological information, no significant signalling pathways were identified. This suggests the two tissues are remarkably similar, with major differences associated with cell-type composition effects. There were some key genes that were hypermethylated in WT compared to NR including *SIX2*, *SIM2*, *LY6G5C*, *ULK2* and *VAX2*. Many were general embryonic developmental genes, validating results of each tissue comparison to EK. There were 10 genes in the Protocadherin gene cluster that were methylated in WT with respect to NR, a pathway previously shown to be hypermethylated in WTs and discussed in chapter 4.

5.8 Comparison to EK reveals EK-specific and cancer-associated transcription factor families

Finally, the transcription factors enriched for binding within the MVPs of all three tissues compared to EK were examined to identify any common features. This analysis found that the CpGs significantly hypomethylated in EK compared to each of the other three tissue types significantly associated with HOXL binding sites in each comparison (NK: $p=6.86x10^{-36}$, 20 genes, NR: $p=7.29x10^{-33}$, 26 genes and WT: $p=2.25x10^{-77}$, 43 genes; total of 47 genes). Therefore, it can be predicted that homeobox family genes are expressed during normal renal development and are

silenced in the post natal normal kidney and are not associated with developmental arrest or malignancy.

Continuing this strategy, it was next observed that the regions hypermethylated in both NR and WT tissues compared to EK showed significant association with four gene families: PRD, SIX, FOX and PAX,PRD (Table 5.4). These gene families were not detected in comparative analysis between EK and NK suggesting these transcription factors are hypomethylated in both EK and NK, and aberrantly hypermethylated in Wilms tumourigenesis.

Furthermore, WTs also showed hypermethylation of WNT and FZD targets compared to EK that was not detected in comparison to NK or NR evidencing that β-catenin/Wnt signalling is aberrant in WTs. As activating *CTNNB1* mutations are frequently seen in WT, the hypermethylation (and predicted silencing) of β-catenin signalling proteins may indicate that activated β-catenin is signalling by a different pathway (i.e. the non-canonical Wnt pathway) or that due to the activation, expression of Wnt and Fzd proteins becomes redundant.

	NR	WT
FOX	$p=1.29x10^{-11}$	$p=2.90x10^{-14}$
	17 genes	25 genes
	59 CpGs	139 CpGs
PAX,PRD	$p=1.05x10^{-7}$	$p=1.21x10^{-7}$
	4 genes	6 genes
	20 CpGs	32 CpGs
PRD	$p=2.11x10^{-5}$	$p=6.58x10^{-23}$
	9 genes	24 genes
	30 CpGs	116 CpGs
SIX	$p=6.43x10^{-5}$	$p=6.18x10^{-21}$
	4 genes	5 genes
	11 CpGs	46 CpGs

Table 5.4: Significant transcription factor families with targets hypomethylated in NRs and WTs compared to EK. Details include p-value for significance, the number of genes targeted and the number of CpGs involved.

5.9 Summary and conclusions

To summarise, analysis of methylation levels in EK, NK, NR and WT has shown that NR retention is associated with aberrant hypermethylation of PCR2 targets at many loci involved in renal development. By comparison of WT to EK and

NR, evidence was shown that WTs are also hypermethylated at genes involved in development, not limited to renal development, which are associated with a wider range of Polycomb group protein binding sites. WTs also showed hypomethylation and predicted activation of genes involved in Ras signalling.

The results of these analyses, although similar to those in chapter 4, reveal novel markers of progression that were previously not detected and therefore masked by cell type composition effects. To make the most of methylation analysis, it is key to perform both corrected and non-corrected analyses to fully understand the methylation changes associated with the mixed tissue histology as well as the underlying changes associated with malignant progression.

As mentioned, a common feature of CpGs methylated in NR and WT tissues in comparison to EK tissue was location in Polycomb target regions. The normal role of Polycomb group proteins is to silence target genes during development, by altering the local chromatin structure, i.e. by depositing H3K27me3. Polycomb proteins make up Polycomb repressive complex 1 and 2 (PRC1 and PRC2) which are comprised of proteins with various functions including chromatin modification. In NR tissue the hypermethylated sites were predominantly at genes associated with renal development but in WT tissue were more general developmental genes. Polycomb target hypermethylation has previously been associated with the cancer phenotype and less well-differentiated tumours (Ben-Porath, Thomson et al. 2008). It has been proposed that the deregulation of normal Polycomb mechanisms is central to tumour initiation (Bracken and Helin 2009), therefore it is perhaps unsurprising that dysregulation at these sites is seen as an early event.

Supporting the role of Polycomb protein dysregulation in WTs, evidence from a mouse model of *in vivo* reprogramming associated formation of WT-like lesions with failure of Polycomb gene targets to be repressed (Ohnishi, Semi et al. 2014). Furthermore, upregulation of Polycomb genes *BMI-1*, *EZH2*, *SUZ12* and *EED* were seen in progressive blastemal-enriched WT xenografts in mice, suggesting their expression correlated with tumourigenesis (Metsuyanim, Pode-Shakked et al. 2008).

Upregulation of Polycomb genes has been hypothesised as a mechanism by which aberrant gain of methylation may occur at target loci (Bracken and Helin 2009). Polycomb group proteins EZH2 (Vire, Brenner et al. 2006) and CBX7 (Mohammad, Cai et al. 2009) can physically associate with DNMTs and upregulation of these genes would cause progressive recruitment of DNMTs to Polycomb targets switching temporary silencing to permanent silencing (Bracken and Helin 2009). Polycomb group targets are often found to accumulate methylation in cancer cells, which has been specifically observed at *WT1* and several other genes (Bracken, Dietrich et al. 2006). PRC targets are around 7-12 times more likely to be aberrantly methylated and silenced in cancer than non-PRC targets (Bracken and Helin 2009). Alternatively, Polycomb genes may be lead astray by the factors that are required for their association to target genes as Polycomb group proteins do not have the ability to bind to specific DNA motifs.

As Polycomb targets include lineage-specific genes, and many genes involved in renal development were associated with aberrant hypermethylation in NR and WT, it can be hypothesised that their methylation in Wilms tumourigenesis silences pro-differentiation genes so that cells fail to normally differentiate. The question that remains is what causes PRC overexpression in the first place? Genetic mutation

could be involved and exome sequencing projects are currently underway that may highlight novel mutations in WT associated with Polycomb gene regulation. However no full genome analysis has been published which is more likely to identify mutations in significant regulatory regions.

Analyses in this chapter showed that NR and WT appeared remarkably similar. The reason for this is likely that the main differences in methylation between the tissues are associated with cell type composition (including the DMRs identified in chapter four). They are still true biomarkers for Wilms tumourigenesis, but present due to varying cellular compositions. Conversely, the small number of differential CpGs may be due to the more stringent cut-off parameters set for these analyses ($p<0.01$ and $\Delta\beta>0.2$ for RefFreeEWAS comparisons and only $p<0.01$ for non-corrected comparisons).

Cell composition corrected analyses did identify some key genes that were hypermethylated in WT with respect to NR (*LY6G5C*, *VAX2*, *ULK2* and *SIX2*). Of these, *LY6G5C* is located in the major histocompatibility complex (MHC) cluster. Methylation of MHC cluster genes has previously been seen in cancer as a way of decreasing expression to avoid immunosurveillance (Jäger, Ringhoffer et al. 1997; Degenhardt, Huang et al. 2010; Garrido, Paco et al. 2012). This interesting mechanism will be discussed further in chapter seven. Activation of the Wnt signalling pathway occurs by activating *CTNNB1* mutation in ~15% WT. VAX2 is another protein associated with this pathway (a potent dominant-negative Wnt antagonist where loss of expression results in failure of VAX2 to prevent Wnt binding and therefore activation of Wnt signalling) which may be another mechanism of activating the same pathway. *ULK2* is a gene involved in autophagy,

which is a normal process of the kidney to avoid kidney damage by mTOR signalling however its methylation in WT compared to NR doesn't indicate an obvious association with tumourigenesis.

Gain of methylation at SIX2 has been previously shown in progressive blastemal-enriched tumour–propagating xenografts which correlated with loss of SIX2 expression which the authors associated with loss of renal differentiation and WT progression diverting the progenitor phenotype away from a renal pathway to a more malignant one (Metsuyanim, Pode-Shakked et al. 2008). However, in general, SIX2 is highly expressed in WTs (Murphy, Pierce et al. 2012). The RNA seq data showed an increase in SIX2 expression from 0 in all NRs to an average of 73 reads in two out of three WTs. Therefore, gain of methylation at *SIX2* may correlate with loss of expression and be a marker of more malignant WTs in only a subset of cases.

The final comparison of transcription factor enrichment in all tissues showed that HOXL proteins were hypomethylated in EK but not other tissues. The fact that HOXL targets were consistently significantly hypomethylated in EK tissue compared to all other tissues suggests that the master transcription factors active in EK tissue are HOX family genes which are not active in properly differentiated NK, aberrant NR or WT. In fact, Hox proteins are known to be essential for metanephric kidney induction (Wellik, Hawkes et al. 2002; Mugford, Sipila et al. 2008; Preger-Ben Noon, Barak et al. 2009).

Four TF families were identified that show hypermethylation in NR and WT; SIX, FOX, PRD and PAX. These transcription factor families are all involved in development. This data further validates that developmental loci required for normal development are hypermethylated in NR and WT.

Chapter 6: Identifying a tumour-specific biomarker

6.1 Introduction

In Europe, most children with a suspected WT commence pre-operative chemotherapy without histological diagnosis based on a presumptive clinical diagnosis. After four weeks of pre-operative chemotherapy, complete or partial nephrectomy provides tumour stage and histological subtype which together dictate the post-operative treatment (Lemerle, Voute et al. 1983). Although overall survival rates are good (>90% for localised disease), there is a clinical need for a blood or urine biomarker that could aid in diagnosis and evaluate patient response to chemotherapy and improve prediction of relapse. Circulating cell-free DNA (cfDNA) isolated from blood has been used to assess tumour burden in other cancers (Gautschi, Bigosch et al. 2004; Mori, O'Day et al. 2005; Ramirez, Rosell et al. 2005; Diehl, Schmidt et al. 2008) and due to the low frequency of genetic mutation in WT, the methylome, rather than the genome, seems more likely to reveal ubiquitous tumour-specific biomarkers. Therefore, the previously derived methylation data was interrogated to identify sites of tumour-specific differential methylation that were then assessed in cfDNA for use as WT biomarkers.

6.2 Identification of biomarker differentially methylated regions

To identify tumour-specific methylation variable positions (MVPs), as previously defined (Rakyan, Beyan et al. 2011), subset within quantile (SWAN) normalisation of β-values was performed on 22 matched pairs of NK and WT using

Bioconductor package ChAMP, leaving 462,537 CpGs. As previously discussed, this normalisation step was repeated to include the maximum number of CpGs for assessment as biomarkers as the analysis pipeline removes any CpGs with a detection p-value of <0.01 and therefore the overall number of CpGs decreases with an increase in the number of samples. A patient-specific paired linear model was performed using Limma, as previously described, however in this instance the algorithm TREAT was included which adjusts p-values according to the CpG-specific Δβ (McCarthy and Smyth 2009). This analysis identified 309 MVPs of genome-wide significance ($P<5x10^{-8}$; Figure 6.1). As this analysis was paired, the possibility of these being the result of genetic polymorphism, ethnicity or age is excluded. The DMR-finder algorithm champ.lasso implemented through the Bioconductor ChAMP package was then applied (Morris, Butcher et al. 2013) which grouped these MVPs into three discrete clusters (DMRs) of genome-wide significance (Table 6.1). All three of these DMRs were hypermethylated in WT with respect to NK.

The DMRs that were identified by this method were indeed biomarkers for WT; however they may have resulted from cell composition effects. For example, NK is composed of around 95% epithelia and 5% stroma cells whereas WTs are composed of varying proportions of blastema, epithelia and stroma. Because of this, the three DMRs were termed putative DMRs (pDMRs).

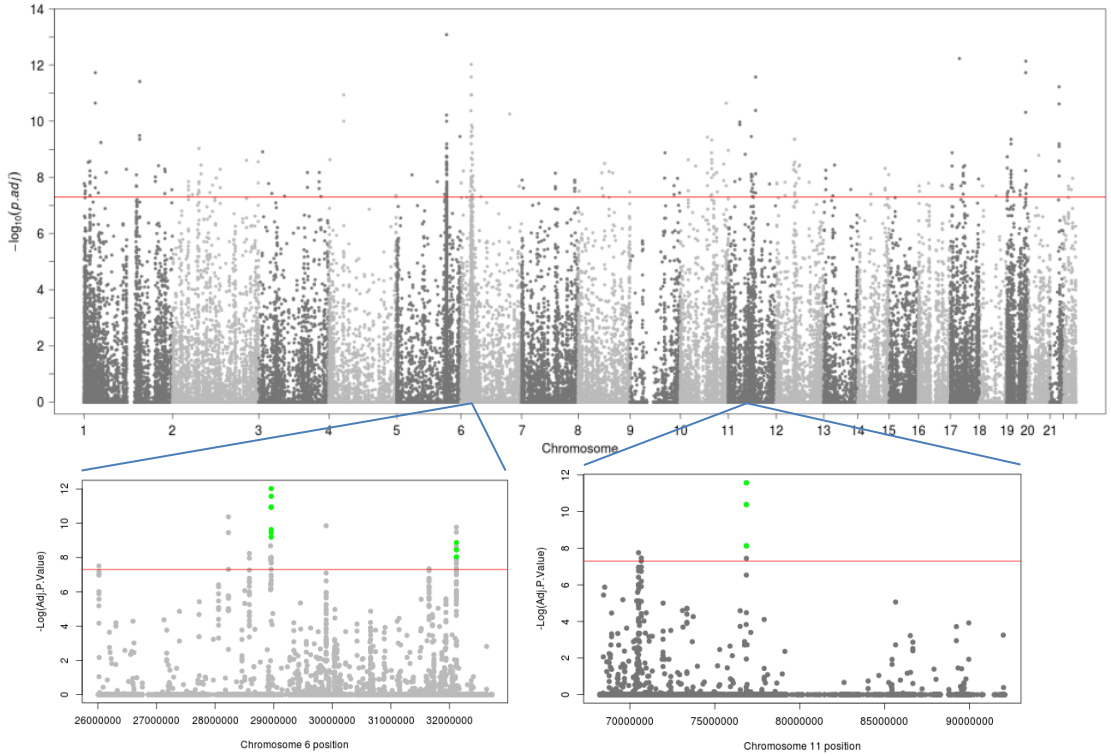


Figure 6.1: Identification of differentially methylated loci with genome-wide significance. Manhattan plot showing $-\log_{10}$ p-values for all CpGs (462,537) generated by comparison of normal kidney (NK) to Wilms tumor (WT). CpGs with genome-wide significance (n = 309) appear above the red line. DMRs 1 and 2 on chromosome 6 and DMR-3 on chromosome 11 are shown on separate plots below in green.

DMR	Location	First CpG	Last CpG	Size (bp)	No. CpGs	DMR p-value	CpG island	Overlapping gene
1	6p22.1	28956226	28956426	200	8	1.58E-10	CpG:42	none
2	6p21.32	32116905	32116963	58	3	4.67E-09	CpG:56	*PRRT1*
3	11q13.5	76858947	76859056	109	3	2.48E-09	CpG:38	*MYO7A*

Table 6.1: Differentially methylated regions hypermethylated in WT compared to NK

6.3 Cell type composition analysis

To assess possible confounding due to differential cell type composition, two further analyses were performed. To begin with, histological analysis was performed by a paediatric pathologist who determined the proportions of each cell type (stroma, epithelia, blastema, necrotic tissue or tissue showing chemotherapy-induced changes, to total 100%) in each sample according to the original H&E stained sections that were used as a guide for microdissection (see chapter 2.2). This confirmed that the samples were composed of the anticipated major cell types, in the proportions expected for NK and with varying proportions in WT samples (Figure 6.2). Next, a recently published algorithm that corrects p-values generated through linear modelling based on estimated cell type contributions; RefFreeEWAS (Houseman, Molitor et al. 2014), was performed. Linear modelling using the RefFreeEWAS algorithm identified 7,272 CpGs with genome-wide significance ($P<5x10^{-8}$) of which 937 had $\Delta\beta>0.3$ and were therefore considered MVPs with 82% being hypermethylated and 18% hypomethylated in WT with respect to NK. As the RefFreeEWAS package is yet to contain a DMR-finder algorithm, the MVPs were focussed on.

For the MVPs that were hypermethylated in WT there was a striking positive enrichment for location within CpG islands (+18% compared to the relative % of each feature in all the background CpGs n=462,537). In total, 483 CpG islands contained at least one MVP. Two CpG islands that showed the greatest number of CpGs were located on chr6 (CpG:56 and CpG:42) with 13 and 11 CpGs respectively which overlap with the pDMRs 1 and 2. Furthermore, by retrospective analysis of the MVPs, an overlap of three CpGs within pDMR3 was identified. The results of this analysis suggest that all three pDMRs are in fact true DMRs and are not mediated by cell composition effects.

To further confirm these results, it was clear from comparing the known cell composition of each sample to the respective methylation level across the three DMRs that methylation levels were high despite the sample composition (Figure 6.2), and in fact the proportion of any of the three cell types showed no correlation with DMR methylation status (Figure 6.3).

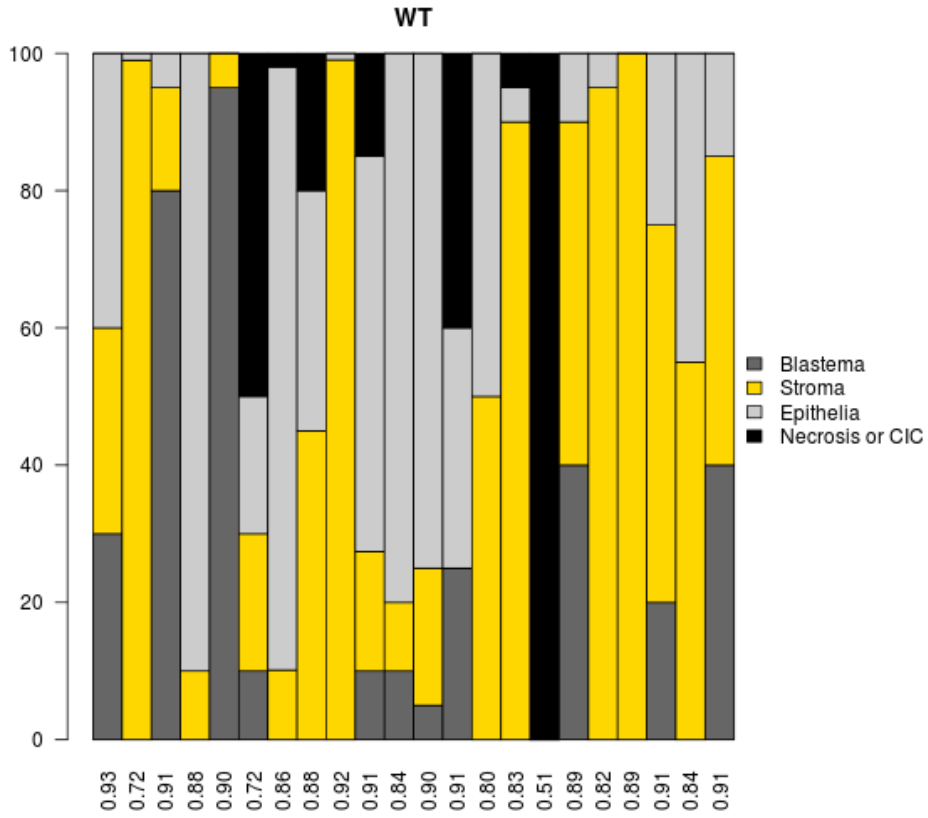


Figure 6.2: Quantification of cell proportions in each micro-dissected section used for DNA extraction. Each haematoxylin and eosin stained section was assessed by a paediatric pathologist who confirmed the proportion of blastema (dark grey), stroma (yellow), epithelia (light grey) or necrosis/chemotherapy induced changes (CIC; black) present in each sample that was micro-dissected. Samples are labelled by the respective average β-value across all DMR CpGs which shows that methylation does not correlate with proportion of any cell type.

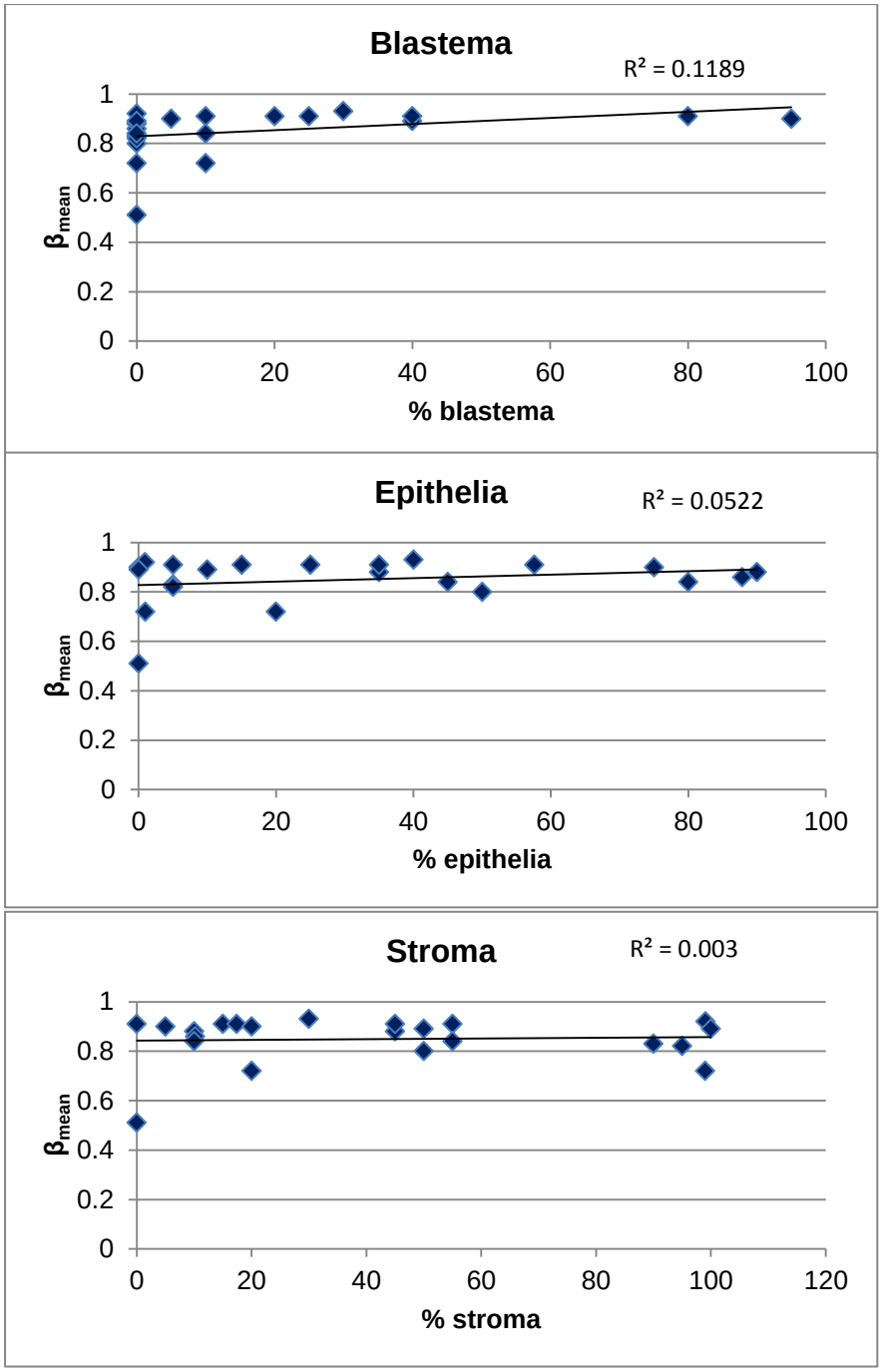


Figure 6.3: Correlative analysis of DMR methylation levels with cell type composition per sample. Correlative analysis between the β_{mean} values of methylation across all three DMRs and the proportion of any of the three main cell types did not show any significant correlation with R^2 values of 0.11, 0.05 and 0.03 for blastema, epithelia and stroma respectively.

6.4 Replication and validation of DMR hypermethylation

After confirming that the DMRs were not associated with cell type composition, the levels of methylation were explored in two independent datasets. The DNA methylation levels across the three DMRs were aggregated to derive β_{mean} for an independent replication dataset of 12 pairs of NK and WT as well as an independent set of 86 WTs encompassing the seven post-chemotherapy histological subtypes classified into two risk groups as defined by the International Society of Paediatric Oncology (SIOP) (Vujanic, Sandstedt et al. 2002).

Comparison of β_{mean} between the NK-WT pairs revealed that WTs were significantly hypermethylated compared to NKs, as expected ($p=9.26x10^{-10}$, 2-tailed T-test, Figure 6.4A). Furthermore, β_{mean} showed consistent high levels across the second independent WT dataset (Figure 6.4B) with a significant difference between WTs classified as high risk (average β_{mean} = 0.87) and intermediate risk (average β_{mean} = 0.78; p=0.0024, 2-tailed T-test). This data confirmed that hypermethylation of the three DMRs was not restricted to the discovery dataset. Next, based on methylation levels in the initial discovery dataset of n=22 pairs, a support vector machine algorithm was applied using R CRAN package e1071. This algorithm used the discovery dataset to differentiate NK from WT then applied the learning algorithm to the two independent cohorts to discern whether each sample was more similar to NK or to WT. In total, the algorithm correctly classified 100% of the replication dataset of 12 pairs and 98% of the 86 independent WT samples. These results showed that the DMRs were very discriminative and potentially useful as biomarkers. Furthermore, as a significant difference in methylation levels between intermediate and high risk WT was identified, this could be of diagnostic use for WT patients.

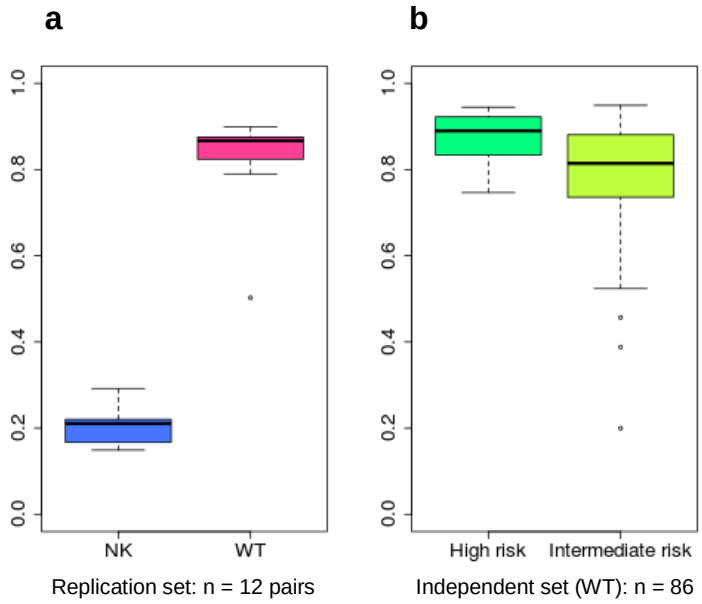


Figure 6.4: Methylation levels for DMRs 1-3 significantly distinguish normal kidney from Wilms tumor and tumour risk groups. (**a**) Aggregated methylation levels across DMRs 1-3 significantly separated normal kidney (blue) and Wilms tumor (pink) in the replication set (n = 12 pairs, $p=1.47x10^{-9}$). (**b**) Methylation levels were high in an independent dataset of WT including high risk (dark green, n= 25) and intermediate risk histological subtypes (light green, n = 61). High risk WT showed significantly higher methylation levels than intermediate risk WT (p=0.0024).

6.5 Validation of methylation levels

Both DMRs 1 and 2 are located within the extended major histocompatibility complex (MHC) region (Horton, Wilming et al. 2004) which is known to be highly polymorphic (de Bakker, McVean et al. 2006), however the matched study design controlled for genetic heterogeneity, ensuring that the observed signal was not confounded by copy number or DNA sequence variation. To validate the levels of methylation in these regions, high depth coverage sequencing of bisulfite-converted DNA was performed in nine pairs of NK and WT. From the number of reads showing "C" and those showing "T", the proportion of methylated reads was ascertained which is analogous to the β-value. Sequencing of DMR-1 showed a very low number of reads with only 6/18 samples giving 10 or more reads whereas DMR-2 performed better with 16/18 samples giving sufficient reads (Table 6.2). Of the samples where sufficient data was derived, the correlation coefficient for DMR-1 bisulfite-sequencing derived methylation levels compared to the 450k β-values results was R^2=0.92 and for DMR-2 was R^2=0.98 with overall correlation coefficient for all data of R^2=0.93 (Figure 6.5). This data shows that the two platforms convincingly give the same methylation levels for each sample.

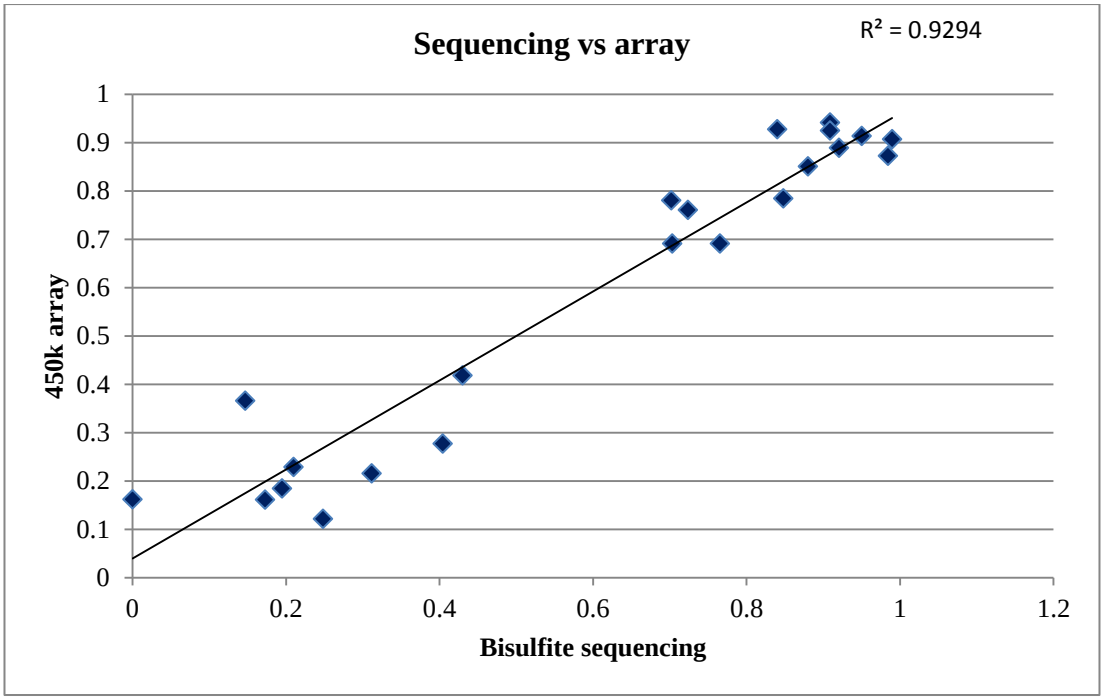


Figure 6.5: Comparison of methylation values assessed by 450k array and bisulfite sequencing. The average β-value for DMR-1 or DMR-2 discerned by the 450k array (y-axis) compared to the average level of methylation detected using bisulfite-sequencing for the same CpGs (x-axis) showed good correlation (R^2 correlation coefficient = 0.929).

DMR	Samples with >10 reads	Correlation coefficient compared to 450k array	Mean difference between 450k value and bisulfite sequencing	Range of difference between 450k value and sequencing value
1	6/18	0.92	0.065	0.01-0.22
2	16/18	0.98	0.073	0.02-0.16

Table 6.2: Validation of 450k methylation signal by bisulfite-sequencing

6.6 Methylation status of embryonic and precursor tissue suggests DMR hypermethylation is associated with malignancy

As embryonic blastema is the WT predicted cell-of-origin, and these DMRs were identified by comparison to normal kidney, it was not clear whether the DMRs existed as a result of retention of an embryonic methylation profile or were solely associated with the malignant state. To explore this question, DNA was extracted from three specimens of human embryonic kidney (EK; at gestational weeks 22, 22 and 23) as well as separately microdissested embryonic blastema from the same specimens (n=3). Bisulfite-sequencing was then carried out on this DNA using the Illumina MiSeq. Bisulfite sequencing identified average levels of methylation (comparative to β-values; on a scale of 0=unmethylated to 1=methylated) for DMRs 1 and 2 of 0.007 and 0.12 for EK and embryonic blastema respectively. This data showed that the embryonic tissue was not methylated. Next, the levels of methylation in 20 cases with matched NRs were examined. These showed intermediate levels of methylation in comparison to NK and WT (Figure 6.6). Put

together, this data suggests that the increase in methylation levels seen in WTs must be associated with transformation of embryonic precursor cells towards a malignant phenotype instead of retention of an embryonic methylome.

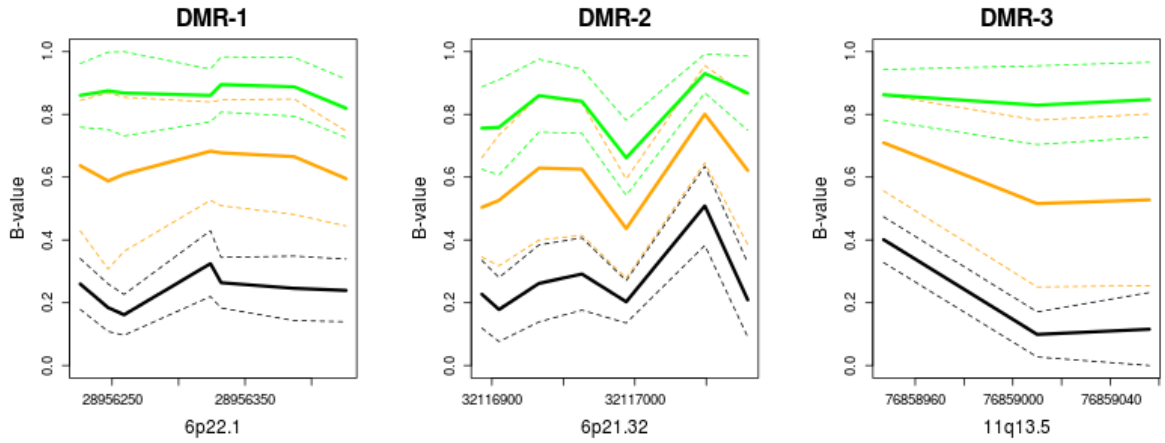


Figure 6.6: WT precursor lesions show intermediate methylation at significant DMRs. Of the 22 cases where genome-wide methylation profiling was carried out on matched normal kidney (NK) and Wilms tumour (WT), 20 had nephrogenic rests (NRs) that were also profiled. The average methylation levels for WT (green), NR (orange) and NK (black) across these 20 cases is shown at each DMR with dotted lines representing the standard deviation. In every case, WT is hypermethylated with respect to NK, and the NR appears at an intermediate state of methylation.

6.7 DMRs as blood biomarkers

As the three DMRs showed high levels of methylation, they were potentially interesting targets for blood biomarkers however if any leukocyte subpopulation was also methylated at these loci, they would mask any methylated DNA released from the tumour (Adalsteinsson, Gudnason et al. 2012). Therefore, to assess the potential influence that the methylomes of each blood subpopulation may have on detection of these DMR biomarkers in patient blood, two analyses were carried out. Firstly, publically available methylation signatures that define normal peripheral blood subgroups were examined (Koestler, Marsit et al. 2012) and secondly, methylation levels for the three DMRs in normal whole blood were examined by extracting 450k methylation data from 411 healthy whole blood samples using Marmal-aid (Lowe and Rakyan 2013) (http://marmal-aid.org/). These analyses showed that DMRs 1-3 did not overlap with any blood-related methylation signature, and that normal blood methylation levels for DMRs 1-3 were extremely low (average β_{mean}= 0.12, Figure 6.7). Therefore, it can be concluded that the hypermethylated DMRs are good targets for detection in the circulation above a low methylation background. Furthermore, they will not be confounded by varying leukocyte populations.

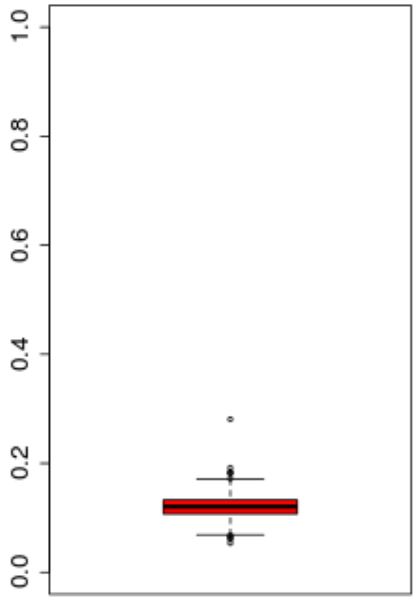


Figure 6.7: Methylation levels of DMRs 1-3 in whole blood. Aggregated methylation levels for DMRs 1-3 in healthy whole blood (n = 411) shows low methylation levels.

6.8 DMRs as biomarkers for other cancers

As previously mentioned, DMRs 1 and 2 are located within the MHC cluster. Methylation of genes in this cluster and reciprocal loss of expression is a common feature of many cancers, as cells use this mechanism to evade immunosurveillance (Doyle, Martin et al. 1985; Jäger, Ringhoffer et al. 1997; Degenhardt, Huang et al. 2010; Garrido, Paco et al. 2012). To see whether other cancers shared the same hypermethylated DMRs, as these appear strong candidates for blood biomarkers over a low blood background, β-values for the three DMRs were extracted for 28 different cancers (n=5,278) using Marmal-aid. The SVM analysis previously used to classify the replication and independent datasets, was used to classify each cancer sample as

"tumour-like" (hypermethylated) or "normal-like" (hypomethylated; Table 6.3). This analysis showed that 8 cancers had >90% cases with hypermethylation at the same loci including prostate adenocarcinoma, uterine corpus endometrial carcinoma, brain low grade glioma, rectal adenocarcinoma, breast invasive carcinoma and bladder urothelial carcinoma. Therefore, these DMRs may act as potential blood biomarkers for several adult and other childhood cancers if the tumour DNA is indeed released into the circulation. Cancer-specific hypermethylation of these loci may in fact be representative of the tissues from which these samples were derived, however as no methylation was detected in any healthy whole blood samples; any tumour-released DNA should still be detected at higher levels in cancer patients compared to cancer-free blood.

Cancer	Total	Hypermethylated	%
Embryocarcinoma	5	5	100.0
Neuroblastoma	3	3	100.0
Prostate Adenocarcinoma	185	183	98.9
Uterine Corpus Endometrioid Carcinoma	383	366	95.6
Rectum Adenocarcinoma	96	90	93.8
Brain Lower Grade Glioma	205	192	93.7
Breast Invasive Carcinoma	586	546	93.2
Bladder Urothelial Carcinoma	153	141	92.2
Lymphoid Neoplasm Diffuse Large B cell Lymphoma	17	15	88.2
Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	122	107	87.7
Glioblastoma Multiforme	323	280	86.7
Colon Adenocarcinoma	330	273	82.7
Liver Hepatocellular Carcinoma	98	80	81.6
Cholangiocarcinoma	50	38	76.0
Stomach Adenocarcinoma	178	126	70.8
Lung Squamous Cell Carcinoma	230	148	64.3
Head and Neck Squamous Cell Carcinoma	412	265	64.3
Lung Adenocarcinoma	380	223	58.7
Skin Cutaneous Melanoma	278	155	55.8
Bone Sarcoma	38	19	50.0
Kidney Renal Papillary Cell Carcinoma	87	38	43.7
Pancreatic Adenocarcinoma	50	17	34.0
Kidney Renal Clear Cell Carcinoma	283	88	31.1
Kidney Chromophobe	65	10	15.4
Thyroid Carcinoma	441	41	9.3
Acute Myeloid Leukaemia	242	20	8.3
Childhood Acute Lymphoblastic Leukaemia	29	2	6.9
Acute Promyelocytic Leukaemia	9	0	0.0

Table 6.3: Classification of 28 tumour types (n=5,278 tumours) showed that other cancers are hypermethylated at DMRs 1-3

6.9 DMR-2 as a biomarker detected in the circulation

As these DMRs appeared to be suitable targets for detection in the circulation, and DMR-2 was successfully sequenced in FFPE tissue, the same assay was performed on DNA extracted from patient serum. Circulating cell-free DNA (cfDNA) is released into the blood by dying cells and can be a marker for a particular disease state. However, only very low concentrations are obtainable.

Serum was taken from 10 patients at serial time points during treatment for WT including pre-chemotherapy (n=5), pre-operative (n=8) and post-operative (n=8) according to the IMPORT study (see chapter 2.4). Due to the difficulty in collecting serum from patients at serial time points, and as the prospective clinical study had only recently opened, not all patients had a full set of samples, and at the time of this experiment, none had an end-of-treatment sample available. Serum samples from age-matched cancer-free controls were also included (n=7) and cfDNA was extracted from these samples using the QIAamp Circulating Nucelic Acid kit (QIAGEN).

In total, 3 pre-operative, 2 post-operative and 3 control samples failed to generate enough DNA for amplification by PCR leaving a set of 20 cfDNA samples for sequencing (Table 6.4). These samples were bisulfite-converted, amplified with primers specific for DMR-2 and sequenced using the Illumina MiSeq. All samples generated sufficient reads for analysis. The entire region sequenced (319 bp: chr6:32,116,940-32,117,259) included 44 CpGs, which did not all follow the same pattern as some showed extreme variation between samples whereas some did not vary at all (i.e. 0.0 across all samples). Those showing no variation were removed as they showed no biological significance.

To identify CpGs that showed consistent methylation at one time-point and differential methylation between time-points, i.e. that gave a reliable, consistent signal, the samples were grouped and an ANOVA test was performed in R which generated CpG-specific statistics for differential methylation. As groups were very small, these statistics were non-significant overall however 14 CpGs showed moderated F-scores greater than 1 which indicated that any of the between-group comparisons were non-zero (i.e. were comparatively different). Therefore, the methylation data for these 14 CpGs was aggregated for each group (to give the % of methylated DMR-2 in patient blood, or $\%M_{mean}$) giving totals of 13.4 for controls, 14.5 for diagnostic samples, 19.9 for samples taken during pre-operative chemotherapy and 19.2 for samples taken in the early post-operative period (Figure 6.8).

These data showed that control samples had the lowest levels of methylated DMR-2 DNA in circulation. Patients had increased levels at diagnosis which then substantially and significantly increased during pre-operative treatment ($p=0.028$). This high level continued into the immediate post-operative phase.

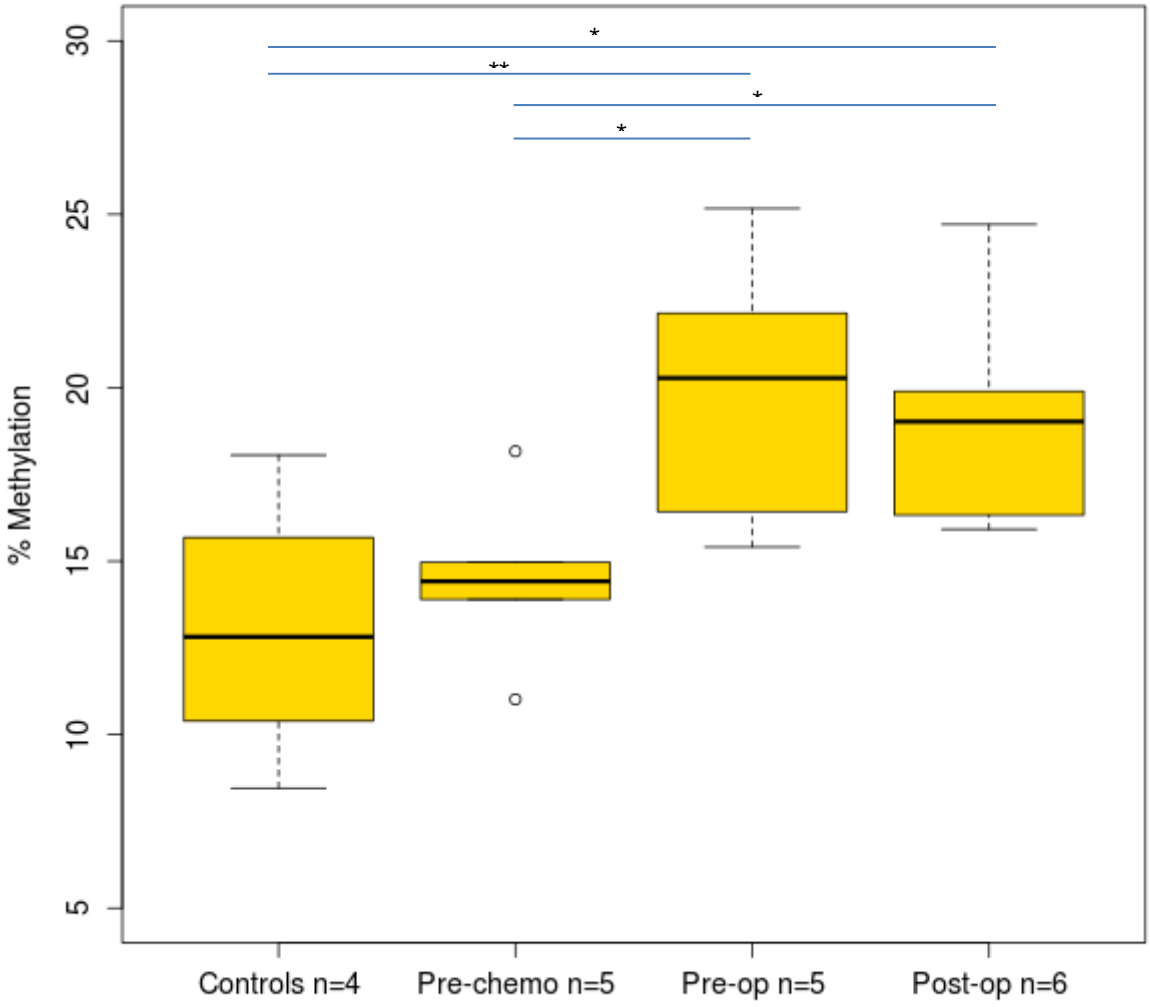


Figure 6.8: Levels of methylation in serum cfDNA during WT treatment (n = 10 patients). Levels of $\%M_{mean}$ (for 14 CpGs within DMR-2) show significant differences between controls and pre-op samples (p=0.010), controls and post-op samples (p=0.017), pre-chemotherapy and pre-op (p=0.028) and pre-chemotherapy and post-op (p=0.047; all by 2-tailed T-test). Data show increasing levels of methylated cfDNA after exposure to chemotherapy. Level of significance is shown by * (<0.05) or ** (<0.01) with horizontal lines indicating the respective comparison.

Patient	$\%M_{mean}$	Time point	Laterality	Days post-surgery serum taken	After pathological review of excised tumour: % of chemotherapy induced changes	Stage	Tumour subtype
Control 1	13.3						
Control 2	8.5						
Control 3	12.3						
Control 4	18.1						
1	14.4	Diagnosis	unilateral		5-10%	1	Epithelial
1	16.4	End of pre-op chemo					
2	13.9	Diagnosis					
2	20.3	Mid pre-op chemo	unilateral		70%	1	Diffuse anaplastic; later bone metastasis
2	24.7	Start of post-op chemo		20			
3	18.2	Diagnosis	unilateral		60%	3	Diffuse anaplastic
4	11	Diagnosis	unilateral		20%	1	Mixed
5	15	Diagnosis	bilateral		right: 40% left upper: 0%	5	right: mixed left upper: mixed
5	18.5	Start of post-op chemo		13	left lower: 80%		left lower: regressive
6	15.4	End of pre-op chemo	unilateral		2-3%	3	Diffuse anaplastic and blastemal
6	16.3	Start of post-op chemo		8			
7	22.2	End of pre-op chemo	unilateral		95%	3	Regressive
8	25.2	End of pre-op chemo	unilateral		90%	3	Regressive
8	19.9	Start of post-op chemo		4			
9	19.6	Start of post-op chemo	unilateral	24	20%	1	Mixed
10	15.9	Start of post-op chemo	unilateral	17	5-10%	1	Epithelial

Table 6.4: Clinical information on serum samples from which cfDNA was isolated. $\%M_{mean}$ = the % of DMR-2 DNA in the circulation that was methylated.

6.10 Summary and conclusions

This chapter described the identification of three differentially methylated regions that were present in 118/120 WTs studied (98%). These biomarkers are therefore far more ubiquitous that any previously identified genetic or epigenetic marker for WT. The RefFreeEWAS algorithm was used to confirm that these were not mediated by cell type composition, which was also shown by comparison of DMR methylation levels to known cell type composition patterns and by the fact that 84/86 WTs that were stratified across all seven post-chemotherapy histological subtypes showed consistently high levels of methylation. It can be concluded that all cells of the tumour acquire gain of methylation at these loci.

Patients treated according to the SIOP protocol generally receive pre-operative chemotherapy, however one patient included in the initial 22 paired discovery set did not, however the tumour still showed high levels of DNA methylation at all three loci indicating that detectable, increased methylation is not chemotherapy-mediated.

To discern the origin of these methylation marks, embryonic kidney, embryonic blastema and nephrogenic rest methylation levels were studied which showed that the embryonic tissue was not methylated while the NRs showed intermediate methylation levels. This analysis confirmed that DMR methylation was not association with retention of an embryonic methylome but associated with malignant transformation.

As a proof-of-principle experiment, it was shown that tumours release methylated DMR-2 DNA into the circulation during treatment. Studying methylation levels in whole blood as well as separated leukocyte population indicated that these DMRs

were not methylated in healthy blood and therefore any methylation observed should be associated with a disease state. The data showed that levels of methylated DMR-2 DNA in the circulation increase during treatment with chemotherapy supporting the idea that the methylated DNA originates from the WT as tumour necrosis is seen in response to chemotherapy. This is validated by the fact that a higher proportion of methylated DNA in the circulation was detected in serum taken from two patients with regressive type WT (defined as >2/3 necrosis at nephrectomy (Vujanic, Sandstedt et al. 2002)) compared to those with non-regressive WTs (23.6% compared to 17.3% respectively).

After nephrectomy, the serum samples showed continued high levels of methylated DNA in the circulation. This was unexpected but may be explained by the fact that serum was sampled between 4 and 24 days post-surgery. The short-term persistence observed may be related to the methylated cfDNA retaining interaction with nucleosome proteins which protect it from degradation (Rumore and Steinman 1990). To fully assess the extent of methylated DNA post-surgery, a full time course up to the end-of treatment included follow-up would be required and is planned for a future study. Indeed, a high level of methylated DNA post-surgery may relate to residual tumour left in the patient. In one patient included in this study, very high levels were observed post-surgery and three months later, bone metastasis was detected.

In conclusion, this study has defined the first epigenetic biomarker applicable to analysis of circulating cfDNA in patients with WT. This biomarker may be useful in determining tumour response during pre-operative chemotherapy. As relapse rate is low in WT patients, a much larger, prospectively collected sample series is needed to

demonstrate clinical utility. This is currently in progress with a multi-centre European clinical trial planned to collect the required samples. These data have been published in Genome Biology, see Appendix 4.

Chapter 7: Final summary

Based on the paucity of frequent genetic mutations in WT, the high frequency of WTs with epigenetic aberration at 11p15 and the embryonic nature of the tumour, six hypotheses were initially proposed. Here, evidence in support of these hypotheses is described and summarised and potential future work is suggested.

The initial aims were *to determine genome-wide methylation profiles for normal kidney, nephrogenic rest and Wilms tumour and to identify epigenetic changes associated with the persistence of NR in post-natal kidney as well as evolution of WT from its presumed precursor in normal embryonic kidney.*

Genome-wide methylation profiles for each tissue were indeed generated showing that DNA methylation can distinguish between tissue types. DNA methylation analysis of NR compared to NK or EK showed that NR formation is associated with hypermethylation of genes involved in renal development and loci that show bivalent chromatin marks in ESCs as well as hypermethylation of PCR2 binding sites. These findings, from analyses performed with and without correction for cell type composition, suggest that the initiating step in Wilms tumourigenesis, i.e. NR retention in post natal kidney, involves PRC2-associated gain of methylation at renal developmental loci, by either an active or passive mechanism, which is not cell composition-mediated. This aberrant methylation at loci required for normal nephrogenic differentiation would therefore result in lesions that cannot complete normal differentiation and remain as embryonic-like tissue in the post natal kidney but with epigenomes arrested from a more primitive stage.

In terms of WT evolution from its associated NR, the evidence presented in this thesis suggests that only a subset of WTs (predominantly those with bilateral disease), show large epigenetic remodelling during evolution from their precursor lesion. Cell composition corrected analysis showed how remarkably similar the two tissues were and how they shared common properties such as hypermethylation of polycomb protein targets. Furthermore, a progressive increase in methylation was shown for genes involved in Wnt/cadherin signalling and for the three biomarker DMRs. Although this evidence suggests that they are true precursor lesions, not every NR transforms into a WT and of the NRs sampled, some may be tumour-associated instead of tumour initiating. In fact, the true NR from which the WT evolved is likely surrounded and compressed by WT and may therefore have been rejected from the stringent selection criteria for this study. This confounding factor may be associated with the identification of two WT groups. Indeed, the WTs in group-1 that appear distinct from their NRs may show a distinct methylation profile due to analysis of a NR that was not a true precursor to the paired WT analysed. Although this seems unlikely, given the strong association between group-1 WTs and bilateral disease, it is however a possible confounder in this analysis. Whether the two WT groups were associated with cell type composition cannot currently be determined as the RefFreeEWAS package does not output a cell type corrected matrix of β-values for MDS analysis or hierarchical clustering.

It is interesting to speculate whether those WTs that are not visibly associated with NRs truly arise without a precursor stage or whether NRs are always present but not visible in many cases. Evidence presented in this thesis shows that both NRs and WTs have hypermethylation of genes involved in renal development which is associated with arrested development. The DNA methylation profile for WTs with

no visible NR would need to be examined to assess whether the hallmarks of NRs described above are present, including hypermethylation of developmental loci. This would suggest that the cells have gone through the step-wise progression from a developmentally arrested precursor prior to transformation.

The next aim was *to identify epigenetic markers that will be useful in distinguishing NR from WT – a particularly challenging differential diagnosis for histopathology, especially in the context of chemotherapy-treated WT.* Although this was detected for group-1 WTs, group-2 tumours showed no significantly different sites of methylation when compared to their precursor NR. The work described in chapter 3 on MDS of the 1% most variable CpGs in the dataset showed an increase in methylation variability between NR and WT samples compared to NK samples. In chapter 4, this was analysed in more detail and the number of probes with significant difference in variance were found to be much higher in WTs than NRs compared to NK. A previous study analysed methylation variability in cancer progression from its precursor in adenocarcinoma and found that it increased during progression towards cancer (Timp and Feinberg 2013) whereas another study found the peak of variability to occur within the intraepithelial neoplasia prior to transformation to cervical cancer (Teschendorff, Liu et al. 2014). The data in this thesis supports the first study as WTs showed increased variability, however if the same comparison were possible for cell composition corrected analyses, the opposite may be seen as the mixed populations seen in WTs will skew variability measures. NRs likely show a great deal of intra-sample variability as they have many different fates, as described in chapter 1.6.2, including dormancy, sclerosis, neoplasia or hyperplasia

(Beckwith, Kiviat et al. 1990) and they have been present since birth and therefore have had time to acquire a variable epigenome. Despite this, the non-corrected findings, showing increase in methylation variability in WT over NR, may prove to be an accurate biomarker for distinguishing WT from NR. Evaluating the clinical use for this biomarker would require a larger study to determine the precise parameters for tissue-specific methylation variability compared to NK.

A pilot analysis of cell-free methylated DNA in the circulation supported the aims: *identify tumour-specific sites of methylation that can be detected in blood serum for use as a biomarker and confirm that the blood methylation biomarkers are discriminative in an independent cohort and have clinical utility*

A biomarker common to 118/120 WTs was identified that was detectable in blood. This biomarker encompassed three regions at 6p22.1, 6p21.32 and 11q13.5 that were hypermethylated in WT compared to NK. In particular, DMR-2 at 6p22.32 was found to be methylated in the circulation of patients with WT, with the highest levels detected after a short period of pre-operative chemotherapy. This marker may be useful for assessing response to chemotherapy or, if sensitivity can be increased, for diagnosis. This requires testing in a much larger series of patients and suitable samples are being collected through the ongoing IMPORT trial for future testing of this hypothesis.

Both regions on chromosome 6 were within the extended major histocompatibility complex (MHC) cluster. Furthermore, *LY6G5C* hypermethylation, also within the MHC cluster, was seen in WTs compared to NR in chapter 5. Hypermethylation of

MHC genes is seen in many cancers associate with avoiding immunosurveillance (Doyle, Martin et al. 1985; Jäger, Ringhoffer et al. 1997; Degenhardt, Huang et al. 2010; Garrido, Paco et al. 2012). Class I and II MHC antigen expression was observed in mature glomeruli and tubule structures in embryonic kidney from 8-13 weeks gestation (Borthwick, Hughes et al. 1988). However in WT, blastemal elements were negative for class I gene expression whereas differentiated elements were positive while class II antigens were not expressed at all in WT (Borthwick, Hughes et al. 1988). This supports that MHC genes show reduced expression in WT. Little further evidence exists on MHC expression or methylation in WTs and is therefore an exciting avenue for future research

The final aim: *To identify changes in methylation associated with progression towards malignancy rather than mixed cell populations by performing both non-corrected and corrected analyses,* was shown to be true as, although some results were common between analyses, different results were also found by both corrected and non-corrected analyses that gain biological insight into WT biology (Figure 7.1).

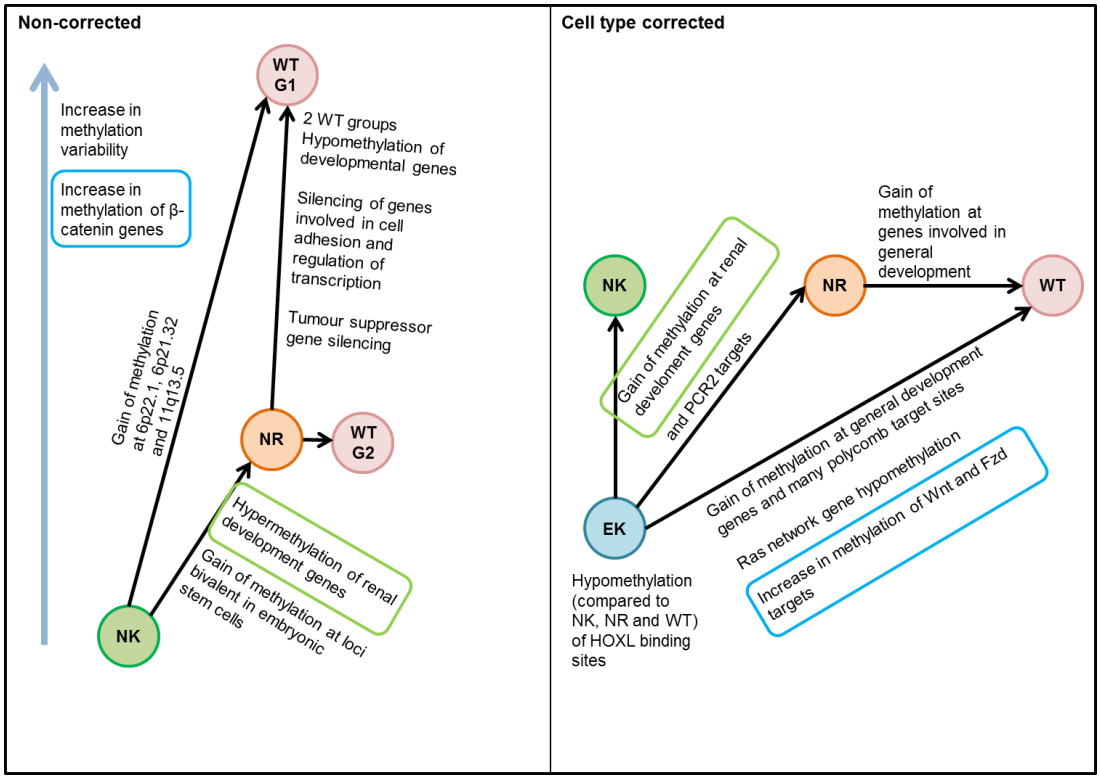


Figure 7.1: A summary of the key findings from both cell type corrected (right panel) and non-corrected (left panel) analyses. Arrows between tissues (EK = embryonic kidney; blue, NK = normal kidney; green, NR = nephrogenic rest; orange, WT = Wilms tumour; red) indicate the direction of the relative description, i.e. an arrow pointing from NK to NR with "hypermethylation" indicates hypermethylation in the NR compared to the NK. In the left panel are findings identifies through analysis of NK, NR and WT whereas in the right panel, the findings of cell-composition corrected analyses including EK are shown. Similar findings between studies are circled in blue or green.

This thesis increases knowledge on NR formation and progression in WTs but also opens up several avenues for future research. Firstly, further study into MHC cluster methylation and expression in WTs is an interesting avenue to aid understanding into why the biomarker DMRs are so significantly hypermethylated in WT. As the costs for 450k analysis are decreasing, a follow-up study to validate the methylation

signature in bilateral WTs may be possible. The limiting element is generation of quality DNA from NRs however with the advance of technology improving the quality of FFPE DNA this may be feasible from less material. By these means it will be possible to test whether any of the biomarkers identified are suitable for clinical use. Combination of methylation data with whole genome sequencing may enlighten which tumours show true epigenetic drivers and which result from genetic mutations and combination with a larger expression dataset by RNA sequencing may further validate the relationship between methylation and gene expression supporting the promising initial evidence in this thesis. Finally, analysing a large cohort of patient serum samples to assess levels of methylated DMR-2 in cell-free DNA will allow for proper interrogation of the potential diagnostic or therapeutic benefits of this blood-borne epigenetic biomarker.

References

Abdel-Wahab, O., A. Mullally, et al. (2009). "Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies." Blood **114**(1): 144-147.

Adalsteinsson, B. T., H. Gudnason, et al. (2012). "Heterogeneity in White Blood Cells Has Potential to Confound DNA Methylation Measurements." PLoS ONE **7**(10): e46705.

Ahmad, K. and S. Henikoff (2002). "The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly." Molecular Cell **9**(6): 1191-1200.

Aiden, A. P., M. N. Rivera, et al. (2010). "Wilms tumor chromatin profiles highlight stem cell properties and a renal developmental network." Cell Stem Cell **6**(6): 591-602.

Akino, K., M. Toyota, et al. (2007). "Identification of DFNA5 as a target of epigenetic inactivation in gastric cancer." Cancer Science **98**(1): 88-95.

Arai, E., Y. Kanai, et al. (2006). "Regional DNA hypermethylation and DNA methyltransferase (DNMT) 1 protein overexpression in both renal tumors and corresponding nontumorous renal tissues." International Journal of Cancer **119**(2): 288-296.

Arcellana-Panlilio, M. Y., R. M. Egeler, et al. (2000). "Decreased expression of the INK4 family of cyclin-dependent kinase inhibitors in Wilms tumor." Genes, Chromosomes and Cancer **29**(1): 63-69.

Balgkouranidou, I., M. Chimonidou, et al. (2014). "Breast cancer metastasis suppressor-1 promoter methylation in cell-free DNA provides prognostic information in non-small cell lung cancer." British journal of cancer.

Bard, J. B., J. E. McConnell, et al. (1994). "Towards a genetic basis for kidney development." Mech Dev **48**(1): 3-11.

Bardeesy, N., D. Falkoff, et al. (1994). "Anaplastic Wilms' tumour, a subtype displaying poor prognosis, harbours p53 gene mutations." Nat Genet **7**(1): 91-97.

Bartel, D. P. (2004). "MicroRNAs: Genomics, Biogenesis, Mechanism, and Function." Cell **116**(2): 281-297.

Beckwith, J. B. (1998). "Nephrogenic rests and the pathogenesis of Wilms tumor: developmental and clinical considerations." Am J Med Genet **79**(4): 268-273.

Beckwith, J. B., N. B. Kiviat, et al. (1990). "Nephrogenic Rests, Nephroblastomatosis, and the Pathogenesis of Wilms' Tumor." Fetal & Pediatric Pathology **10**(1-2): 1-36.

Ben-Porath, I., M. W. Thomson, et al. (2008). "An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors." Nature genetics **40**(5): 499-507.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) **57**(1): 289-300.

Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells." Cell **125**(2): 315-326.

Bielen, A., G. Box, et al. (2012). "Dependence of Wilms tumor cells on signaling through insulin-like growth factor 1 in an orthotopic xenograft model targetable by specific receptor inhibition." Proceedings of the National Academy of Sciences **109**(20): E1267–E1276.

Biniszkiewicz, D., J. Gribnau, et al. (2002). "Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality." Mol Cell Biol **22**(7): 2124-2135.

Borthwick, G. M., L. Hughes, et al. (1988). "Expression of class I and II major histocompatibility complex antigens in Wilms' tumour and normal developing human kidney." British journal of cancer **58**(6): 753-761.

Bourc'his, D., G.-L. Xu, et al. (2001). "Dnmt3L and the Establishment of Maternal Genomic Imprints." Science **294**(5551): 2536-2539.

Boyle, S., A. Misfeldt, et al. (2008). "Fate mapping using Cited1-CreERT2 mice demonstrates that the cap mesenchyme contains self-renewing progenitor cells and gives rise exclusively to nephronic epithelia." Developmental Biology **313**(1): 234-245.
Bracken, A. P., N. Dietrich, et al. (2006). "Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions." Genes & Development **20**(9): 1123-1136.
Bracken, A. P. and K. Helin (2009). "Polycomb group proteins: navigators of lineage pathways led astray in cancer." Nature reviews. Cancer **9**(11): 773-784.
Brennan, K., M. Garcia-Closas, et al. (2012). "Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk." Cancer Research **72**(9): 2304-2313.
Breslow, N., A. Olshan, et al. (1993). "Epidemiology of Wilms tumor." Med Pediatr Oncol **21**(3): 172-181.
Breslow, N. E., J. B. Beckwith, et al. (2006). "Age distributions, birth weights, nephrogenic rests, and heterogeneity in the pathogenesis of Wilms tumor." Pediatric blood & cancer **47**(3): 260-267.
Brodbeck, S., B. Besenbeck, et al. (2004). "The transcription factor Six2 activates expression of the Gdnf gene as well as its own promoter." Mech Dev **121**(10): 1211-1222.
Burke, Z. D. and D. Tosh (2005). "Therapeutic potential of transdifferentiated cells." Clin Sci (Lond) **108**(4): 309-321.
Calin, G. A. and C. M. Croce (2006). "MicroRNA-Cancer Connection: The Beginning of a New Tale." Cancer Research **66**(15): 7390-7394.
Chang, H. M., R. Triboulet, et al. (2013). "A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway." Nature **497**(7448): 244-248.
Chédin, F., M. R. Lieber, et al. (2002). "The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a." Proceedings of the National Academy of Sciences **99**(26): 16916-16921.
Chen, P. Y., S. Feng, et al. (2011). "A comparative analysis of DNA methylation across human embryonic stem cell lines." Genome biology **12**(7): R62.
Chilukamarri, L., A. L. Hancock, et al. (2007). "Hypomethylation and aberrant expression of the glioma pathogenesis-related 1 gene in Wilms tumors." Neoplasia **9**(11): 970-978.
Clark, P. E., D. Polosukhina, et al. (2011). "beta-Catenin and K-RAS synergize to form primitive renal epithelial tumors with features of epithelial Wilms' tumors." The American journal of pathology **179**(6): 3045-3055.
Creyghton, M. P., A. W. Cheng, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proceedings of the National Academy of Sciences.
Dalgliesh, G. L., K. Furge, et al. (2010). "Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes." Nature **463**(7279): 360-363.
Dallosso, A. R., A. L. Hancock, et al. (2009). "Frequent long-range epigenetic silencing of protocadherin gene clusters on chromosome 5q31 in Wilms' tumor." PLoS Genet **5**(11): e1000745.
de Bakker, P. I., G. McVean, et al. (2006). "A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC." Nature genetics **38**(10): 1166-1172.
Degenhardt, Y., J. Huang, et al. (2010). "Distinct MHC gene expression patterns during progression of melanoma." Genes, chromosomes & cancer **49**(2): 144-154.
Diehl, F., K. Schmidt, et al. (2008). "Circulating mutant DNA to assess tumor dynamics." Nat Med **14**(9): 985-990.
Doi, A., I. H. Park, et al. (2009). "Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts." Nat Genet **41**(12): 1350-1353.

Dome, J. S., C. A. Bockhold, et al. (2005). "High telomerase RNA expression level is an adverse prognostic factor for favorable-histology Wilms' tumor." Journal of clinical oncology : official journal of the American Society of Clinical Oncology **23**(36): 9138-9145.
Doyle, A., W. J. Martin, et al. (1985). "Markedly decreased expression of class I histocompatibility antigens, protein, and mRNA in human small-cell lung cancer." The Journal of experimental medicine **161**(5): 1135-1151.
Eggert, A., M. A. Grotzer, et al. (2001). "Expression of the neurotrophin receptor TrkB is associated with unfavorable outcome in Wilms' tumor." Journal of clinical oncology : official journal of the American Society of Clinical Oncology **19**(3): 689-696.
Ehrlich, M., N. E. Hopkins, et al. (2003). "Satellite DNA hypomethylation in karyotyped Wilms tumors." Cancer Genet Cytogenet **141**(2): 97-105.
Ernst, J., P. Kheradpour, et al. (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types." Nature **473**(7345): 43-49.
Ernst, T., A. J. Chase, et al. (2010). "Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders." Nat Genet **42**(8): 722-726.
Esteller, M. (2002). "CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future." Oncogene **21**(35): 5427-5440.
Esteller, M., P. G. Corn, et al. (2001). "A Gene Hypermethylation Profile of Human Cancer." Cancer Research **61**(8): 3225-3229.
Fedoriw, A. M., P. Stein, et al. (2004). "Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting." Science **303**(5655): 238-240.
Feinberg, A. P., R. Ohlsson, et al. (2006). "The epigenetic progenitor origin of human cancer." Nat Rev Genet **7**(1): 21-33.
Feinberg, A. P. and B. Tycko (2004). "The history of cancer epigenetics." Nat Rev Cancer **4**(2): 143-153.
Fitzpatrick, G. V., P. D. Soloway, et al. (2002). "Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1." Nat Genet **32**(3): 426-431.
Francke, U., L. B. Holmes, et al. (1979). "Aniridia-Wilms' tumor association: evidence for specific deletion of 11p13." Cytogenet Cell Genet **24**(3): 185-192.
Fukuzawa, R., R. W. Heathcott, et al. (2007). "Sequential WT1 and CTNNB1 mutations and alterations of β-catenin localisation in intralobar nephrogenic rests and associated Wilms tumours: two case studies." Journal of Clinical Pathology **60**(9): 1013-1016.
Gadd, S., V. Huff, et al. (2012). "Clinically relevant subsets identified by gene expression patterns support a revised ontogenic model of Wilms tumor: a Children's Oncology Group Study." Neoplasia **14**(8): 742-756.
Gallagher, E. J. and D. LeRoith (2010). "The proliferating role of insulin and insulin-like growth factors in cancer." Trends Endocrinol Metab **21**(10): 610-618.
Garrido, C., L. Paco, et al. (2012). "MHC class I molecules act as tumor suppressor genes regulating the cell cycle gene expression, invasion, and intrinsic tumorigenicity of melanoma cells." Carcinogenesis.
Gautschi, O., C. Bigosch, et al. (2004). "Circulating Deoxyribonucleic Acid As Prognostic Marker in Non–Small-Cell Lung Cancer Patients Undergoing Chemotherapy." Journal of Clinical Oncology **22**(20): 4157-4164.
Gerlinger, M., A. J. Rowan, et al. (2012). "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing." N Engl J Med **366**(10): 883-892.
Gonzalez-Gomez, P., M. J. Bello, et al. (2003). "CpG island methylation status and mutation analysis of the RB1 gene essential promoter region and protein-binding pocket domain in nervous system tumours." Br J Cancer **88**(1): 109-114.
Gonzalez-Zulueta, M., C. M. Bender, et al. (1995). "Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing." Cancer Res **55**(20): 4531-4535.
Gowher, H., K. Liebert, et al. (2005). "Mechanism of Stimulation of Catalytic Activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L." Journal of Biological Chemistry **280**(14): 13341-13348.

Gratias, E. J., L. J. Jennings, et al. (2013). "Gain of 1q is associated with inferior event-free and overall survival in patients with favorable histology Wilms tumor: a report from the Children's Oncology Group." Cancer **119**(21): 3887-3894.
Grundy, P. E., N. E. Breslow, et al. (2005). "Loss of heterozygosity for chromosomes 1p and 16q is an adverse prognostic factor in favorable-histology Wilms tumor: a report from the National Wilms Tumor Study Group." J Clin Oncol **23**(29): 7312-7321.
Grundy, P. E., P. E. Telzerow, et al. (1994). "Loss of Heterozygosity for Chromosomes 16q and 1p in Wilms' Tumors Predicts an Adverse Outcome." Cancer Research **54**(9): 2331-2333.
Guerrero-Setas, D., N. Perez-Janices, et al. (2013). "RASSF2 hypermethylation is present and related to shorter survival in squamous cervical cancer." Mod Pathol.
Hansen, K. D., W. Timp, et al. (2011). "Increased methylation variation in epigenetic domains across cancer types." Nat Genet **43**(8): 768-775.
Harada, K., S. Toyooka, et al. (2002). "Deregulation of caspase 8 and 10 expression in pediatric tumors and cell lines." Cancer Res **62**(20): 5897-5901.
Haruta, M., Y. Arai, et al. (2008). "Duplication of paternal IGF2 or loss of maternal IGF2 imprinting occurs in half of Wilms tumors with various structural WT1 abnormalities." Genes Chromosomes Cancer **47**(8): 712-727.
Hauser, S., M. Kogej, et al. (2013). "Serum DNA hypermethylation in patients with bladder cancer: results of a prospective multicenter study." Anticancer research **33**(3): 779-784.
Hauser, S., T. Zahalka, et al. (2013). "Serum DNA hypermethylation in patients with kidney cancer: results of a prospective study." Anticancer research **33**(10): 4651-4656.
Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.
Hing, S., Y. J. Lu, et al. (2001). "Gain of 1q is associated with adverse outcome in favorable histology Wilms' tumors." Am J Pathol **158**(2): 393-398.
Honda, S., Y. Arai, et al. (2008). "Loss of imprinting of IGF2 correlates with hypermethylation of the H19 differentially methylated region in hepatoblastoma." British journal of cancer **99**(11): 1891-1899.
Horton, R., L. Wilming, et al. (2004). "Gene map of the extended human MHC." Nat Rev Genet **5**(12): 889-899.
Houghton, P. J., C. L. Morton, et al. (2010). "Initial testing of a monoclonal antibody (IMC-A12) against IGF-1R by the Pediatric Preclinical Testing Program." Pediatr Blood Cancer **54**(7): 921-926.
Houseman, E. A., W. P. Accomando, et al. (2012). "DNA methylation arrays as surrogate measures of cell mixture distribution." BMC bioinformatics **13**: 86.
Houseman, E. A., J. Molitor, et al. (2014). "Reference-free cell mixture adjustments in analysis of DNA methylation data." Bioinformatics **30**(10): 1431-1439.
Hu, Q., F. Gao, et al. (2011). "Wt1 ablation and Igf2 upregulation in mice result in Wilms tumors with elevated ERK1/2 phosphorylation." J Clin Invest **121**(1): 174-183.
Hubertus, J., M. Lacher, et al. (2011). "Altered expression of imprinted genes in Wilms tumors." Oncol Rep **25**(3): 817-823.
Huelsken, J. and J. Behrens (2002). "The Wnt signalling pathway." Journal of Cell Science **115**(21): 3977-3978.
Iorio, M. V., M. Ferracin, et al. (2005). "MicroRNA Gene Expression Deregulation in Human Breast Cancer." Cancer Research **65**(16): 7065-7070.
Irizarry, R. A., C. Ladd-Acosta, et al. (2009). "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores." Nat Genet **41**(2): 178-186.
Issa, J. P. (2011). "Epigenetic variation and cellular Darwinism." Nature genetics **43**(8): 724-726.
Ito, S., L. Shen, et al. (2011). "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine." Science **333**(6047): 1300-1303.

Jaffe, A. E. and R. A. Irizarry (2014). "Accounting for cellular heterogeneity is critical in epigenome-wide association studies." Genome biology **15**(2): R31.
Jäger, E., M. Ringhoffer, et al. (1997). "Immunoselection in vivo: Independent loss of MHC class I and melanocyte differentiation antigen expression in metastatic melanoma." International Journal of Cancer **71**(2): 142-147.
Janssen, B. L., L. T. van de Locht, et al. (1999). "Transcription of the MAGE-1 gene and the methylation status of its Ets binding promoter elements: a quantitative analysis in melanoma cell lines using a real-time polymerase chain reaction technique." Melanoma Res **9**(3): 213-222.
Jeong, M., D. Sun, et al. (2014). "Large conserved domains of low DNA methylation maintained by Dnmt3a." Nat Genet **46**(1): 17-23.
Jjingo, D., A. B. Conley, et al. (2012). "On the presence and role of human gene-body DNA methylation." Oncotarget **3**(4): 462-474.
Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond." Nature reviews. Genetics **13**(7): 484-492.
Kang, M. J., B. J. Park, et al. (2000). "Loss of imprinting and elevated expression of wild-type p73 in human gastric adenocarcinoma." Clin Cancer Res **6**(5): 1767-1771.
Kim, D., G. Pertea, et al. (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biology **14**(4): R36.
Kim, M. S., X. Chang, et al. (2008). "Aberrant promoter methylation and tumor suppressive activity of the DFNA5 gene in colorectal carcinoma." Oncogene **27**(25): 3624-3634.
Kim, M. S., C. Lebron, et al. (2008). "Methylation of the DFNA5 increases risk of lymph node metastasis in human breast cancer." Biochemical and Biophysical Research Communications **370**(1): 38-43.
Ko, M., Y. Huang, et al. (2010). "Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2." Nature **468**(7325): 839-843.
Koesters, R., R. Ridder, et al. (1999). "Mutational activation of the beta-catenin proto-oncogene is a common event in the development of Wilms' tumors." Cancer Res **59**(16): 3880-3882.
Koestler, D. C., C. J. Marsit, et al. (2012). "Peripheral Blood Immune Cell Methylation Profiles Are Associated with Nonhematopoietic Cancers." Cancer Epidemiology Biomarkers & Prevention **21**(8): 1293-1302.
Kohli, R. M. and Y. Zhang (2013). "TET enzymes, TDG and the dynamics of DNA demethylation." Nature **502**(7472): 472-479.
Kolb, E. A., R. Gorlick, et al. (2011). "Initial testing (stage 1) of the IGF-1 receptor inhibitor BMS-754807 by the pediatric preclinical testing program." Pediatr Blood Cancer **56**(4): 595-603.
Kort, E. J., L. Farber, et al. (2008). "The E2F3-Oncomir-1 axis is activated in Wilms' tumor." Cancer Res **68**(11): 4034-4038.
Kreidberg, J. A., H. Sariola, et al. (1993). "WT-1 is required for early kidney development." Cell **74**(4): 679-691.
Kulis, M., S. Heath, et al. (2012). "Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia." Nat Genet **44**(11): 1236-1242.
Lakomy, R., J. Sana, et al. (2011). "MiR-195, miR-196b, miR-181c, miR-21 expression levels and O-6-methylguanine-DNA methyltransferase methylation status are associated with clinical outcome in glioblastoma patients." Cancer Science **102**(12): 2186-2190.
Lee, E. J., Y. Gusev, et al. (2007). "Expression profiling identifies microRNA signature in pancreatic cancer." International Journal of Cancer **120**(5): 1046-1054.
Lee, J. J., E. Ko, et al. (2012). "Methylation and Immunoexpression of p16(INK4a) Tumor Suppressor Gene in Primary Breast Cancer Tissue and Their Quantitative p16(INK4a) Hypermethylation in Plasma by Real-Time PCR." Korean J Pathol **46**(6): 554-561.

Lemerle, J., P. A. Voute, et al. (1983). "Effectiveness of preoperative chemotherapy in Wilms' tumor: results of an International Society of Paediatric Oncology (SIOP) clinical trial." Journal of Clinical Oncology **1**(10): 604-609.
Li, C. M., M. Guo, et al. (2002). "Gene expression in Wilms' tumor mimics the earliest committed stage in the metanephric mesenchymal-epithelial transition." Am J Pathol **160**(6): 2181-2190.
Li, C. M., C. E. Kim, et al. (2004). "CTNNB1 mutations and overexpression of Wnt/beta-catenin target genes in WT1-mutant Wilms' tumors." Am J Pathol **165**(6): 1943-1953.
Li, D., Y. Zhao, et al. (2011). "Analysis of MiR-195 and MiR-497 Expression, Regulation and Role in Breast Cancer." Clinical Cancer Research **17**(7): 1722-1730.
Li, Y. M., G. Franklin, et al. (1998). "The H19 transcript is associated with polysomes and may regulate IGF2 expression in trans." The Journal of biological chemistry **273**(43): 28247-28252.
Lin, R.-K., H.-S. Hsu, et al. (2007). "Alteration of DNA methyltransferases contributes to 5′CpG methylation and poor prognosis in lung cancer." Lung Cancer **55**(2): 205-213.
Lonergan, G. J., M. I. Martinez-Leon, et al. (1998). "Nephrogenic rests, nephroblastomatosis, and associated lesions of the kidney." Radiographics : a review publication of the Radiological Society of North America, Inc **18**(4): 947-968.
Lowe, R. and V. K. Rakyan (2013). "Marmal-aid--a database for Infinium HumanMethylation450." BMC bioinformatics **14**: 359.
Lu, J., G. Getz, et al. (2005). "MicroRNA expression profiles classify human cancers." Nature **435**(7043): 834-838.
Ludgate, J. L., G. Le Mee, et al. (2012). "Global demethylation in loss of imprinting subtype of wilms tumor." Genes Chromosomes Cancer.
Maiti, S., R. Alam, et al. (2000). "Frequent Association of β-Catenin and WT1 Mutations in Wilms Tumors." Cancer Research **60**(22): 6288-6292.
Major, M. B., N. D. Camp, et al. (2007). "Wilms Tumor Suppressor WTX Negatively Regulates WNT/ß-Catenin Signaling." Science **316**(5827): 1043-1046.
Malekzadeh, K., R. C. Sobti, et al. (2009). "Methylation patterns of Rb1 and Casp-8 promoters and their impact on their expression in bladder cancer." Cancer Invest **27**(1): 70-80.
Malik, K., A. Salpekar, et al. (2000). "Identification of Differential Methylation of the WT1 Antisense Regulatory Region and Relaxation of Imprinting in Wilms' Tumor." Cancer Res **60**(9): 2356-2360.
Mann, J. R. (2001). "Imprinting in the germ line." Stem Cells **19**(4): 287-294.
Martens, J. H. A., H. G. Stunnenberg, et al. (2011). "The Decade of the Epigenomes?" Genes & Cancer **2**(6): 680-687.
Mayer, W., A. Niveleau, et al. (2000). "Embryogenesis: Demethylation of the zygotic paternal genome." Nature **403**(6769): 501-502.
McCarthy, D. J. and G. K. Smyth (2009). "Testing significance relative to a fold-change threshold is a TREAT." Bioinformatics **25**(6): 765-771.
McDonald, J. M., E. C. Douglass, et al. (1998). "Linkage of Familial Wilms' Tumor Predisposition to Chromosome 19 and a Two-Locus Model for the Etiology of Familial Tumors." Cancer Research **58**(7): 1387-1390.
McDonald, O. G., H. Wu, et al. (2011). "Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition." Nat Struct Mol Biol **18**(8): 867-874.
McLean, C. Y., D. Bristor, et al. (2010). "GREAT improves functional interpretation of cis-regulatory regions." Nature biotechnology **28**(5): 495-501.
Meneghini, M. D., M. Wu, et al. (2003). "Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin." Cell **112**(5): 725-736.
Merlet-Benichou, C., T. Gilbert, et al. (1999). "Nephron number: variability is the rule. Causes and consequences." Lab Invest **79**(5): 515-527.

Merlo, A., J. G. Herman, et al. (1995). "5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers." Nat Med **1**(7): 686-692.
Merrill, B. J. (2012). "Wnt pathway regulation of embryonic stem cell self-renewal." Cold Spring Harbor perspectives in biology **4**(9): a007971.
Messahel, B., R. Williams, et al. (2009). "Allele loss at 16q defines poorer prognosis Wilms tumour irrespective of treatment approach in the UKW1-3 clinical trials: a Children's Cancer and Leukaemia Group (CCLG) Study." European Journal of Cancer **45**(5): 819-826.
Metsuyanim, S., N. Pode-Shakked, et al. (2008). "Accumulation of Malignant Renal Stem Cells Is Associated with Epigenetic Changes in Normal Renal Progenitor Genes." Stem Cells **26**(7): 1808-1817.
Miki, T., S. Y. Yasuda, et al. (2011). "Wnt/beta-catenin signaling in embryonic stem cell self-renewal and somatic cell reprogramming." Stem cell reviews **7**(4): 836-846.
Mikkelsen, T. S., M. Ku, et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature **448**(7153): 553-560.
Mohammad, H. P., Y. Cai, et al. (2009). "Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation." Cancer Research **69**(15): 6322-6330.
Moisan, A., Miguel N. Rivera, et al. (2011). "The WTX Tumor Suppressor Regulates Mesenchymal Progenitor Cell Fate Specification." Developmental Cell **20**(5): 583-596.
Moore, A. W., L. McInnes, et al. (1999). "YAC complementation shows a requirement for Wt1 in the development of epicardium, adrenal gland and throughout nephrogenesis." Development **126**(9): 1845-1857.
Mori, H., S. M. Colman, et al. (2002). "Chromosome translocations and covert leukemic clones are generated during normal fetal development." Proceedings of the National Academy of Sciences of the United States of America **99**(12): 8242-8247.
Mori, T., S. J. O'Day, et al. (2005). "Predictive Utility of Circulating Methylated DNA in Serum of Melanoma Patients Receiving Biochemotherapy." Journal of Clinical Oncology **23**(36): 9351-9358.
Morison, I. M., C. J. Paton, et al. (2001). "The imprinted gene and parent-of-origin effect database." Nucleic Acids Res **29**(1): 275-276.
Morris, M. R., D. Astuti, et al. (2013). "Perlman syndrome: overgrowth, Wilms tumor predisposition and DIS3L2." Am J Med Genet C Semin Med Genet **163C**(2): 106-113.
Morris, M. R., L. B. Hesson, et al. (2003). "Multigene methylation analysis of Wilms' tumour and adult renal cell carcinoma." Oncogene **22**(43): 6794-6801.
Morris, T., L. Butcher, et al. (2013). "450k Chip Analysis Methylation Pipeline (ChAMP)." Bioinformatics.
Mugford, J. W., P. Sipila, et al. (2008). "Hoxd11 specifies a program of metanephric kidney development within the intermediate mesoderm of the mouse embryo." Dev Biol **319**(2): 396-405.
Mundlos, S., J. Pelletier, et al. (1993). "Nuclear localization of the protein encoded by the Wilms' tumor gene WT1 in embryonic and adult tissues." Development **119**(4): 1329-1341.
Murphy, A. J., J. Pierce, et al. (2012). "SIX2 and CITED1, markers of nephronic progenitor self-renewal, remain active in primitive elements of Wilms' tumor." Journal of pediatric surgery **47**(6): 1239-1249.
Nakamura, M., Y. Yonekawa, et al. (2001). "Promoter hypermethylation of the RB1 gene in glioblastomas." Lab Invest **81**(1): 77-82.
Nakanishi, H., T. Suda, et al. (2004). "Loss of imprinting of PEG1/MEST in lung cancer cell lines." Oncol Rep **12**(6): 1273-1278.
Natrajan, R., S. E. Little, et al. (2006). "Amplification and overexpression of CACNA1E correlates with relapse in favorable histology Wilms' tumors." Clinical cancer

research : an official journal of the American Association for Cancer Research **12**(24): 7284-7293.
Natrajan, R., J. S. Reis-Filho, et al. (2006). "Blastemal expression of type I insulin-like growth factor receptor in Wilms' tumors is driven by increased copy number and correlates with relapse." Cancer Research **66**(23): 11148-11155.
Natrajan, R., R. D. Williams, et al. (2006). "Array CGH profiling of favourable histology Wilms tumours reveals novel gains and losses associated with relapse." The Journal of Pathology **210**(1): 49-58.
Nikoloski, G., S. M. Langemeijer, et al. (2010). "Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes." Nat Genet **42**(8): 665-667.
Nishigaki, M., K. Aoyagi, et al. (2005). "Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays." Cancer Res **65**(6): 2115-2124.
Nishinakamura, R. (2003). "Kidney development conserved over species: essential roles of Sall1." Semin Cell Dev Biol **14**(4): 241-247.
Ohlsson, R., H. Cui, et al. (1999). "Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity." Cancer Res **59**(16): 3889-3892.
Ohm, J. E., K. M. McGarvey, et al. (2007). "A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing." Nature genetics **39**(2): 237-242.
Ohnishi, K., K. Semi, et al. (2014). "Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation." Cell **156**(4): 663-677.
Okamoto, K., I. M. Morison, et al. (1997). "Epigenetic changes at the insulin-like growth factor II/H19 locus in developing kidney is an early event in Wilms tumorigenesis." Proceedings of the National Academy of Sciences **94**(10): 5367-5371.
Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-257.
Pan, G., S. Tian, et al. (2007). "Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells." Cell Stem Cell **1**(3): 299-312.
Park, J.-S., M. T. Valerius, et al. (2007). "Wnt/β-catenin signaling regulates nephron induction during mouse kidney development." Development **134**(13): 2533-2539.
Park, J. Y., D. Kim, et al. (2013). "Gene silencing of SLC5A8 identified by genome-wide methylation profiling in lung cancer." Lung Cancer **79**(3): 198-204.
Paull, T. T., E. P. Rogakou, et al. (2000). "A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage." Current biology : CB **10**(15): 886-895.
Pelletier, J., W. Bruening, et al. (1991). "Germline mutations in the Wilms' tumor suppressor gene are associated with abnormal urogenital development in Denys-Drash syndrome." Cell **67**(2): 437-447.
Pidsley, R., Y. W. CC, et al. (2013). "A data-driven approach to preprocessing Illumina 450K methylation array data." BMC Genomics **14**: 293.
Pode-Shakked, N., O. Harari-Steinberg, et al. (2011). "Resistance or sensitivity of Wilms' tumor to anti-FZD7 antibody highlights the Wnt pathway as a possible therapeutic target." Oncogene **30**(14): 1664-1680.
Pode-Shakked, N., R. Shukrun, et al. (2013). "The isolation and characterization of renal cancer initiating cells from human Wilms' tumour xenografts unveils new therapeutic targets." EMBO molecular medicine **5**(1): 18-37.
Ponomaryova, A. A., E. Y. Rykova, et al. (2013). "Potentialities of aberrantly methylated circulating DNA for diagnostics and post-treatment follow-up of lung cancer patients." Lung Cancer **81**(3): 397-403.

Preger-Ben Noon, E., H. Barak, et al. (2009). "Interplay between activin and Hox genes determines the formation of the kidney morphogenetic field." Development **136**(12): 1995-2004.
Price, M. E., A. M. Cotton, et al. (2013). "Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array." Epigenetics & chromatin **6**(1): 4.
Rada-Iglesias, A., R. Bajpai, et al. (2011). "A unique chromatin signature uncovers early developmental enhancers in humans." Nature **470**(7333): 279-283.
Rahman, N., L. Arbour, et al. (1996). "Evidence for a familial Wilms' tumour gene (FWT1) on chromosome 17q12-q21." Nat Genet **13**(4): 461-463.
Raisner, R. M., P. D. Hartley, et al. (2005). "Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin." Cell **123**(2): 233-248.
Rakyan, V. K., H. Beyan, et al. (2011). "Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis." PLoS genetics **7**(9): e1002300.
Ramirez, J. L., R. Rosell, et al. (2005). "14-3-3σ Methylation in Pretreatment Serum Circulating DNA of Cisplatin-Plus-Gemcitabine-Treated Advanced Non–Small-Cell Lung Cancer Patients Predicts Survival: The Spanish Lung Cancer Group." Journal of Clinical Oncology **23**(36): 9105-9112.
Ramsahoye, B. H., D. Biniszkiewicz, et al. (2000). "Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a." Proceedings of the National Academy of Sciences **97**(10): 5237-5242.
Reik, W., K. W. Brown, et al. (1995). "Imprinting mutations in the Beckwith—Wiedemann syndrome suggested by an altered imprinting pattern in the IGF2–H19 domain." Human Molecular Genetics **4**(12): 2379-2385.
Reik, W., K. W. Brown, et al. (1995). "Imprinting mutations in the Beckwith—Wiedemann syndrome suggested by an altered imprinting pattern in the IGF2–H19 domain." Hum Mol Genet **4**(12): 2379-2385.
Reya, T., A. W. Duncan, et al. (2003). "A role for Wnt signalling in self-renewal of haematopoietic stem cells." Nature **423**(6938): 409-414.
Rivera, M. N. and D. A. Haber (2005). "Wilms' tumour: connecting tumorigenesis and organ development in the kidney." Nat Rev Cancer **5**(9): 699-712.
Rivera, M. N., W. J. Kim, et al. (2009). "The tumor suppressor WTX shuttles to the nucleus and modulates WT1 activity." Proceedings of the National Academy of Sciences **106**(20): 8338-8343.
Roll, J. D., A. G. Rivenbark, et al. (2008). "DNMT3b overexpression contributes to a hypermethylator phenotype in human breast cancer cell lines." Mol Cancer **7**: 15.
Rumore, P. M. and C. R. Steinman (1990). "Endogenous circulating DNA in systemic lupus erythematosus. Occurrence as multimeric complexes bound to histone." The Journal of clinical investigation **86**(1): 69-74.
Sakai, T., J. Toguchida, et al. (1991). "Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene." Am J Hum Genet **48**(5): 880-888.
Sakatani, T., A. Kaneda, et al. (2005). "Loss of Imprinting of Igf2 Alters Intestinal Maturation and Tumorigenesis in Mice." Science **307**(5717): 1976-1978.
Sato, N., H. Matsubayashi, et al. (2005). "Epigenetic down-regulation of CDKN1C/p57KIP2 in pancreatic ductal neoplasms identified by gene expression profiling." Clin Cancer Res **11**(13): 4681-4688.
Sato, N., L. Meijer, et al. (2004). "Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor." Nat Med **10**(1): 55-63.
Schedl, A. (2007). "Renal abnormalities and their developmental origin." Nat Rev Genet **8**(10): 791-802.
Schlesinger, Y., R. Straussman, et al. (2007). "Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer." Nature genetics **39**(2): 232-236.

Schmitt, J., C. Backes, et al. (2012). "Treatment-independent miRNA signature in blood of wilms tumor patients." BMC Genomics **13**: 379.
Scott, R. H., A. Murray, et al. (2012). "Stratification of Wilms tumor by genetic and epigenetic analysis." Oncotarget **3**(3): 327-335.
Scott, R. H., C. A. Stiller, et al. (2006). "Syndromes and constitutional chromosomal abnormalities associated with Wilms tumour." Journal of Medical Genetics **43**(9): 705-715.
Simpson, D. J., N. A. Hibberts, et al. (2000). "Loss of pRb Expression in Pituitary Adenomas Is Associated with Methylation of the RB1 CpG Island." Cancer Research **60**(5): 1211-1216.
Sims, R. J., 3rd and D. Reinberg (2008). "Is there a code embedded in proteins that is based on post-translational modifications?" Nature reviews. Molecular cell biology **9**(10): 815-820.
Smith, Z. D., M. M. Chan, et al. (2012). "A unique regulatory phase of DNA methylation in the early mammalian embryo." Nature **484**(7394): 339-344.
Smyth, G. K. (2005). limma: Linear Models for Microarray Data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. R. Gentleman, V. Carey, W. Huber, R. Irizarry and S. Dudoit, Springer New York**:** 397-420.
Sneeringer, C. J., M. P. Scott, et al. (2010). "Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas." Proc Natl Acad Sci U S A **107**(49): 20980-20985.
Sokol, S. Y. (2011). "Maintaining embryonic stem cell pluripotency with Wnt signaling." Development **138**(20): 4341-4350.
Song, D. J., L. F. Yue, et al. (2013). "[Relationship between mRNA expression and promoter methylation status of p73 gene in peripheral blood among children with Wilms' tumor]." Zhongguo dang dai er ke za zhi = Chinese journal of contemporary pediatrics **15**(8): 638-643.
Steenman, M. J., S. Rainier, et al. (1994). "Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour." Nature genetics **7**(3): 433-439.
Tahiliani, M., K. P. Koh, et al. (2009). "Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1." Science **324**(5929): 930-935.
Takai, D. and P. A. Jones (2002). "Comprehensive analysis of CpG islands in human chromosomes 21 and 22." Proc Natl Acad Sci U S A **99**(6): 3740-3745.
Teodoridis, J. M., C. Hardie, et al. (2008). "CpG island methylator phenotype (CIMP) in cancer: causes and implications." Cancer letters **268**(2): 177-186.
Teschendorff, A. E., A. Jones, et al. (2012). "Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation." Genome medicine **4**(3): 24.
Teschendorff, A. E., X. Liu, et al. (2014). "The dynamics of DNA methylation covariation patterns in carcinogenesis." PLoS computational biology **10**(7): e1003709.
Thirlwell, C., M. Eymard, et al. (2010). "Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip." Methods **52**(3): 248-254.
Tian, F., S. P. Yip, et al. (2013). "Promoter hypermethylation of tumor suppressor genes in serum as potential biomarker for the diagnosis of nasopharyngeal carcinoma." Cancer epidemiology **37**(5): 708-713.
Timp, W. and A. P. Feinberg (2013). "Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host." Nat Rev Cancer **13**(7): 497-510.
Torres, M., E. Gomez-Pardo, et al. (1995). "Pax-2 controls multiple steps of urogenital development." Development **121**(12): 4057-4065.

Torrezan, G. T., E. N. Ferreira, et al. (2014). "Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour." Nature communications **5**: 4039.
Turnbull, C., E. R. Perdeaux, et al. (2012). "A genome-wide association study identifies susceptibility loci for Wilms tumor." Nat Genet **44**(6): 681-684.
Urbach, A., A. Yermalovich, et al. (2014). "Lin28 sustains early renal progenitors and induces Wilms tumor." Genes & Development.
van Haaften, G., G. L. Dalgliesh, et al. (2009). "Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer." Nat Genet **41**(5): 521-523.
Vire, E., C. Brenner, et al. (2006). "The Polycomb group protein EZH2 directly controls DNA methylation." Nature **439**(7078): 871-874.
Volinia, S., G. A. Calin, et al. (2006). "A microRNA expression signature of human solid tumors defines cancer gene targets." Proceedings of the National Academy of Sciences of the United States of America **103**(7): 2257-2261.
Vujanic, G. M., B. Sandstedt, et al. (2002). "Revised International Society of Paediatric Oncology (SIOP) working classification of renal tumors of childhood." Medical and Pediatric Oncology **38**(2): 79-82.
Vuononvirta, R., N. J. Sebire, et al. (2008). "Perilobar nephrogenic rests are nonobligate molecular genetic precursor lesions of insulin-like growth factor-II-associated Wilms tumors." Clin Cancer Res **14**(23): 7635-7644.
Wagner, K. J., W. N. Cooper, et al. (2002). "Frequent RASSF1A tumour suppressor gene promoter methylation in Wilms' tumour and colorectal cancer." Oncogene **21**(47): 7277-7282.
Wang, J., L. Wang, et al. (2012). "Identification of potential serum biomarkers for Wilms tumor after excluding confounding effects of common systemic inflammatory factors." Molecular Biology Reports **39**(5): 5095-5104.
Wang, Q., M. Williamson, et al. (2007). "Hypomethylation of WNT5A, CRIP1 and S100P in prostate cancer." Oncogene **26**(45): 6560-6565.
Watson, J. A., K. Bryan, et al. (2013). "miRNA profiles as a predictor of chemoresponsiveness in Wilms' tumor blastema." PLoS ONE **8**(1): e53417.
Weaver, J. R., M. Susiarjo, et al. (2009). "Imprinting and epigenetic changes in the early embryo." Mamm Genome **20**(9-10): 532-543.
Wegert, J., S. Bausenwein, et al. (2011). "Retinoic acid pathway activity in Wilms tumors and characterization of biological responses in vitro." Molecular cancer **10**: 136.
Wellik, D. M., P. J. Hawkes, et al. (2002). "Hox11 paralogous genes are essential for metanephric kidney induction." Genes & Development **16**(11): 1423-1432.
Widschwendter, M., H. Fiegl, et al. (2007). "Epigenetic stem cell signature in cancer." Nature genetics **39**(2): 157-158.
Williams, R. D., R. Al-Saadi, et al. (2010). "Subtype-specific FBXW7 mutation and MYCN copy number gain in Wilms' tumor." Clin Cancer Res **16**(7): 2036-2045.
Williams, R. D., R. Al-Saadi, et al. (2011). "Molecular profiling reveals frequent gain of MYCN and anaplasia-specific loss of 4q and 14q in wilms tumor." Genes, Chromosomes and Cancer **50**(12): 982-995.
Wright, K. D., D. M. Green, et al. (2009). "Late effects of treatment for wilms tumor." Pediatric hematology and oncology **26**(6): 407-413.
Wu, M. K., N. Sabbaghian, et al. (2013). "Biallelic DICER1 mutations occur in Wilms tumours." The Journal of Pathology **230**(2): 154-164.
Xie, W., M. D. Schultz, et al. (2013). "Epigenomic analysis of multilineage differentiation of human embryonic stem cells." Cell **153**(5): 1134-1148.
Xu, P.-X., J. Adams, et al. (1999). "Eya1-deficient mice lack ears and kidneys and show abnormal apoptosis of organ primordia." Nat Genet **23**(1): 113-117.
Yu, Y., F. Xu, et al. (1999). "NOEY2 (ARHI), an imprinted putative tumor suppressor gene in ovarian and breast carcinomas." Proc Natl Acad Sci U S A **96**(1): 214-219.
Zhang, L., M. S. Anglesio, et al. (2007). "The E3 ligase HACE1 is a critical chromosome 6q21 tumor suppressor involved in multiple cancers." Nat Med **13**(9): 1060-1069.

Ziller, M. J., F. Muller, et al. (2011). "Genomic distribution and inter-sample variation of non-CpG methylation across human cell types." PLoS genetics **7**(12): e1002389.
Ziller, M. J., F. Müller, et al. (2011). "Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types." PLoS Genet **7**(12): e1002389.
Zirn, B., O. Hartmann, et al. (2006). "Expression profiling of Wilms tumors reveals new candidate genes for different clinical parameters." Int J Cancer **118**(8): 1954-1962.
Zirn, B., B. Samans, et al. (2005). "All-trans retinoic acid treatment of Wilms tumor cells reverses expression of genes associated with high risk and relapse in vivo." Oncogene **24**(33): 5246-5251.
Zitzmann, F., D. Mayr, et al. (2014). "Frequent hypermethylation of a CTCF binding site influences Wilms tumor 1 expression in Wilms tumors." Oncol Rep **31**(4): 1871-1876.
Zou, J., C. Lippert, et al. (2014). "Epigenome-wide association studies without the need for cell-type composition." Nature methods **11**(3): 309-311.

Appendix

1: Table A1 – Wilms tumour staging classifications for patients treated according to the SIOP protocol, from Vujanic et al (Vujanic, Sandstedt et al. 2002)

Stage	Criteria
I	(a) The tumour limited to kidney or surrounded with fibrous pseudocapsule if outside of the normal contours of the kidney, the renal capsule or pseudocapsule may be infiltrated with the tumour but it does not reach the outer surface, and it is completely resected (resection margins 'clear')
	(b) The tumour may be protruding ('bulging') into the pelvic system and 'dipping' into the ureter (but it is not infiltrating their walls)
	(c) The vessels of the renal sinus are not involved
	(d) Intra-renal vessel involvement may be present
II	(a) The tumour extends beyond kidney or penetrates through the renal capsule and/or fibrous pseudocapsule into perirenal fat but is completely resected (resection margins 'clear')
	(b) The tumour infiltrates the renal sinus and/or invades blood and lymphatic vessels outside the renal parenchyma but it is completely resected
	(c) The tumour infiltrates adjacent organs or vena cava but is completely resected
III	(a) Incomplete excision of the tumour which extends beyond resection margins (gross or microscopical tumour remains postoperatively)
	(b) Any abdominal lymph nodes are involved
	(c) Tumour rupture before or intraoperatively (irrespective of other criteria for staging)
	(d) The tumour has penetrated through the peritoneal surface
	(e) Tumour implants are found on the peritoneal surface
	(f) The tumour thrombi present at resection margins of vessels or ureter, transsected or removed piecemeal by surgeon
	(g) The tumour has been surgically biopsied (wedge biopsy) prior to preoperative chemotherapy or surgery
IV	Haematogenous metastases (lung, liver, bone, brain, etc.) or lymph node metastases outside the abdomino-pelvic region
V	Bilateral renal tumours at diagnosis. Each side should be substaged according to the above criteria.

2: Lin28 sustains early renal progenitors and induces Wilms tumor. Achia Urbach, Alena Yermalovich, Jin Zhang, Catherine S. Spina, Hao Zhu, Antonio R. Perez-Atayde, Rachel Shukrun, Jocelyn Charlton, Neil Sebire, William Mifsud, Benjamin Dekel, Kathy Pritchard-Jones and George Q. Daley. *Genes and Development*, doi:10.1101/gad.237149.113 (2014).

Lin28 sustains early renal progenitors and induces Wilms tumor

Achia Urbach, Alena Yermalovich, Jin Zhang, et al.

Genes Dev. published online April 14, 2014
Access the most recent version at doi:10.1101/gad.237149.113

Lin28 sustains early renal progenitors and induces Wilms tumor

Achia Urbach,[1,2,3,4] Alena Yermalovich,[1,2,3,4] Jin Zhang,[1,2,3,4] Catherine S. Spina,[1,2,3,4] Hao Zhu,[5,6,7] Antonio R. Perez-Atayde,[8] Rachel Shukrun,[9,10] Jocelyn Charlton,[11] Neil Sebire,[12] William Mifsud,[12] Benjamin Dekel,[9,10] Kathy Pritchard-Jones,[11] and George Q. Daley[1,2,3,4,13]

[1]Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Children's Hospital Boston, Boston, Massachusetts 02115, USA; [2]Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA; [3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA; [4]Harvard Stem Cell Institute, Boston, Massachusetts 02115, USA; [5]Children's Research Institute, [6]Department of Pediatrics, [7]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA; [8]Department of Pathology, Boston Children's Hospital, Boston, Massachusetts 02115, USA; [9]Pediatric Stem Cell Research Institute, [10]Division of Pediatric Nephrology, Sheba Medical Center, Sackler School of Medicine, Tel Aviv University, Tel Aviv 52621, Israel; [11]Institute of Child Health, University College London, London WC1H 0AJ, United Kingdom; [12]Department of Histopathology, Camelia Botnar Laboratories, Great Ormond Street Hospital for Children, London WC1N 3JH, United Kingdom

Wilms Tumor, the most common pediatric kidney cancer, evolves from the failure of terminal differentiation of the embryonic kidney. Here we show that overexpression of the heterochronic regulator *Lin28* during kidney development in mice markedly expands nephrogenic progenitors by blocking their final wave of differentiation, ultimately resulting in a pathology highly reminiscent of Wilms tumor. Using lineage-specific promoters to target *Lin28* to specific cell types, we observed Wilms tumor only when *Lin28* is aberrantly expressed in multiple derivatives of the intermediate mesoderm, implicating the cell of origin as a multipotential renal progenitor. We show that withdrawal of *Lin28* expression reverts tumorigenesis and markedly expands the numbers of glomerulus-like structures and that tumor formation is suppressed by enforced expression of *Let-7* microRNA. Finally, we demonstrate overexpression of the *LIN28B* paralog in a significant percentage of human Wilms tumor. Our data thus implicate the *Lin28/Let-7* pathway in kidney development and tumorigenesis.

[*Keywords*: kidney development; Wilms tumor; Lin28]

Supplemental material is available for this article.

Received December 24, 2013; revised version accepted March 25, 2014.

Wilms tumor, a common pediatric kidney cancer affecting one in 10,000 children in North America, arises from the failure of embryonic nephrogenic cells to undergo terminal differentiation (Rivera and Haber 2005). The development of the kidney is a complex process that requires reciprocal inductive interactions between the ureteric bud (UB) and metanephric mesenchyme (MM), which leads to proliferation and expansion of the primitive cap mesenchyme (CM) (Grobstein 1955, 1956; Hatini et al. 1996; Davidson 2009). The CM cells differentiate into mature nephrons by a mesenchymal-to-epithelial transition (MET). As they also possess self-renewal capacity, CM cells represent embryonic kidney stem cells (Kobayashi et al. 2008; Pleniceanu et al. 2010). CM cells proliferate and differentiate in the outer nephrogenic zone of the kidney until the second postnatal day in mice (Hartman et al. 2007) and the 36th week of gestation in humans (Hinchliffe et al. 1991), after which time all remaining CM cells synchronously differentiate to establish the final number of nephrons in the adult kidney (Hartman et al. 2007). Wilms tumor shares histological features with the developing kidney and is frequently associated with persistent areas of embryonic tissue known as nephrogenic rests, which contain blastemal cells with varying degrees of differentiation (Rivera and Haber 2005).

Lin28 is an RNA-binding protein that regulates gene expression via two different mechanisms: one that blocks the processing of the *Let-7* family of microRNAs (miRNAs) (Heo et al. 2008; Newman et al. 2008; Rybak et al. 2008; Viswanathan et al. 2008) and another that involves direct

[13]**Corresponding author**
E-mail george.daley@childrens.harvard.edu
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.237149.113.

binding to a wide array of mRNA targets (for review, see Shyh-Chang and Daley 2013). The *Let-7*-dependent mechanism entails binding to the terminal loop of pri/pre-*Let-7* miRNAs, which prevents their maturation and thus enables the translation of genes that are suppressed by *Let-7* miRNAs (Viswanathan et al. 2008). Oncogenes such as *K-Ras* and *c-Myc* are prominent *Let-7* targets (Viswanathan and Daley 2010). In mammals, *Lin28A* and its closely related paralog, *Lin28B*, are highly expressed in pluripotent cells, where they play an important role in the maintenance of self-renewal and proliferation (Shyh-Chang and Daley 2013). Both Lin28 proteins are highly expressed in early embryonic development but become down-regulated over time, while levels of mature *Let-7* family members rise as stem cells differentiate into specialized tissue types (Viswanathan and Daley 2010). Overexpression of *LIN28* is common in various tumor types and facilitates cellular transformation (Viswanathan et al. 2009). *Lin28* also promotes reprogramming of somatic cells into induced pluripotent cells (Yu et al. 2007).

Given that *Lin28* is highly active in embryonic tissues and was originally described as a heterochronic gene that regulates developmental timing in *Caenorhabditis elegans* (Ambros and Horvitz 1984; Moss et al. 1997), we hypothesized that *Lin28* overexpression might play a role in pediatric tumor formation by altering the timing of tissue differentiation and organogenesis during embryonic development. Indeed *LIN28A* overexpression has been implicated in type II germ cell tumors (Gillis et al. 2011), which result from a failure of differentiation of primordial germ cells (PGCs) (Oosterhuis and Looijenga 2005), while *LIN28B* has been linked to neuroblastoma (Diskin et al. 2012), a pediatric tumor derived from neural crest tissues that fail to complete their differentiation program (Maris 2010; Molenaar et al. 2012). Last, we reported previously that in rare cases of human Wilms tumor, *LIN28B* overexpression is caused by translocation at the *LIN28B* locus (Viswanathan et al. 2009). Here we describe a novel murine model of Wilms tumor caused by enforced overexpression of *Lin28* during embryonic kidney development and demonstrate by immunohistochemistry that LIN28B is overexpressed in up to 30% of cases of human Wilms tumor. These data, together with recent insights from whole-genome sequencing of Wilms tumor, implicate defects in miRNA regulation as a major mechanism of kidney tumorigenesis.

Results

Lin28 overexpression during embryonic kidney development leads to Wilms tumor

Previously, we and others have shown that *LIN28* plays an important role in germ cell development (West et al. 2009; Shinoda et al. 2013a) and is associated with human germ cell tumors (Gillis et al. 2011; Murray et al. 2013). Thus, we endeavored to overexpress *Lin28* in PGCs by crossing mice containing a *Lox-stop-Lox-Lin28a* cassette (LSL-*Lin28a*) (Supplemental Fig. S1A) with mice carrying a Vasa-Cre transgene, which we anticipated would express Cre in PGCs when transmitted paternally, allowing us to test the potential for *Lin28* to induce germ cell tumors (Gallardo et al. 2007). Contrary to expectations, however, the cross between a LSL-*Lin28a* female and *Vasa*Cre male did not yield the predicted germ cell phenotype (zero out of 50) but unexpectedly produced renal tumors in 10% of the offspring (five out of 50; four bilateral and one unilateral) (Fig. 1A, top left). Tumors expressed the *Lin28a* transgene, apparently a consequence of aberrant "leaky" activation, whereas normal kidneys showed no transgene expression (Fig. 1B). Crosses of LSL-*Lin28a* males with females carrying the Vasa-Cre allele resulted in constitutional overexpression in all tissues by virtue of Cre expression in oocytes (Gallardo et al. 2007) and perinatal lethality. Interestingly, the kidneys of transgenic embryonic day 18.5 (E18.5) embryos were larger than the kidneys of their littermate controls and contained fewer mature proximal tubules (Supplemental Fig. S1B). When we harvested the kidneys from E18.5 transgenic and control embryos and transplanted them under the kidney capsule of immunodeficient mice, tumors developed in a high percentage of recipients (seven out of 10) (Fig. 1A, top right panel; Supplemental Fig. S1C). No tumors formed in transplant recipients of control kidneys (zero out of nine). Analysis of tumor gene expression (Fig. 1C) and histology (Fig. 1D) indicated that the *Lin28a*-derived tumors were highly similar to human Wilms tumor.

Previously, we reported two cases of human Wilms tumor in which *LIN28B* was overexpressed as a result of chromosomal translocation (Viswanathan et al. 2009). To determine whether human *LIN28B* overexpression would replicate Wilms tumor formation in mice, we engineered a transgenic strain that afforded spatial and temporal control of human *LIN28B* (or mouse *Lin28a*) overexpression by crossing the Rosa26-Lox-stop-Lox-rtTA allele with the Col1A1-TRE-*LIN28B* allele (Lox-TetOn-*LIN28B* mice) (Supplemental Fig. S1D; Zhu et al. 2010). To achieve global *LIN28B* overexpression in the developing and/or adult kidney, we crossed Lox-TetOn-*LIN28B* mice with *Wt1*Cre mice (Zhou et al. 2008), as *Wt1* is expressed in the intermediate mesoderm (Huff 2011), the origin of the metanephric kidney (Davidson 2009). All *Wt1*Cre-*LIN28B* mice (15 out of 15) developed kidney tumors (Fig. 1A, bottom panel) within the first 2 wk of life when exposed to doxycycline (Dox) induction during embryonic development (E0, E14.5, or even as late as E18.5) (see below). Importantly, the histology of the *LIN28B*-derived tumors was similar to *Lin28a*-derived tumors (Fig. 1D). Taken together, these results establish that overexpression of either murine *Lin28a* or human *LIN28B* during kidney development in transgenic strains of mice leads to kidney tumor formation that is highly reminiscent of human Wilms tumor (Fig. 1E).

Lin28 overexpression sustains the CM cells in the adult kidney

During kidney development, the nephronogenic progenitor cells of the CM cells differentiate into pretubular epithelial aggregates by a MET at around E12.5 (Rivera

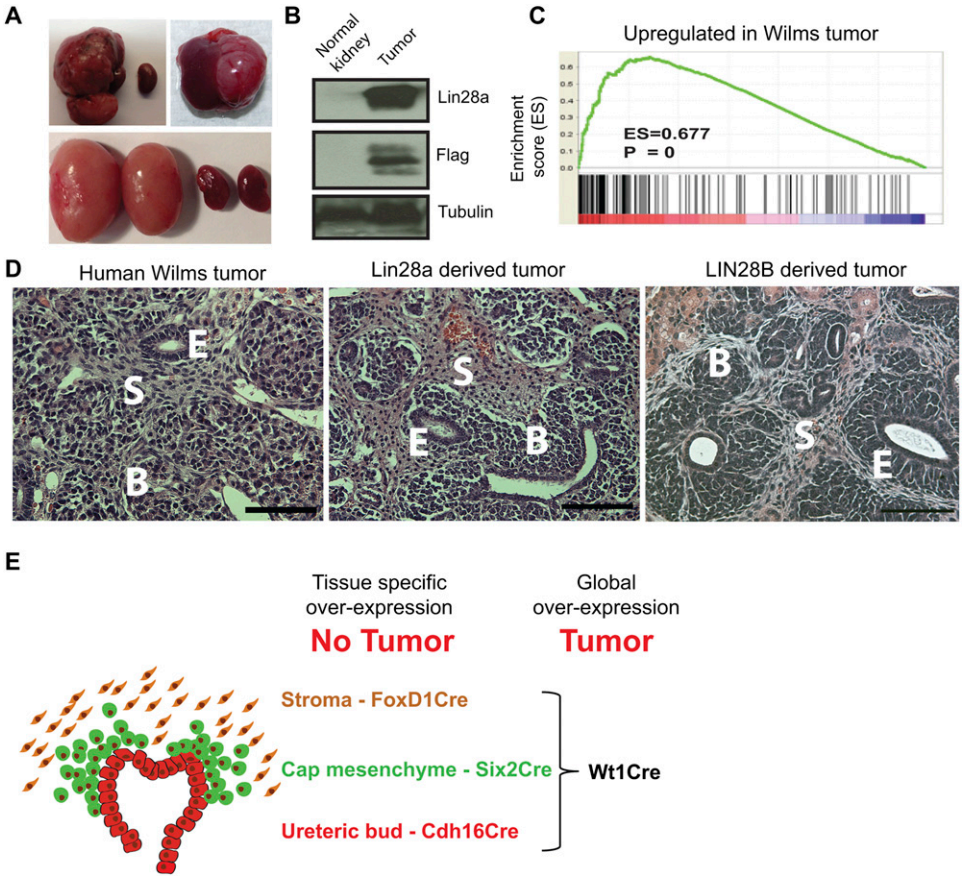


Figure 1. *Lin28* overexpression in embryonic kidneys leads to Wilms tumor. (*A*) Kidney tumors as a result of *Lin28* overexpression in the kidney. (*Top left* panel) Renal tumor in a 17-wk-old kidney from the crossing between a LSL-*Lin28a* female and a *Vasa*Cre male. The smaller kidney is the normal kidney from the opposite side in the same mouse. (*Top right* panel) Tumor derived by transplantation of a *Lin28a*-overexpressing kidney (from the crossing between a LSL-*Lin28a* male and a *Vasa*Cre female) under the kidney capsule of an immunodeficient mouse 17 wk post-transplantation. (*Bottom* panel) *LIN28B*-derived tumors in 3-wk-old kidneys (Dox induction [1g/L] from E0). (*B*) Western blot analysis of tumors and age-matched normal kidneys with antibodies against Lin28a, Flag, and Tubulin. The expression of the Flag tag shows activation of the transgenic *Lin28a* in the tumor. (*C*) Gene set enrichment analysis (GSEA) of microarray data from *Lin28a*-derived tumors and control kidneys showing statistically significant up-regulation of "Wilms tumor signature genes" in the tumor compared with the control. (*D*) Typical histology of a human Wilms tumor, a *Lin28a*-derived tumor, and a *LIN28B*-derived tumor. The tumors are triphasic and contain structures of blastema (B), epithelium (E), and mesenchymal stromal (S) cells. Bar, 100 μm. (*E*) Schematic representation of the UB and CM cells, which normally exist in the nephrogenic zone of the mouse kidney only until P2. Global *Lin28a* or *LIN28B* overexpression in the developing kidney leads to a pathology similar to human Wilms tumor. However, lineage-specific overexpression did not cause renal tumor formation (see below).

and Haber 2005). Normally, a balance between differentiation and proliferation of CM cells is sustained in the nephrogenic zone of the developing kidney until postnatal day 2 (P2), after which time all CM cells undergo a terminal wave of differentiation (Rumballe et al. 2011). In contrast to normal kidneys, the *Lin28*-derived tumors continue to sustain proliferating CM cells, as evidenced by the expression of CM-specific transcription factors (e.g., *Six2*, *Cited1*, and *Eya1*) and Ki67 staining (Fig. 2A,B; Supplemental Fig. S2A). Moreover, H&E (Fig. 2C) and immunofluorescence (Fig. 2D; Supplemental Fig. S2B,C) staining of the tumors demonstrates that the tumor consists of keratin8-positive UB cells surrounded by Six2-positive CM cells, similar to the structures that normally exist in the nephrogenic zone of the developing kidney. Taken together, these data indicate that *Lin28* overexpression prolongs the timing of kidney development, sustaining proliferation of the CM cells into adulthood. Interestingly, the CM cells of the tumor retain their differentiation capacity, as evident by gene expression for markers of epithelialization, such as *Wnt4* and *CDH6* (K-cadherin) (Fig. 3A; Supplemental Fig S3A); gross histology (Fig. 3B; Supplemental Fig S3B); and histologic staining for Lotus tetragonolobus lectin (LTL), a specific marker for mature proximal tubules (Fig. 3C). Importantly, the fact that the differentiated epithelial cells are Flag-positive (Supplemental Fig. S3C) indicates that these cells are indeed derived from the *Lin28*-overexpressing cells. Furthermore, the *Lin28*-derived tumors also contain structures resembling differentiated glomeruli (Supplemental Fig. S3D,E). Thus, the CM cells within the tumor retain a differentiation capacity that recapitulates normal kidney development.

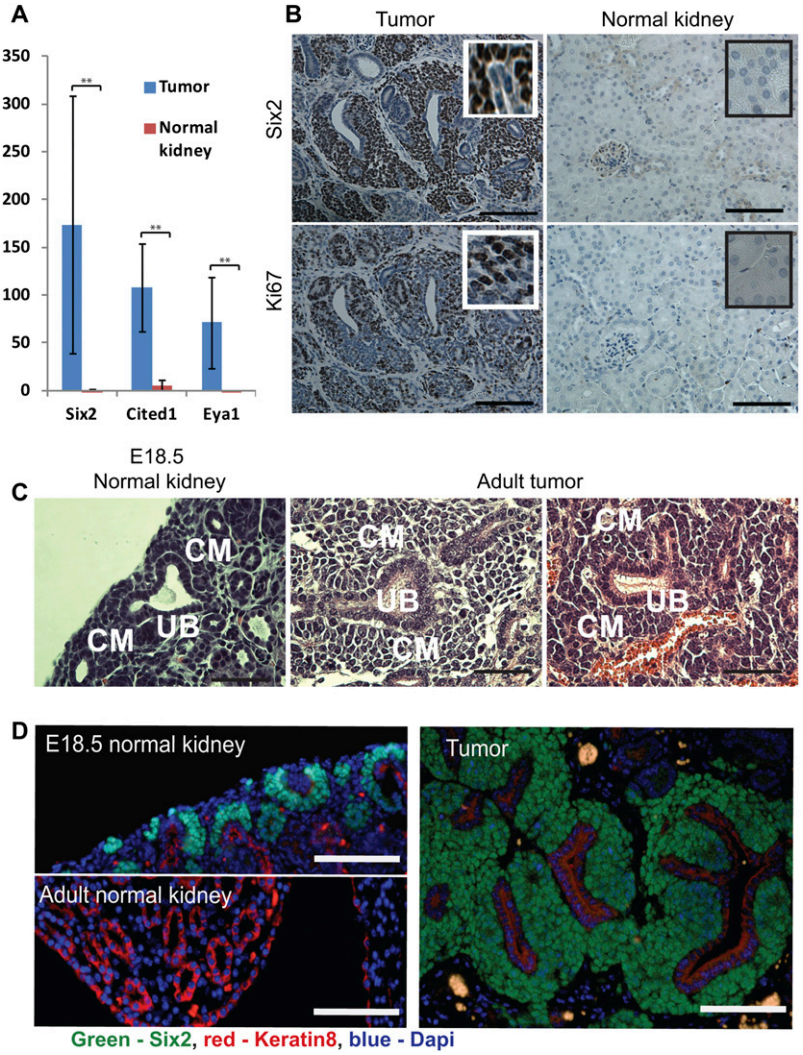


Figure 2. *Lin28*-derived tumors harbor proliferating CM cells. (*A*) Overexpression of CM-specific transcription factors in the tumor. Microarray data (tumor, $n = 4$; control, $n = 4$). (*B*) Six2 and Ki67 immunohistochemistry in normal kidneys and *LIN28a*-derived tumors. Bar, 100 μm. (*C*) Histology of the E18.5 nephrogenic zone and adult tumors. Note that embryonic structures consisting of a branched UB surrounded by CM cells appears in the adult tumors. Bar, 50 μm. (*D*) Six2 and Keratin8 coimmunostaining in a normal embryonic kidney, a normal adult kidney, and an adult tumor. (Note: Keratin 8 is a marker for UB cells during kidney development and for the collecting duct in the adult kidney). In contrast to normal adult kidneys, in which Six2 is not expressed and keratin8 expression is restricted to the collecting duct, the expression pattern of Six2 and Keratin8 in the tumors is similar to the nephrogenic zone during kidney development. Bar, 100 μm.

Lin28 prevents the postnatal wave of differentiation of the CM cells

We documented that *Lin28* overexpression leads to persistent proliferation of CM cells in adult mice. During normal kidney development, *Lin28a* is expressed in CM cells until E13.5, after which expression wanes, while *Lin28b* is not expressed at all (Fig. 4A; Supplemental Fig. S4A). Interestingly, however, the nephrogenic zone of E18.5 transgenic embryos appears normal, without aberrant expansion of the CM cells (Fig. 4B; Supplemental Fig. S4B), suggesting normal proliferation of the CM cells in the nephrogenic zone during embryonic development in the presence of *Lin28*. Importantly, induction of *Lin28* overexpression as late as E18.5 was enough to sustain proliferation of CM cells into adulthood (Supplemental Fig. S4C). To discern whether ectopic *Lin28* expression could reactivate proliferation of nephrogenic cells after the early postnatal period of terminal differentiation, we induced *LIN28B* overexpression in renal tissues by virtue of gene activation via *Wt1*Cre at P10, when no CM cells exist in the normal kidney. When *LIN28B* was overexpressed at P10, tumors failed to develop, and expansion of *Six2*-positive cells was not detected (Supplemental Fig. S4D). Instead, late *LIN28B* induction resulted in a cystic kidney phenotype (Supplemental Fig. S4E). Collectively, these data suggest that the CM cells retain their differentiation capacity in the presence of *Lin28* expression but that *Lin28* delays the timing of the final postnatal wave of synchronous differentiation, allowing the nephrogenic process to persist, eventually producing Wilms tumor.

Wilms tumor arises when Lin28 is overexpressed in multiple early kidney lineages

We then sought to determine in which cells of the developing kidney overexpression of *Lin28* was required to promote Wilms tumor formation. We overexpressed *LIN28B* or *Lin28a* specifically in CM cells by crossing *Lox-tetOn-LIN28B* and Lox-TetOn-*Lin28a* transgenic mice with *Six2*-Cre mice (Kobayashi et al. 2008). Contrary to expectation, overexpression of *Lin28a* or *LIN28B* in the CM cells failed to induce tumor forma-

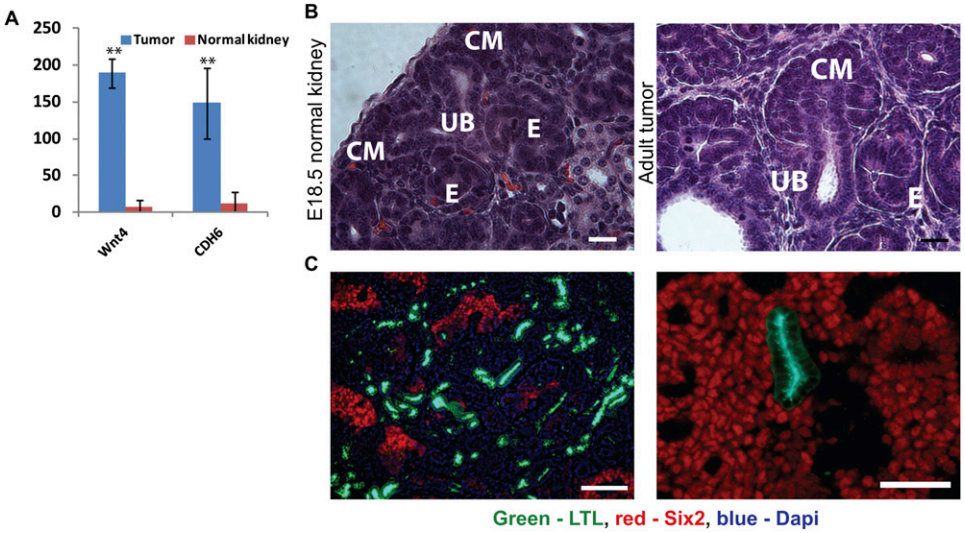


Figure 3. Differentiation capacity of the CM cells in the tumor. (*A*) Overexpression of epithelialization markers Wnt4 and Cdh16 in the tumor. Microarray data (tumor, $n = 4$; normal kidney, $n = 4$). (*B*) Epithelial structures in a normal embryonic kidney and a *Lin28*-derived tumor. (E) Epithelial structures of comma-shaped/S-shaped bodies differentiated from the CM cells. Bar, 20 μm. (*C*) Mixture of progenitor cells (Six2-positive) and mature proximal tubule cells (LTL-positive) in the tumor. Bar, 100 μm.

tion (zero out of 15) but instead produced a cystic kidney phenotype. The cystic phenotype appeared in the *Six2*-Cre-*Lin28* mice when *Lin28* overexpression was induced early in embryonic development (Supplemental Figs. S5A,B) or in adult mice (Supplemental Fig. S5C). We crossed Lox-TetOn-*Lin28a/LIN28B* mice with mice carrying *FoxD1*Cree ($n = 10$) to effect stromal cell-specific expression (Humphreys et al. 2010) and with mice carrying *Cdh16*Cre ($n = 7$) to effect UB cell-specific expression (Shao et al. 2002), but neither of these crossings reproduced the tumor phenotype. Interestingly, however, overexpression of *LIN28B* in stromal cells (*FoxD1*Cre) led to hydronephrosis in the adult kidney (Supplemental Fig. S5D), while no pathology was seen in the *Cdh16*Cre mice. Therefore, in the murine model, *Lin28* activation is required in the earliest renal progenitor that gives rise to the multiple cell types of the developing kidney, implying that *Lin28* promotes a coordinated prolongation of nephrogenesis, which ultimately progresses to Wilms tumor formation (see Fig. 1E).

Lin28 expression is required for the maintenance of CM cells within the tumor

Our data firmly establish a role for *Lin28* overexpression in tumor initiation, but to determine whether continued

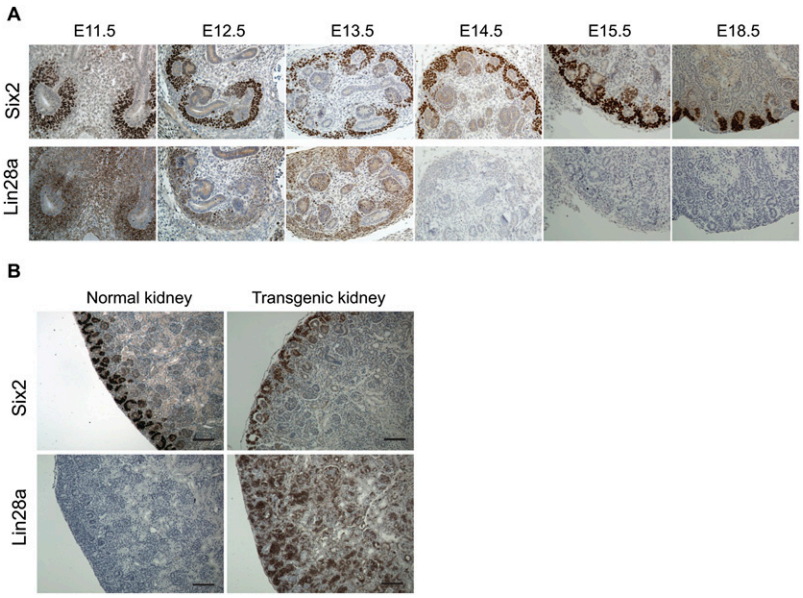


Figure 4. *Lin28* overexpression prevents the postnatal wave of differentiation of the CM cells. (*A*) Analysis of Lin28a expression during mouse embryonic kidney development. Lin28a is expressed in the CM (Six2-positive) cells of the developing kidney until E13.5. All panels are at the same scale. (*B*, *top* panel) Six2 expression in E18.5 normal and transgenic kidneys. (*Bottom* panels) Lin28a expression in normal and transgenic kidneys. Note that there is no expansion in CM cells of the transgenic kidney compared with the control. Bar, 100 μm.

Lin28 expression is necessary for tumor maintenance, we studied the effect of *Lin28* down-regulation by withdrawal of Dox commencing at P7. While significant numbers of *Six2/Eya1*-positive CM cells persist 2 wk after Dox withdrawal (Fig. 5A; Supplemental Fig S6), at 3 wk after Dox withdrawal, there was a significant decrease in CM cells in the kidneys of the transgenic mice, and not overt tumors, but a markedly increased number of well differentiated glomerulus-like structures (Fig. 5B,C). Thus, we conclude that expression of *Lin28* is sufficient for tumor initiation and necessary for tumor maintenance.

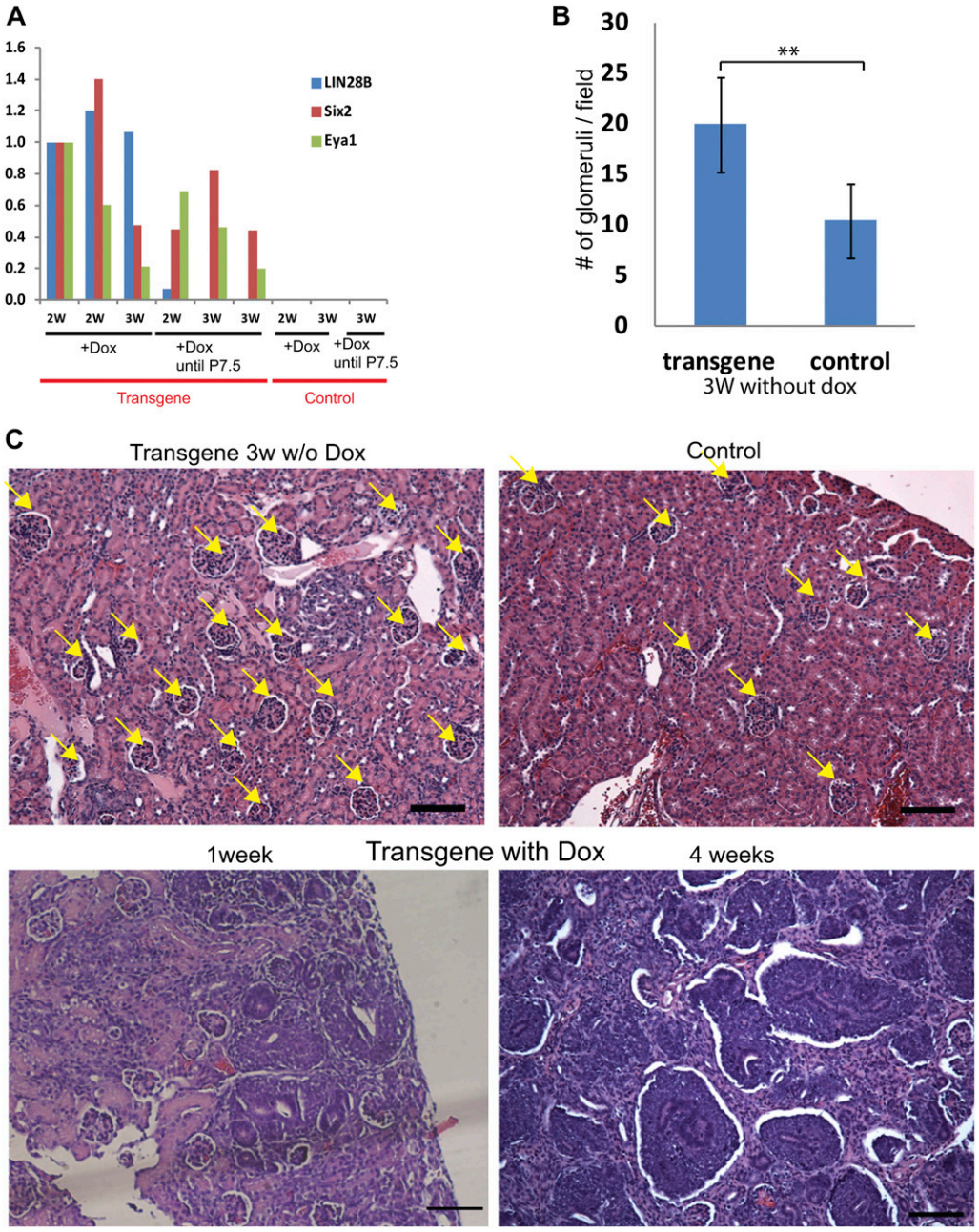


Figure 5. *Lin28* down-regulation leads to differentiation of the CM cells in the tumor. (*A*) qRT–PCR analysis of *LIN28B*, *Six2*, and *Eya1* (markers form CM cells) in transgenic and control kidneys 1 and 2 wk after Dox withdrawal. (2W) 2-wk-old mice (1 wk without Dox); (3W) 3-wk-old mice (2 wk without Dox). Note that *Six2* and *Eya1* are still expressed 2 wk after Dox is withdrawn. (*B*) Average glomerulus number per 10× magnification field in the microscope in transgenic and control kidneys 3 wk after Dox withdrawal (n = 12). (*C*) Representative H&E staining of transgenic kidneys 3 wk after Dox withdrawal (transgene 3 wk w/o Dox) compared with a control 4-wk-old kidney and transgenic 1-wk-old and 4-wk-old kidneys maintained on Dox (transgene with Dox). Dox was induced from E0. Note that 3 wk after Dox withdrawal, the transgenic kidneys contain many glumerulus-like structures and do not contain CM cells. Yellow arrows point to glomerulus-like structures. Bar, 100 μm.

Lin28-induced Wilms tumor is suppressed by enforced expression of Let-7

Analysis of miRNA expression in *Lin28*-induced tumors demonstrated significant suppression of mature *Let-7* species (Supplemental Fig. S7A) but no significant change in steady-state levels of pri/pre-*Let7* (Supplemental Fig. S7B). In accordance with this, Let7 target genes were up-regulated in the tumors compared with control kidneys (Supplemental Fig S7C). To determine whether *Lin28*-induced tumorigenesis could be suppressed by enforced restoration of *Let-7*, we crossed the tumor-prone *LIN28B* transgenic mice with a strain that expresses a chimeric *Let-7g* (i7s) species whose processing is not inhibited by *Lin28* (Piskounova et al. 2008). We verified by immunohistochemistry staining that the transgenic kidneys overexpressed LIN28B (Supplemental Fig. S7D), that endogenous *Let-7* was down-regulated when the chimeric *Let-7g* transgene was overexpressed in the *LIN28B-i7s* kidneys (Fig. 6A), and that Let-7 targets were down-regulated in the *Lin28B;i7s* kidneys compared with *Lin28B* kidneys (Supplemental Fig. S7E). We showed previously that the "i7s" mice are smaller than their littermate controls (Zhu

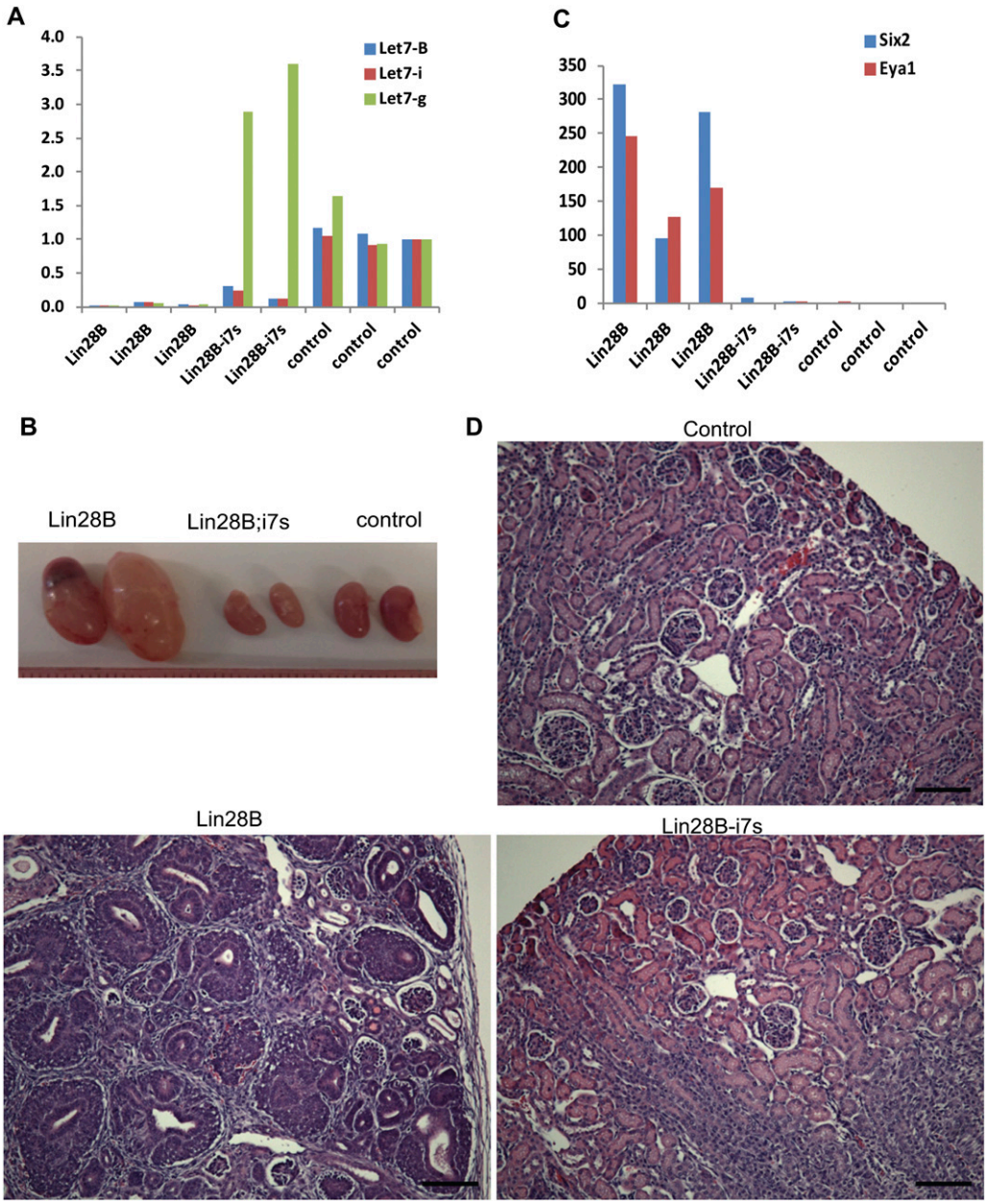


Figure 6. *Lin28* acts thorough the *Let7* pathway to prevent normal kidney development. (*A*) qRT–PCR analysis of mature *Let7* levels in *LIN28B*, *LIN28B;i7s*, and control kidneys. Dox was provided from E14.5 until the end of the experiment. (*B*) Morphology of *LIN28B*, *LIN28B;i7s*, and control kidneys. (*C*) qRT–PCR analysis of the CM markers *Six2* and *Eya1* in *LIN28B*, *LIN28B;i7s*, and control kidneys. (*D*) H&E staining of *LIN28B*, *LIN28B;i7s*, and control kidneys. Note that the *Lin28B;i7s* kidneys have a normal histology. Bar, 100 μm.

et al. 2011). Accordingly, kidneys of compound *LIN28B-i7s* mice appeared smaller than normal kidneys (Fig. 6B). Importantly, however, there were no evidence of persistent CM cells in the *LIN28B-i7s* kidneys, as demonstrated by quantitative RT–PCR (qRT–PCR) (Fig. 6C) and histology (Fig. 6D). These results demonstrate that enforced *Let-7* expression can counteract the effect of *LIN28B* overexpression, suggesting that *LIN28B* induces Wilms tumor at least in part by suppressing *Let-7* miRNAs.

LIN28B expression in human Wilms tumor

Cancer-initiating cells (CICs) have recently been isolated from human Wilms tumor (Pode-Shakked et al. 2013). We detected overexpression of *LIN28B* (but not *LIN28A*) in $NCAM1^{+}ALDH1^{+}$ wild-type CICs isolated from early-generation Wilms tumor xenografts (harboring a blastemal predominant phenotype) and to a lesser extent in primary Wilms tumor when compared with developing human kidneys (Supplemental Fig. S8A). To determine whether *LIN28* expression is a prominent feature of human Wilms tumor, we analyzed the expression of *LIN28A* and *LIN28B* in human Wilms tumor samples from Boston Children's Hospital (USA, $n = 28$) and an independent set from Great Ormond Street Hospital/University College London Institute of Child Health (UK, $n = 77$). Indeed, immunohistochemical staining of these samples revealed overexpression of LIN28B (Fig. 7A) in eight out of 28 and 10 out of 77 samples, respectively, compared with normal kidneys. Conversely, LIN28A expression was not detected in any sample. This observation was further supported by our analysis of published microarray data that indicated frequent expression of *LIN28B* but not *LIN28A* in human Wilms tumor (Fig. 7B; Supplemental Fig. S8B).

In our tissue microarray analyses (UK samples), we noted that LIN28B expression was restricted to blastemal cells, which are the most undifferentiated tumor component. Clinical outcome data were available for 76 out of 77 UK patients who had been uniformly treated with prenephrectomy chemotherapy. Among the nine LIN28B-positive tumors, five patients relapsed, and three died. Among the 67 LIN28B-negative tumors, eight patients relapsed, and two died, thus suggesting a significant association of LIN28B expression with relapse and death ($P = 0.0059$ and $P = 0.0105$, respectively, two-tailed Fisher's test). Detailed histological and statistical analysis of the LIN28B-positive and LIN28B-negative tumors can be found in Supplemental Tables 1 and 2. Taken together, these data indicate that expression of *LIN28B* is a feature of a significant minority of cases of human Wilms tumor, which, together with our prior report of activation of *LIN28B* by chromosomal translocation in two cases, implicates *LIN28* in the pathogenesis of human Wilms tumor.

Discussion

It has been argued that pediatric cancer can arise from the failure of embryonic cells to complete their differentiation program (Maris and Denny 2002; Messahel et al. 2005), indicating that the pathogenesis of pediatric tumors is

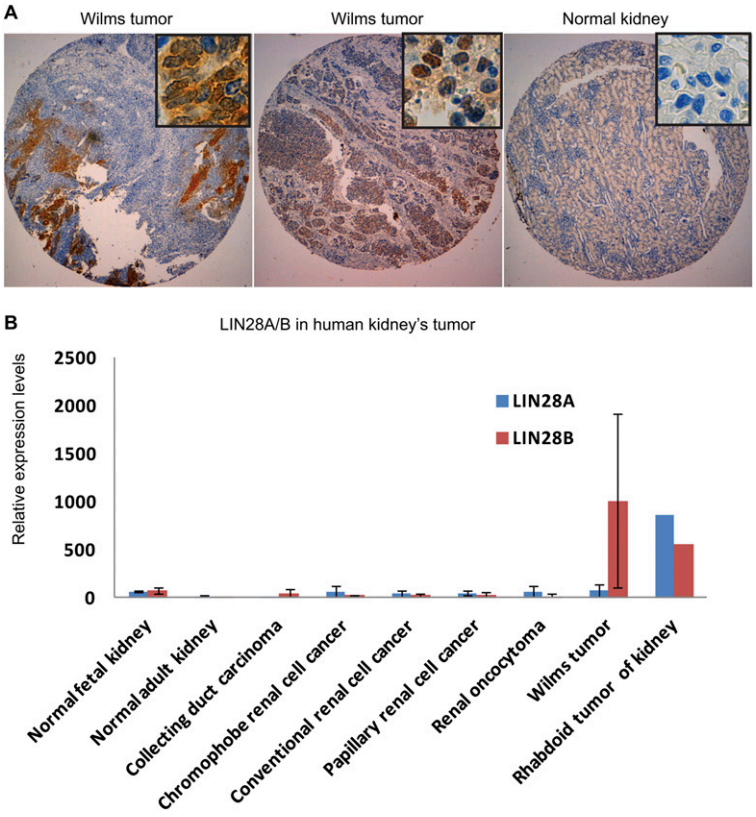


Figure 7. *LIN28B* expression in human Wilms tumor. (*A*) LIN28B immunohistochemistry in human Wilms tumor. (*B*) *LIN28B* and *LIN28A* expression levels in diverse types of human renal tumors based on published microarray data (GSE11151).

directly linked to dysregulated embryonic development and organogenesis (Scotting et al. 2005). Here we demonstrate that overexpression of *Lin28*, an RNA-binding protein linked to pluripotency, stem cell self-renewal, and delayed larval development in *C. elegans*, prolongs kidney development and promotes Wilms tumor formation. Reminiscent of its association with heterochronic phenotypes in *C. elegans*, *Lin28* overexpression prevents the synchronous wave of differentiation of CM cells in the developing kidney, which, under normal conditions, is complete by P2 in mice (Hartman et al. 2007) and by 36 wk of gestation in humans (Hinchliffe et al. 1991). In transgenic mice engineered for kidney-specific *Lin28* overexpression, CM cells continue to proliferate into adulthood, resulting in conversion to a tumor highly reminiscent of human Wilms tumor. When activation of *Lin28* was targeted to specific cellular compartments such as UB, stroma, or CM cells, we failed to observe tumor formation; instead, in our model, Cre excision in the intermediate mesoderm, the origin of the entire kidney, was required to induce tumor formation. Theoretically, this observation suggests that *Lin28* activation is required in either a stage-specific manner in the earliest kidney progenitor cells or more than one kidney cell lineage. The fact that tumors formed when *Lin28* overexpression was induced as late as E18.5, when the earliest progenitors no longer exist, suggests that the second explanation is more likely. Our model implies that aberrant *Lin28* expression produces a coordinated expansion of the nephrogenic zone, resulting in proliferating blastema and nephrogenic rests, which are characteristic of human Wilms tumor.

Previously, a murine model of Wilms tumor was generated by *Wt1* ablation and *Igf2* overexpression (Hu et al. 2011), which established that *Wt1* ablation prevents the MET of CM cells that is essential for nephrogenesis. In contrast, in our model, the CM cells persist beyond the period when synchronous differentiation typically occurs and retain their capacity to undergo MET, resulting in a markedly expanded period of nephrogenesis that ultimately progresses to frank tumorigenesis. The differences between these two models demonstrate that pediatric Wilms tumor formation can occur at diverse stages of development and through different molecular mechanisms.

Lin28 has profound effects on both the proliferative and metabolic machinery of tumor cells. By virtue of blocking *Let-7* biogenesis, *Lin28* leads to derepression of known oncogenic targets of *Let-7* such as *Myc* (Sampson et al. 2007), *Ras* (Johnson et al. 2005), *Hmga2* (Lee and Dutta 2007; Mayr et al. 2007), and cyclins (Legesse-Miller et al. 2009; Chang et al. 2012). Moreover, *Lin28* has been shown to promote glycolytic metabolism in tissues and cancer cells and thus is a central regulator of cellular bioenergetics (Shyh-Chang et al. 2013). It appears that *Lin28* functions to balance the proliferative and metabolic needs of rapidly growing cells in the early embryo (Shinoda et al. 2013b) and that this function becomes reactivated in many adult tumors. In cases of pediatric malignancy, *Lin28* appears to prolong the embryonic patterns of tissue growth (as we showed here for Wilms tumor), which is likely the case for germ cell tumors and neuroblastoma.

Point mutations in *WT1*, *WTX*, β-catenin, and abnormalities involving translocations in chromosome 6 have been linked to Wilms tumor (Knudson and Strong 1972; Rivera and Haber 2005; Hu et al. 2011), and several susceptibility loci for Wilms tumor were recently identified by a genome-wide association study (Turnbull et al. 2012). However, the underlying genetic basis of most cases of Wilms tumor remains unknown (Turnbull et al. 2012). Here we report that up to a third of human Wilms tumors overexpress *LIN28B*. Moreover, our data suggest an association of *LIN28B* expression with a "high-risk" subtype of Wilms tumor called "blastemal type" (Vujanic et al. 2002) that is defined by the persistence of a large proportion of blastemal cells in the viable tumor component after prenephrectomy chemotherapy. Therefore, *LIN28B* expression may be a marker of such therapy-resistant blastemal cells, which currently cannot be identified in chemo-naïve tumors treated according to the North American approach of immediate nephrectomy, but a much larger unselected cohort of tumors would be required to investigate this properly. Prior studies have implicated aberrant expression of the *HACE1* locus on chromosome 6 in Wilms tumor (Slade et al. 2010), but *LIN28B* is tightly linked to this locus, and our detection of aberrant overexpression of *LIN28B* in a significant minority of human Wilms tumors suggests the possibility of coordinate dysregulation. Previously, we reported that rare cases of Wilms tumor result from activation of *LIN28B* by chromosomal translocation and that amplification of the *LIN28B* locus occurs in only ~2% of tumors (Viswanathan et al. 2009); thus, the mechanism of *LIN28B* activation remains unexplained in most cases. Based on our demonstration that both murine *Lin28a* and human *LIN28B* are competent to induce Wilms tumor in our mouse model, we speculate that the prevalence of *LIN28B* activation and the absence of aberrant *LIN28A* expression in human Wilms tumor is likely due to a specific mechanism of *LIN28B* up-regulation and not differences in the transforming potential of the genes. In support of this hypothesis, it has been shown recently (Aiden et al. 2010) that the promoter of *LIN28A* in human Wilms tumors is enriched with histone K27me3.

The miRNA pathway is a common target for dysregulation in different types of tumors (Merritt et al. 2008; Bahubeshi et al. 2011). Recently, it was demonstrated that *DIS3L2*, the gene responsible for Perlman syndrome, which entails a predisposition to Wilms tumor, is a nuclease responsible for degrading *Let-7* miRNAs that have become polyuridylated due to *Lin28*-mediated recruitment of a terminal uridylyl transferase (Chang et al. 2013). Taken together with the observation of *LIN28B* overexpression in up to 30% of human Wilms tumors and our demonstration that enforced expression of *Let-7* abrogates *Lin28*-induced kidney tumorigenesis, these data suggest that a common mechanism of Wilms tumor pathophysiology is dysregulation of *LET-7* miRNA biogenesis or function. This hypothesis would be reinforced if indeed Drosha and Dicer1 mutations or other mutation in the miRNA processing machinery are likewise found in Wilms tumor. The *LIN28*/*LET-7* pathway represents an appealing therapeutic target

for Wilms tumor through either inhibition of *LIN28* function or delivery of *LET-7* to tumor cells.

Materials and methods

Mice

All animal procedures were conducted according to animal care guidelines approved by the Institutional Animal Care and Use Committee at Boston Children's Hospital.

LSL-Lin28a mice

A Lox-stop (four PGK-polyA and three sv40 polyA) Lox cassette and a Flag-tagged murine *Lin28a* ORF were cloned into pEF6 plasmid (Invitrogen, catalog no. V962-20) downstream from a PEF-1α promoter, and targeting was performed into V6.5 embryonic stem cells. Chimeric mice were generated by injection of embryonic stem cells into BALB/c blastocysts and then bred to CD-1 females to generate germline-transmitted pups.

Lox-TetOn-*Lin28* mice were previously generated in our laboratory as described in Zhu et al. (2010, 2011).

The following Cre mice were obtained from Jackson laboratory: *Vasa*Cre (stock no. 006954), *Six2*Cre (stock no. 009606), *FoxD1*Cre (stock no. 012463), and *Cdh16*Cre (stock no. 012237). The *Wt1*Cre mice were contributed by the laboratory of Dr. William Pu at Boston Children's Hospital. Cre mice were crossed to the Lox-TetOn-Lin28 mice, and 1 mg/mL Dox was administered to the drinking water at different time points to induce *LIN28B/Lin28a* expression. To achieve co-overexpression of *Lin28* and *Let7*, we crossed *Wt1*Cre mice with TRE-7S21L ("i7s" mice) mice and then crossed the *Wt1*Cre;i7s mice with Lox-TetOn-*Lin28* mice.

For the transplantation experiments, embryos were harvested by cesarean section at E18.5. Kidneys were harvested from transgenic and control embryos, dissected to smaller pieces, and then transplanted under the kidney capsules of NSG mice.

qRT–PCR

RNA was isolated by TRIzol and reverse-transcribed using SuperScript III (Invitrogen, catalog no. 18080-051) or miScript II RT kit (Qiagen, catalog no. 218161). mRNA expression was measured by qPCR using the ΔΔCT method with the following primers: m*Lin28a*: m*Six2* (forward primer, GCAAGTCAGCA ACTGGTTCA; reverse primer, CTTCTCATCCTCGGAACT GC), m*Eya1* (forward primer, TTTCCCTGGGACTACGAATG; reverse primer, GGAAAGCCATCTGTTCCAAA), m*Gapdh* (forward primer, GCAGTGGCAAAGTGGAGATTG; reverse primer, AATTTGCCGTGAGTGGAGTCATC), and m*bActin* (forward primer, TACTCCTGCTTGCTGATCCAC; reverse primer, CA GAAGGAGATTACTGCTCTGGCT); and h*LIN28B* (forward primer, GCCCCTTGGATATTCCAGTC; reverse primer, TGACT CAAGGCCTTTGGAAG). For qRT–PCR of mature *Let7* miRNA and pre/pri-*Let7*, we used Qiagen miScript target as described in the miScript protocol.

Microarray

RNA from four *Lin28a*-derived Wilms tumor samples and four control kidneys was harvested and processed using TRIzol. The Illumina Ref-8 microarray platform was used by the Boston Children's Hospital Intellectual and Developmental Disabilities Research Center (IDDRC) Molecular Genetics Core Facility. The microarray data have been deposited in Gene Expression Omnibus (GEO) and given the series accession number GSE56323. Gene set enrichment analysis (GSEA) was used to identify gene sets and pathways associated with a set of up-regulated or down-regulated genes. Published microarray data from GEO (GSE3822, GSE6890, and GSE12588) were used for the analysis of *Lin28a* and *Lin28b* during mouse kidney development. Data from Oncomine (https://www.oncomine.org) were used for the analysis of *LIN28A* and *LIN28B* expression in human Wilms tumor samples.

Histological analysis

Tissue samples were fixed in 10% buffered formalin and embedded in paraffin. Immunostaining was performed using the following antibodies: LIN28A (1:250; Cell Signaling, catalog no. 8641S), LIN28B (1:250; Cell Signaling, catalog no. 4196S), SIX2 (1:400; Proteintech Group, catalog no. 11562-1-AP), LTL (1:500; Vector Laboratories, catalog no. FL-1321), and Keratin 8 (1:50; Developmental Studies Hybridoma Bank [DSHB], catalog no. TROMA-I-s [TROMA-I-s]).

Slides were dewaxed with xylene and rehydrated through a series of washes with decreasing percentages of ethanol. Antigen retrieval was performed in 10 mM sodium citrate buffer (pH 6.0) by placement in a decloaking chamber for 30 min on high temperature.

Immunohistochemistry was performed with Elite ABC kit and DAB substrate (Vector Laboratories) according to the manufacturer's protocol. For immunofluorescence, Alexa 488- or Alexa568-conjugated goat anti-rabbit or goat anti-mouse secondary antibodies were used. Lin28B immunohistochemistry was scored as positive versus absent staining. In all cases, positive staining was seen in groups of blastemal cells.

To compare the nephron number between transgenic mice and controls, we count the number of glomerulus-like structures in 12 random fields (from the kidney cortex) under 10× magnification.

Statistical analysis

Data are presented as mean ± SEM, and Student's *t*-test (two-tailed distribution, two-sample unequal variance) was used to calculate *P*-values. Statistical significance is displayed as $P < 0.05$ (*) or $P < 0.01$ (**) unless specified otherwise. The tests were performed using Microsoft Excel, with the test type always set to two-sample equal variance.

Acknowledgments

We thank all of the clinicians at Boston Children's Hospital and the UK Children's Cancer and Leukaemia Group (CCLG) Centres who managed the care of the children entered into clinical and biology studies. We especially thank Dr. Richard Williams and Mr. Tasnim Chagtai (University College London [UCL] Institute of Child Health) for access to clinical and microarray expression data, and Dr. Sergey Popov (Institute of Cancer Research, University of London) and Professor Gordan Vujanic (University of Wales Medical School, Cardiff University) for assistance with construction of TMAs from patients treated in the SIOP WT 2001 trial in the UK. We thank the CCLG Tissue Bank for access to samples. G.Q.D. is an investigator of the Howard Hughes Medical Institute and the Manton Center for Orphan Disease Research and an affiliate member of the Broad Institute. This work was funded by the Ellison Medical Foundation and private funds of the Children's Hospital. K.P.-J. receives support from Cancer Research UK, Great Ormond Street Children's Charity, Children with Cancer Charity, and the National Institute for Health Research, Great Ormond Street Hospital, UCL Biomedical Research Centre award. The CCLG Tissue Bank is funded by Cancer Research UK and CCLG.

References

Aiden AP, Rivera MN, Rheinbay E, Ku M, Coffman EJ, Truong TT, Vargas SO, Lander ES, Haber DA, Bernstein BE. 2010. Wilms tumor chromatin profiles highlight stem cell properties and a renal developmental network. *Cell Stem Cell* **6:** 591–602.

Ambros V, Horvitz HR. 1984. Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* **226:** 409–416.

Bahubeshi A, Tischkowitz M, Foulkes WD. 2011. miRNA processing and human cancer: DICER1 cuts the mustard. *Sci Transl Med* **3:** 111ps146.

Chang HM, Martinez NJ, Thornton JE, Hagan JP, Nguyen KD, Gregory RI. 2012. Trim71 cooperates with microRNAs to repress Cdkn1a expression and promote embryonic stem cell proliferation. *Nature Commun* **3:** 923.

Chang HM, Triboulet R, Thornton JE, Gregory RI. 2013. A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28–let-7 pathway. *Nature* **497:** 244–248.

Davidson AJ. 2009. Mouse kidney development. In *StemBook*, (ed. The Stem Cell Research Community), doi: 10.3824/stembook.1.34.1, http://www.stembook.org.

Diskin S, Capasso M, Schnepp R, Cole K, Attiyeh E, Hou C, Diamond M, Carpenter E, Winter C, Lee H, et al. 2012. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. *Nat Genet* **445:** 1126–1130.

Gallardo T, Shirley L, John GB, Castrillon DH. 2007. Generation of a germ cell-specific mouse transgenic Cre line, Vasa-Cre. *Genesis* **45:** 413–417.

Gillis AJ, Stoop H, Biermann K, van Gurp RJ, Swartzman E, Cribbes S, Ferlinz A, Shannon M, Oosterhuis JW, Looijenga LH. 2011. Expression and interdependencies of pluripotency factors LIN28, OCT3/4, NANOG and SOX2 in human testicular germ cells and tumours of the testis. *Int J Androl* **34:** e160–e174.

Grobstein C. 1955. Inductive interaction in the development of the mouse metanephros. *J Exp Zool* **130:** 319–339.

Grobstein C. 1956. Trans-filter induction of tubules in mouse metanephrogenic mesenchyme. *Exp Cell Res* **10:** 424–440.

Hartman HA, Lai HL, Patterson LT. 2007. Cessation of renal morphogenesis in mice. *Dev Biol* **310:** 379–387.

Hatini V, Huh SO, Herzlinger D, Soares VC, Lai E. 1996. Essential role of stromal mesenchyme in kidney morphogenesis revealed by targeted disruption of Winged Helix transcription factor BF-2. *Genes Dev* **10:** 1467–1478.

Heo I, Joo C, Cho J, Ha M, Han J, Kim VN. 2008. Lin28 mediates the terminal uridylation of let-7 precursor microRNA. *Mol Cell* **32:** 276–284.

Hinchliffe SA, Sargent PH, Howard CV, Chan YF, van Velzen D. 1991. Human intrauterine renal growth expressed in absolute number of glomeruli assessed by the disector method and Cavalieri principle. *Lab Invest* **64:** 777–784.

Hu Q, Gao F, Tian W, Ruteshouser E, Wang Y, Lazar A, Stewart J, Strong L, Behringer R, Huff V. 2011. Wt1 ablation and Igf2 upregulation in mice result in Wilms tumors with elevated ERK1/2 phosphorylation. *J Clin Invest* **121:** 174–257.

Huff V. 2011. Wilms' tumours: about tumour suppressor genes, an oncogene and a chameleon gene. *Nat Rev Cancer* **11:** 111–132.

Humphreys BD, Lin SL, Kobayashi A, Hudson TE, Nowlin BT, Bonventre JV, Valerius MT, McMahon AP, Duffield JS. 2010. Fate tracing reveals the pericyte and not epithelial origin of myofibroblasts in kidney fibrosis. *Am J Pathol* **176:** 85–97.

Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ. 2005. RAS is regulated by the let-7 microRNA family. *Cell* **120:** 635–647.

Knudson AG Jr, Strong LC. 1972. Mutation and cancer: a model for Wilms' tumor of the kidney. *J Natl Cancer Inst* **48:** 313–324.

Kobayashi A, Valerius M, Mugford J, Carroll T, Self M, Oliver G, McMahon A. 2008. Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development. *Cell Stem Cell* **3:** 169–250.

Lee YS, Dutta A. 2007. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev* **21:** 1025–1030.

Legesse-Miller A, Elemento O, Pfau SJ, Forman JJ, Tavazoie S, Coller HA. 2009. let-7 Overexpression leads to an increased fraction of cells in G2/M, direct down-regulation of Cdc34, and stabilization of Wee1 kinase in primary fibroblasts. *J Biol Chem* **284:** 6605–6609.

Maris JM. 2010. Recent advances in neuroblastoma. *N Engl J Med* **362:** 2202–2211.

Maris JM, Denny CT. 2002. Focus on embryonal malignancies. *Cancer Cell* **2:** 447–450.

Mayr C, Hemann MT, Bartel DP. 2007. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* **315:** 1576–1579.

Merritt WM, Lin YG, Han LY, Kamat AA, Spannuth WA, Schmandt R, Urbauer D, Pennacchio LA, Cheng JF, Nick AM, et al. 2008. Dicer, Drosha, and outcomes in patients with ovarian cancer. *N Engl J Med* **359:** 2641–2650.

Messahel B, Hing S, Nash R, Jeffrey I, Pritchard-Jones K. 2005. Clinical features of molecular pathology of solid tumours in childhood. *Lancet Oncol* **6:** 421–430.

Molenaar J, Domingo-Fernández R, Ebus M, Lindner S, Koster J, Drabek K, Mestdagh P, van Sluis P, Valentijn L, van Nes J, et al. 2012. LIN28B induces neuroblastoma and enhances MYCN levels via let-7 suppression. *Nat Genet* **44:** 1199–1206.

Moss EG, Lee RC, Ambros V. 1997. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell* **88:** 637–646.

Murray MJ, Saini HK, Siegler CA, Hanning JE, Barker EM, van Dongen S, Ward DM, Raby KL, Groves IJ, Scarpini CG, et al. 2013. LIN28 Expression in malignant germ cell tumors downregulates let-7 and increases oncogene levels. *Cancer Res* **73:** 4872–4884.

Newman MA, Thomson JM, Hammond SM. 2008. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* **14:** 1539–1549.

Oosterhuis JW, Looijenga LH. 2005. Testicular germ-cell tumours in a broader perspective. *Nat Rev Cancer* **5:** 210–222.

Piskounova E, Viswanathan SR, Janas M, LaPierre RJ, Daley GQ, Sliz P, Gregory RI. 2008. Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J Biol Chem* **283:** 21310–21314.

Pleniceanu O, Harari-Steinberg O, Dekel B. 2010. Concise review: Kidney stem/progenitor cells: differentiate, sort out, or reprogram? *Stem Cells* **28:** 1649–1709.

Pode-Shakked N, Shukrun R, Mark-Danieli M, Tsvetkov P, Bahar S, Pri-Chen S, Goldstein RS, Rom-Gross E, Mor Y, Fridman E et al. 2013. The isolation and characterization of renal cancer initiating cells from human Wilms' tumour xenografts unveils new therapeutic targets. *EMBO Mol Med* **5:** 18–37.

Rivera M, Haber D. 2005. Wilms' tumour: connecting tumorigenesis and organ development in the kidney. *Nat Rev Cancer* **5:** 699–712.

Rumballe B, Georgas K, Combes A, Ju A, Gilbert T, Little M. 2011. Nephron formation adopts a novel spatial topology at cessation of nephrogenesis. *Dev Biol* **360:** 110–122.

Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, Wulczyn FG. 2008. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol* **10:** 987–993.

Sampson VB, Rong NH, Han J, Yang Q, Aris V, Soteropoulos P, Petrelli NJ, Dunn SP, Krueger LJ. 2007. MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. *Cancer Res* **67:** 9762–9770.

Scotting P, Walker D, Perilongo G. 2005. Childhood solid tumours: a developmental disorder. *Nat Rev Cancer* **5:** 481–488.

Shao X, Somlo S, Igarashi P. 2002. Epithelial-specific Cre/lox recombination in the developing kidney and genitourinary tract. *J Am Soc Nephrol* **13:** 1837–1846.

Shinoda G, De Soysa TY, Seligson MT, Yabuuchi A, Fujiwara Y, Huang PY, Hagan JP, Gregory RI, Moss EG, Daley GQ. 2013a. Lin28a regulates germ cell pool size and fertility. *Stem Cells* **31:** 1001–1009.

Shinoda G, Shyh-Chang N, Soysa TY, Zhu H, Seligson MT, Shah SP, Abo-Sido N, Yabuuchi A, Hagan JP, Gregory RI, et al. 2013b. Fetal deficiency of lin28 programs life-long aberrations in growth and glucose metabolism. *Stem Cells* **31:** 1563–1573.

Shyh-Chang N, Daley G. 2013. Lin28: primal regulator of growth and metabolism in stem cells. *Cell Stem Cell* **12:** 395–406.

Shyh-Chang N, Zhu H, Yvanka de Soysa T, Shinoda G, Seligson MT, Tsanov KM, Nguyen L, Asara JM, Cantley LC, Daley GQ. 2013. Lin28 enhances tissue repair by reprogramming cellular metabolism. *Cell* **155:** 778–792.

Slade I, Stephens P, Douglas J, Barker K, Stebbings L, Abbaszadeh F, Pritchard-Jones K, Cole R, Pizer B, Stiller C, et al. 2010. Constitutional translocation breakpoint mapping by genome-wide paired-end sequencing identifies HACE1 as a putative Wilms tumour susceptibility gene. *J Med Genet* **47:** 342–347.

Turnbull C, Perdeaux ER, Pernet D, Naranjo A, Renwick A, Seal S, Munoz-Xicola RM, Hanks S, Slade I, Zachariou A, et al. 2012. A genome-wide association study identifies susceptibility loci for Wilms tumor. *Nat Genet* **44:** 681–684.

Viswanathan SR, Daley GQ. 2010. Lin28: a microRNA regulator with a macro role. *Cell* **140:** 445–449.

Viswanathan SR, Daley GQ, Gregory RI. 2008. Selective blockade of microRNA processing by Lin28. *Science* **320:** 97–100.

Viswanathan SR, Powers JT, Einhorn W, Hoshida Y, Ng TL, Toffanin S, O'Sullivan M, Lu J, Phillips LA, Lockhart VL, et al. 2009. Lin28 promotes transformation and is associated with advanced human malignancies. *Nat Genet* **41:** 843–848.

Vujanic GM, Sandstedt B, Harms D, Kelsey A, Leuschner I, de Kraker J. 2002. Revised International Society of Paediatric Oncology (SIOP) working classification of renal tumors of childhood. *Med Pediatr Oncol* **38:** 79–82.

West JA, Viswanathan SR, Yabuuchi A, Cunniff K, Takeuchi A, Park IH, Sero JE, Zhu H, Perez-Atayde A, Frazier AL, et al. 2009. A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* **460:** 909–913.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, et al. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318:** 1917–1920.

Zhou B, Ma Q, Rajagopal S, Wu SM, Domian I, Rivera-Feliciano J, Jiang D, von Gise A, Ikeda S, Chien KR, et al. 2008. Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart. *Nature* **454:** 109–113.

Zhu H, Shah S, Shyh-Chang N, Shinoda G, Einhorn WS, Viswanathan SR, Takeuchi A, Grasemann C, Rinn JL, Lopez MF, et al. 2010. Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat Genet* **42:** 626–630.

Zhu H, Shyh-Chang N, Segrè AV, Shinoda G, Shah SP, Einhorn WS, Takeuchi A, Engreitz JM, Hagan JP, Kharas MG, et al. 2011. The Lin28/let-7 axis regulates glucose metabolism. *Cell* **147:** 81–94.

3: Table A2: General laboratory equipment used in this project

Equipment	Details	Supplier
Benchtop centrifuge	Eppendorf Microcentrifuge 5415D	Eppendorf AG, Hamburg, Germany
Eppendorfs	RNase-free: 0.2-1.5ml	Anachem Ltd, Bedfordshire, UK
Heat block	Techne Dri-Block DB.2A	Bibby Scientific France, Paris, France
Large centrifuge	Hettich, Rotina 46R	Gemini BV, Apeldoorn, The Netherlands
Light microscope	Invertoskop ID03	Carl Zeiss Microscopy GmbH, Gottingen, Germany
Microtome	Accu-cut SRM 200 Rotary Microtome	Sakura Finetek USA Inc, California, USA
Nanodrop	ND-1000	Thermo Fisher Scientific, Waltham, MA, USA
Nuclease free water	W4502	Sigma Aldrich, St Louis, MO, USA
PCR thermal cycler	Biorad Tetrad 2 Thermal Cycler	Life Science, Hertfordshire, UK
Pipette tips	0.1-1000µl	ELKAY, Basingstoke, UK
Pipettes	Discovery Comfort P2-P1000	HTL Lab Solutions, Warsaw, Poland
Qubit machine	Qubit 2.0 Fluorometer	Life Technologies Ltd, Paisley, UK
Vortex	Vortex Genie 2	Scientific Industries Inc, New York, USA
Water bath	Grant SUB6	Keison Products, Chelmsford, UK

4: Methylome Analysis Identifies a Wilms Tumour Epigenetic Biomarker Detectable in Blood. Jocelyn Charlton, Richard D. Williams, Mark Weeks, Neil J. Sebire, Sergey Popov, Gordan Vujanic, William Mifsud, Marisa Alcaide-German, Lee M. Butcher, Stephan Beck, Kathy Pritchard-Jones. Submitted to *Genome Biology*: 8th July 2014.

Genome Biology

BioMed Central
The Open Access Publisher

Methylome analysis identifies a Wilms tumor epigenetic biomarker detectable in blood

Genome Biology 2014, **15**:434 doi:10.1186/s13059-014-0434-y

Jocelyn Charlton (jocelyn.charlton.11@ucl.ac.uk)
Richard Williams (richard.williams@ucl.ac.uk)
Mark Weeks (m.weeks@ucl.ac.uk)
Neil Sebire (neil.sebire@ucl.ac.uk)
Sergey Popov (sergey.popov@icr.ac.uk)
Gordan Vujanic (vujanic@cardiff.ac.uk)
William Mifsud (w.mifsud@ucl.ac.uk)
Marisa Alcaide-German (m.alcaide-german@ucl.ac.uk)
Lee Butcher (l.butcher@ucl.ac.uk)
Stephan Beck (s.beck@ucl.ac.uk)
Kathy Pritchard-Jones (k.pritchard-jones@ucl.ac.uk)

Published online: 19 August 2014

ISSN 1465-6906

Article type Research

Submission date 8 July 2014

Acceptance date 8 August 2014

Article URL http://genomebiology.com/2014/15/8/434

Methylome analysis identifies a Wilms tumor epigenetic biomarker detectable in blood

Jocelyn Charlton[1]
Email: jocelyn.charlton.11@ucl.ac.uk

Richard D Williams[1]
Email: richard.williams@ucl.ac.uk

Mark Weeks[1]
Email: m.weeks@ucl.ac.uk

Neil J Sebire[1]
Email: neil.sebire@ucl.ac.uk

Sergey Popov[2]
Email: sergey.popov@icr.ac.uk

Gordan Vujanic[3]
Email: vujanic@cardiff.ac.uk

William Mifsud[1]
Email: w.mifsud@ucl.ac.uk

Marisa Alcaide-German[1]
Email: m.alcaide-german@ucl.ac.uk

Lee M Butcher[4]
Email: l.butcher@ucl.ac.uk

Stephan Beck[4*]
[*] Corresponding author
Email: s.beck@ucl.ac.uk

Kathy Pritchard-Jones[3*]
[*] Corresponding author
Email: k.pritchard-jones@ucl.ac.uk

[1] UCL Institute of Child Health, University College London, 30 Guilford Street, London WC1N 1EH, UK

[2] The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK

[3] Department of Pathology, Cardiff University School of Medicine, Heath Park, Cardiff CF14 4XN, UK

[4] UCL Cancer Institute, University College London, 72 Huntley Street, London WC1E 6BT, UK

Abstract

Background

Wilms tumor is the most common pediatric renal malignancy and there is a clinical need for a molecular biomarker to assess treatment response and predict relapse. The known mutated genes in this tumor type show low mutation frequencies, whereas aberrant methylation at 11p15 is by far the most common aberration. We therefore analyzed the epigenome, rather than the genome, to identify ubiquitous tumor-specific biomarkers.

Results

Methylome analysis of matched normal kidney and Wilms tumor identifies 309 preliminary methylation variable positions which we translate into three differentially methylated regions (DMR) for use as tumor-specific biomarkers. Using two novel algorithms we show that these three DMRs are not confounded by cell type composition. We further show that these DMRs are not methylated in embryonic blastema but are intermediately methylated in Wilms tumor precursor lesions. We validate the biomarker DMRs using two independent sample sets of normal kidney and Wilms tumor and seven Wilms tumor histological subtypes, achieving 100% and 98% correct classification, respectively. As proof-of-principle for clinical utility, we successfully use biomarker DMR-2 in a pilot analysis of cell-free circulating DNA to monitor tumor response during treatment in ten patients.

Conclusions

These findings define the most common methylated regions in Wilms tumor known to date which are not associated with their embryonic origin or precursor stage. We show that this tumor-specific methylated DNA is released into the blood circulation where it can be detected non-invasively showing potential for clinical utility.

Background

Wilms tumour (WT) is the most common paediatric renal cancer with a prevalence of one in 10,000 children [1]. In Europe, most patients receive four weeks of pre-operative chemotherapy prior to complete or partial nephrectomy, followed by tumour stage and histology-dependent post-operative treatment [2]. Although overall survival rates are good, there is a clinical need for a biomarker to evaluate patient response to chemotherapy and improve prediction of relapse.

Circulating cell-free DNA (cfDNA) isolated from blood has been used to assess tumour burden in other cancers [3–6]. In WT, the few genes that are recurrently mutated show low mutation frequencies; *WTX* (18%) [7], *CTNNB1* (15%) [8] and *WT1* (12%) [8] and do not account for the majority of WTs. However, epimutation affecting the *IGF2*/*H19* locus at 11p15.5 is much more common, (69%) [8]. Additional genes and regions known to be affected by methylation in WT include *GLIPR1* [9], imprinted genes *NNAT* [10] and the *WT1*-antisense region [11], various satellite regions [12,13], *HACE1* [14], *RASSF1A* [15], *P16* and the protocadherin cluster at 5q31 [16]. Consequently, we concluded that interrogation of the methylome rather than the genome may be more likely to reveal

ubiquitous tumour-specific biomarkers. Therefore, we performed genome-wide methylome analysis of matched WT and surrounding normal kidney (NK) to identify WT-specific sites of methylation which we then assessed in cfDNA for use as WT biomarkers.

Results

To identify tumour-specific methylation variable positions (MVPs), as previously defined [17], we derived methylation levels (β; 0 = unmethylated to 1 = methylated) for 462,537 CpG sites at single base-pair resolution using the Illumina Infinium HumanMethylation450 platform and performed linear modelling to compare 22 matched pairs of NK and WT (full clinical details in Additional file 1: Table S1). We identified 309 MVPs of genome-wide significance ($P < 5x10^{-8}$; Figure 1). Due to the matched study design there was no need to adjust for age, race or gender and we can exclude the possibility of MVPs being confounded by genetic polymorphism(s). We then applied the novel pipeline Lasso, recently developed for analysis of Illumina 450 k data, which considers the local CpG density to group MVPs into functionally more relevant differentially methylated regions (DMRs) [18]. Using this method, we identified 3 DMRs which were hypermethylated in WT with respect to NK (Table 1).

Figure 1 Identification of differentially methylated loci with genome-wide significance. Manhattan plot showing $-\log_{10}$ p-values for all CpGs (462,537) generated by linear modelling of normal kidney (NK) against Wilms tumour (WT). CpGs with genome-wide significance (n = 309) appear above the red line. DMRs 1 and 2 on chromosome 6 and DMR-3 on chromosome 11 are shown in green.

Table 1 Differentially methylated regions hypermethylated in WT compared to NK

DMR	Location	First CpG	Last CpG	Size (bp)	No. CpGs	DMR p-value	CpG island	Nearest gene
1	6p22.1	28956226	28956426	200	8	1.58E-10	CpG:42	ZNF311
2	6p21.32	32116905	32116963	58	3	4.67E-09	CpG:56	*PRRT1*
3	11q13.5	76858947	76859056	109	3	2.48E-09	CpG:38	*MYO7A*

To assess possible confounding due to differential cell type composition, we carried out two analyses. First, we conducted histological analysis confirming our samples to be composed of the expected major cell types, consisting of 95% epithelia and 5% stroma in NKs while WTs showed varying proportions of immature stroma, epithelia and blastema (Additional file 2: Figure S1). Second, we used a recently published algorithm, RefFreeEWAS [19] which corrects p-values based on estimated cell type contributions. Performing linear modelling using this algorithm, we identified 7,272 CpGs with genome-wide significance ($P < 5x10^{-8}$; Figure 2) of which 937 had $\Delta\beta > 0.3$ and were therefore considered cell-composition corrected MVPs. Of these, 766 were hyperMVPs and 171 were hypoMVPs in WT with respect to NK (Additional file 3: Table S2). For hyperMVPs in particular, we saw a striking positive enrichment for location within CpG islands (+18% compared to background). There were 483 CpG islands targeted by aberrant methylation in total (Additional file 4: Table S3) with a varying number of corrected MVPs per island. The greatest enrichment of MVPs occurred in two CpG islands on chr6 (CpG:56 and CpG:42) with 13 and 11 MVPs respectively which overlapped with DMRs 1 and 2. Using a threshold of ≥3 MVPs per DMR, we then mapped the cell-type corrected MVPs onto the DMRs identified with Lasso, confirming that all three DMRs were not confounded by cell composition effects. We therefore continued our analysis focussing on DMRs 1–3.

Figure 2 Methylation levels for DMRs 1–3 significantly distinguish normal kidney from Wilms tumour and whole blood. (a) Aggregated methylation levels across DMRs 1–3 significantly separate normal kidney (blue) and Wilms tumour (pink) in the test set (n = 22 pairs, $P = 3.85x10^{-17}$) and in **(b)** the replication set (n = 12 pairs, $P = 1.47x10^{-9}$). **(c)** Methylation levels were high in an independent dataset of WT including high risk (dark green, n = 25) and intermediate risk histological subtypes (light green, n = 61). High risk WT showed significantly higher methylation levels than intermediate risk WT (P = 0.0024). **(d)** Whole blood (n = 411) shows low methylation levels (red).

Aggregating DNA methylation across DMRs 1–3 (β_{mean}), we found that WT had significantly greater levels of methylation with respect to NK for the discovery dataset as well as an independent dataset of 12 pairs ($P = 3.85x10^{-17}$ and $P = 9.26x10^{-10}$ respectively, 2-tailed t-test, Figure 3a and b). Furthermore, β_{mean} was consistently high across an independent cohort of fresh frozen WT (n = 86; Additional file 2: Table S4) encompassing the seven post-chemotherapy WT histological subtypes classified into two risk groups, as defined by the International Society of Paediatric Oncology [20] (SIOP; Figure 3c). Within this cohort, a significant difference was seen between WT risk groups (P = 0.0024, 2-tailed T-test) with more elevated methylation levels observed in high risk WT (average β_{mean} = 0.87 vs. 0.78). Based on methylation levels in the discovery dataset, a support vector machine (SVM; using R package e1071) correctly classified 100% and 98% samples within each replication set respectively indicating the discriminative potential of these DMRs as biomarkers. Clinical details for both of these cohorts can be found in Additional file 1: Table S1.

Figure 3 Levels of methylation in serum cfDNA during WT treatment. Levels of $\%M_{mean}$ (for 14 CpGs within DMR-2) show significant differences between controls and pre-op samples (p = 0.010), controls and post-op samples (p = 0.017), pre-chemotherapy and pre-op (p = 0.028) and pre-chemotherapy and post-op (p = 0.047; all by 2-tailed T-test). Data shows increasing levels of methylated cfDNA associated with WT necrosis. Level of significance is shown by * (<0.05) or ** (<0.01) with horizontal lines indicating the respective comparison).

DMRs 1–2 are located within the extended major histocompatibility complex (MHC) region [21]. Although the MHC is highly polymorphic [22], our matched study design controlled for any genetic heterogeneity, ensuring that the observed signal was not confounded by copy number or other DNA sequence variation. We validated methylation levels for these DMRs using bisulfite-sequencing (Additional file 2: Figure S3 and Table S5) and confirmed the absence of C➔T mutation. MHC cluster hypermethylation and reciprocal loss of gene expression is common across cancers as a mechanism to evade immunosurveillance and increase oncogenic potential [23–26]. To further explore the association between DMR methylation and tumourigenesis, we extracted DNA from three specimens of human embryonic kidney (EK, gestational age = 22, 22 and 23 weeks) and separately microdissected embryonic blastema (EB, n = 3; the predicted WT cell-of-origin). Bisulfite sequencing of DMRs 1 and 2 showed average β-values of 0.007 for EK and 0.12 for EB. Furthermore, analysis of methylation levels in 20 cases with matched WT precursor lesions termed nephrogenic rests (NR) showed intermediate methylation levels (Additional file 2: Figure S2). Put together, this data suggests that sequential increase in methylation levels is associated with transformation of embryonic precursor cells towards a malignant phenotype.

As all DMRs were methylated in WT in comparison to NK, we predicted that levels of methylated DNA in the circulation may increase with tumour burden. When assessing levels of a methylated blood biomarker, varying proportions of leukocyte subpopulations can alter

the overall methylation signal giving false yield [27]. Therefore, to assess the potential influence that blood populations may have on our dataset, we examined publically available methylation signatures that define normal peripheral blood subgroups [28] as well as methylation levels of normal whole blood (extracted using Marmal-aid [29]; n = 411). We found that DMRs 1–3 did not overlap with any blood-related methylation signature, and that normal blood methylation levels for DMRs 1–3 were extremely low (average β_m = 0.12, Figure 3d). Therefore, we concluded that the WT-specific hypermethylated DMRs were not detected as a result of shifts in leukocyte populations in chemotherapy-treated tissue. Hence, we explored the potential of DMRs 1–3 as tumour-specific blood biomarkers, as they should be detectable above a low background and should not be confounded by shifts in leukocyte populations in the circulation.

To test the potential efficacy of these biomarkers, we performed bisulfite-sequencing of DMR-2 on cfDNA isolated from serum samples taken from 10 children with WT at diagnosis, during pre-operative chemotherapy and following nephrectomy as well as four cancer-free age-matched control serum samples (Additional file 2: Table S6). The entire sequenced region (chr6:32,116,940-32,117,259) spanned 319 bp and included 44 CpGs, many of which showed either no variation or extreme variation in methylation between samples. Therefore, to identify CpGs that showed consistent methylation at one time-point and differential methylation between time-points, we grouped samples within each time-point and performed an ANOVA test. From this, we selected a subset of CpGs (n = 14) that showed differential methylation between at least one pair of groups, irrespective of directionality of methylation changes, avoiding those with very low variance. Aggregating the percentage of methylated cfDNA across these 14 CpGs ($\%M_{mean}$) showed that control samples had the lowest levels of methylated cfDNA (13.4) and that children with WT had relatively higher levels taken at diagnosis before treatment starts (14.5). There was then a substantial and significant increase in the level of methylated cfDNA taken after the pre-operative chemotherapy phase (19.9) that persisted into the immediate post-operative period (19.2; Figure 3). This proof-of-principle experiment establishes DMR-2 as a potential blood-based biomarker for WT.

Discussion

This study has identified three differentially methylated regions (DMRs) with genome-wide significance in Wilms tumour (WT) that are present in nearly all WT (118 out of a total of 120 WTs examined) and are not mediated by cell type composition. In comparison to previous mutational studies, our epigenetic biomarker is far more ubiquitous and even supersedes 11p15 epimutation (in ~70% WT) as the most common biomarker in WT. Therefore, our results suggest that analysis of an epigenetic mark in cell-free DNA could be more successful than analysis of tumour-specific somatic mutations which are much less common. Furthermore, our study revealed a significant difference in methylation levels between intermediate and high risk WT which could be of diagnostic use and suggest a use for novel therapeutic approaches to high risk WT.

Although the NK samples do not contain blastema but WT samples do, our correction for cell type composition compensates for these effects allowing for identification of three true DMRs that are not mediated by cell mixture effects. Methylation analysis of DMRs 1–2 in embryonic kidney and embryonic blastema showed that these regions are not methylated in the presumed WT cell-of-origin. Analysis of nephrogenic rests showed that the precursor lesion had intermediate levels of methylation at the same regions which were then further

methylated in the associated tumours (Additional file 2: Figure S2). We therefore concluded that increased methylation at these loci is associated with tumourigenic transformation and is not simply a manifestation of the embryonic origin of these tumours. Furthermore, as 84/86 WTs were correctly classified by methylation status in our independent cohort, which included tumours stratified into all 7 histological subtypes, and consistent high levels of methylation were seen in the test set of tumours with detailed assessment of their variable cellular composition (Additional file 2: Figure S1), we can conclude that all cells constituting the tumour contributed to the observed gain of methylation.

Our data suggests that WT release methylated DMR-2 cfDNA (M-cfDNA) following exposure to chemotherapy. Supporting this, a higher proportion of M-cfDNA ($\%M_{mean}$ of 23.6 compared to 17.3) was observed in serum sampled after pre-operative chemotherapy in patients with the regressive subtype of WT, defined by >2/3 necrosis in the nephrectomy specimen, compared to patients with other WT subtypes that show less than 2/3 necrosis in response to chemotherapy [20]. In this small 'proof of principle' cases series, we noted that post-operative M-cfDNA levels remained high in the immediate post-operative period, sampled at day 4 up to 24 days post-surgery. This short term persistence may be due to the fact that serum cfDNA can retain interaction with nucleosome proteins which protects the DNA from degradation rendering it relatively stable [30]. This may also explain our ability to sequence such a long fragment (319 bp). A full post-surgery time course is required to assess this rate of degradation. Interestingly, one patient showed post-operative M-cfDNA levels 1.3-fold greater than the group mean; a 22% increase post-surgery. We hypothesise that this post-surgery increase in M-cfDNA may be due to residual tumour within the patient and indeed, three months later, bone metastasis was detected.

Conclusions

In conclusion, we have defined the first epigenetic biomarker for the analysis of circulating cfDNA in Wilms tumour patients. We show that this may be useful to improve the accuracy of determining tumour response during pre-operative chemotherapy and predicting the histological risk group. This could allow appropriate modification of treatment prior to planned nephrectomy, particularly important for surgical planning in bilateral WT, where maximising tumour response to allow partial nephrectomy is the goal. Due to the low relapse rate in WT, a much larger, prospectively collected sample series of patients is required to demonstrate clinical utility as a prognostic biomarker for relapse-free survival. Based on these 'proof of principle' findings, a European multi-centre clinical trial with appropriate sampling is planned to rigorously test whether analysis of this epigenetic biomarker would improve the accuracy of prediction of relapse in all cases.

Methods

Sample selection

Use of patient samples in this study was conducted with appropriate parental written consent and ethical approval granted by the NHS London Bridge Research Ethics Committee (ref: 12/LO/0101) with experiments performed in compliance with the Helsinki Declaration. Patients were enrolled in the UK either into the International Society of Paediatric Oncology (SIOP) Wilms Tumour 2001 Clinical Trial and Study or the Improving Patient Outcomes for Renal Tumours of Childhood (IMPORT) study, with appropriate parental written consent.

Blood serum samples from age-matched controls without cancer were taken from Great Ormond Street Hospital Department of Chemical Pathology with parental written consent. Marmal-aid v1.2.1 [31] was used to extract publically available methylation data annotated as disease = 'Healthy' and tissue = blood (n = 411). Embryonic kidney was obtained from fetal post-mortem examinations carried out at the Fetal Pathology Unit, University Hospital of Wales with written parental consent.

DNA extraction

As NR can only be identified by pathological review of haematoxylin and eosin (H&E) stained formalin-fixed paraffin embedded (FFPE) sections, 3 μm H&E sections from post-nephrectomy FFPE blocks were studied by two independent paediatric pathologists who marked out regions of normal kidney (NK) nephrogenic rest (NR) and Wilms tumour (WT). A total of 22 matched trios (NK, NR and WT) and 12 matched pairs of NK and WT (90 samples) were microdissected by cutting multiple 5 μm sections and removing the desired region with a scalpel. Embryonic blastema (EB) was also microdissected from 5 μm FFPE sections following a master H&E section as a guide. Tissue was taken from the whole section without microdissection to extract DNA from whole embryonic kidney (EK). Fresh frozen (FF) tissue (n = 86 samples) was taken from 83 patient nephrectomies that were classified according to centralised SIOP pathology review including stromal (n = 15), epithelial (n = 10), blastemal (n = 11), mixed (n = 23), diffuse anaplastic (n = 14), focal anaplasia (n = 2) or regressive (n = 11) type. DNA was extracted from FFPE and FF tissue using the DNeasy Blood and Tissue Kit (QIAGEN), however manufacturer's instructions were modified for FFPE DNA: samples were heated to 90 °C for 1 hour post-incubation at 56 °C and incubated at 70 °C for 10 minutes with buffer AL. For cfDNA analysis DNA was extracted from pre-chemotherapy (n = 5), pre-operative (n = 8) or post-operative (n = 8) patient serum and age-matched cancer-free control serum (n = 7) using the QIAamp Circulating Nucleic Acid kit (QIAGEN).

Genome-wide methylation analysis

DNA extracted from FFPE (n = 90) was first treated using the REPLIg FFPE kit (QIAGEN) [32]. Both FF (n = 86) and treated FFPE DNA was then bisulfite-converted using the EZ DNA Methylation Kit (Zymo Research) and interrogated using the Illumina 450 k platform. Two FFPE NR samples failed stringent quality control metrics and were excluded. For all analyses, open source software packages implemented in R [33] or Bioconductor [34] were used as indicated. Raw data was filtered to exclude samples with detection p-value <0.01 and normalised using subset within quantile normalisation (SWAN) using the Bioconductor R package ChAMP version 2.14 [18,35]. For initial MVP detection the normalised data matrix for 22 pairs (NK and WT) was included. Bayesian framework linear modelling using the Bioconductor R package Limma [36] version 3.20.4 [37] was performed to find sites of differential methylation that varied between NK and WT pairs and that were common across patients which avoids false positives from patient-specific SNPs or age effects. To this model, the TREAT function was applied to adjust p-values based on the Δβ-value (>0.1) [38] which were further adjusted to correct for multiple testing [39]. The DMR-lasso algorithm in the ChAMP package was then used to find differentially methylated regions (DMRs) [18] with settings adjusted to include only CpGs that reach genome-wide significance ($p < 5x10^{-8}$). To the same model, we applied the RefFreeEWAS algorithm [19] which uses single value decomposition to estimate the number of cell types contributing to overall histology. The algorithm then deconvoluted the β-values based on the estimated number of cell types (d

= 3) and a design matrix specifying patient pairs and sample histology, and generated bootstrap-derived CpG-specific p-values and covariates that correspond to a 'true' methylation signal with no cell mixture effects.

After confirming our three DMRs were not due to cell composition effects, the DMR values were compared to levels in a replication dataset of 12 pairs, the independent set of 86 FF WT and the matched NRs for 20 cases. Sample classification (by SVM) was performed using the 22 pairs as a training set and 12 pairs and 86 WT as separate test sets using R CRAN package e1071 [40].

Assessment of DMR methylation by bisulfite-sequencing

Of the three DMRs, we chose to validate methylation levels for DMRs 1–2 by sequencing both bisulfite-converted and normal DNA from 9 NK and WT pairs. We performed the same experiment to assess methylation levels in EK and EB. Bisulfite reads for DMR-1 showed poor coverage with only 6/18 validation samples giving sufficient reads (Additional file 2: Table S5). We therefore chose to focus on DMR-2 for detection in cfDNA from 28 serum samples. For cfDNA analysis, 5/21 patient (3 pre-op, 2 post-op) and 3/7 control samples failed to generate sufficient sequence reads for analysis.

Primers were designed (Additional file 2: Table S7) using Primer 3 [41] and MethPrimer [42] and optimised using commercial DNA. DNA for bisulfite-sequencing was converted using the EZ DNA methylation kit (Zymo Research). Library preparation PCRs were performed using NEBNext (New England Biolabs) and KAPA HiFi Uracil + (KAPA Biosystems). Products were cleaned using magnetic beads (Beckman Coulter) and quantified using Picogreen reagents. Sample-specific tags were added prior to sequencing using the Illumina Mi-Seq. Raw bisulfite-converted paired-end reads were mapped to human genome build hg19 with Bismark v0.9.0 [43] using Bowtie 2 [44] as the aligner. Methylated and unmethylated base counts were generated with the bismark_methylation_extractor utility and exported as BedGraph files for further analysis and display in IGV [45]. Aligned BAM files were sorted and indexed with SAMtools [46] for assessment of the regions of interest in IGV. The number of C reads divided by total reads per CpG site was then calculated to discern the % level of methylation per sample.

To generate allele counts for the full sequence of DMR-2, we used the ANGSD package [47]. ANOVA was performed in R using Bioconductor package Limma to make all possible contrasts between groups. CpGs were selected for further analysis if the Limma Toptable moderated F score >1 indicating that any of the contrasts between groups were non-zero and if group variance for that CpG was >1.

Data access

450 k methylation data described in this study are available from GEO [48] with accession ID GSE59157.

Abbreviations

cfDNA, cell-free DNA; DMR, Differentially methylated region; EB, Embryonic blastema; EK, Embryonic kidney; FFPE, Formalin fixed paraffin embedded; IMPORT, Improving

population outcomes for children with renal tumours; M-cfDNA, Methylated cell-free DNA; NK, Normal kidney; NR, Nephrogenic rest; PCR, Polymerase chain reaction; SIOP, International society of paediatric oncology; WT, Wilms tumour.

Competing interests

The authors declare that they have no competing financial interests.

Authors' contributions

J.C. performed DNA extraction from FFPE tissue and serum samples and carried out library preparation for bisulfite sequencing. N.J.S, S.P and W.M. marked out regions of NK, NR and WT for DNA extraction from FFPE tissue. G.V. and N.J.S are the National Reference Panel pathologists for the UK cases in the SIOP WT 2001 and IMPORT trials. G.V also supplied embryonic fetal kidney specimens. J.C. performed bioinformatics analyses with help from R.D.W. LMB provided analysis tools. M.A-G contributed data from a cohort of 86 fresh frozen tumours. MW co-ordinates the clinical study ("IMPORT") that collects serial blood samples of patients with WT during therapy. S.B. and K.P-J conceived the study. J.C. wrote the manuscript with input from all authors. All authors approved the final version.

Acknowledgements

J.C was funded by the UCL Grand Challenges Scheme and the Olivia Hodson fund. The Pritchard-Jones laboratory was funded by Cancer Research UK (C1188/A4614), Great Ormond Street Hospital (GOSH) Children's Charity and Children with Cancer (11MH16). NJS and KPJ are part supported by the NIHR GOSH UCL Biomedical Research Centre. The Beck laboratory was funded by the Wellcome Trust (99148), the UCL Biomedical Research Centre (BRC84/CN/SB/5984), a Royal Society Wolfson Research Merit Award (WM100023), IMI-JU OncoTrack (115234) and EU-FP7 projects IDEAL (259679), EPIGENESYS (257082) and BLUEPRINT (282510).
The authors would like to thank all the investigators at the contributing Children's Cancer and Leukaemia Group (CCLG) treatment centres and the CCLG tissue bank, which is funded by Cancer Research UK, for provision of samples. We thank Tasnim Chagtai and Becky West for processing of frozen tumour samples, Tessa Kasia, Lucy Bailey and Darren Edwards of the Great Ormond Street Hospital Translational Cancer Research Team for collection of tissue samples, the Department of Chemical Pathology at Great Ormond Street Hospital including Helen Aitkenhead, Principal Clinical Scientist, and Daley Aofolaju, Chief Biomedical Scientist for providing control serum samples. We thank Kerra Pearce and Tony Brooks from UCL Genomics for their help with processing the Illumina 450 k arrays and the Illumina sequencing.

References

1. Breslow N, Olshan A, Beckwith JB, Green DM: **Epidemiology of Wilms tumor.** *Med Pediatr Oncol* 1993, **21:**172.

2. Lemerle J, Voute PA, Tournade MF, Rodary C, Delemarre JF, Sarrazin D, Burgers JM, Sandstedt B, Mildenberger H, Carli M: **Effectiveness of preoperative chemotherapy in**

Wilms' tumor: results of an International Society of Paediatric Oncology (SIOP) clinical trial. *J Clin Oncol* 1983, **1:**604. published online EpubOctober 1, 1983

3. Gautschi O, Bigosch C, Huegli B, Jermann M, Marx A, Chassé E, Ratschiller D, Weder W, Joerger M, Betticher DC, Stahel RA, Ziegler A: **Circulating deoxyribonucleic acid as prognostic marker in non–small-cell lung cancer patients undergoing chemotherapy.** *J Clin Oncol* 2004, **22:**4157. published online EpubOctober 15, 2004 (10.1200/jco.2004.11.123).

4. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, Thornton K, Agrawal N, Sokoll L, Szabo SA, Kinzler KW, Vogelstein B, Diaz LA Jr: **Circulating mutant DNA to assess tumor dynamics.** *Nat Med* 2008, **14:**985. http://www.nature.com/nm/journal/v14/n9/suppinfo/nm.1789_S1.html.

5. Ramirez JL, Rosell R, Taron M, Sanchez-Ronco M, Alberola V, de las Peñas R, Sanchez JM, Moran T, Camps C, Massuti B, Sanchez JJ, Salazar F, Catot S: **14-3-3σ Methylation in pretreatment serum circulating DNA of cisplatin-plus-gemcitabine-treated advanced Non–small-cell lung cancer patients predicts survival: the Spanish Lung Cancer Group.** *J Clin Oncol* 2005, **23:**9105. published online EpubDecember 20, 2005 (10.1200/jco.2005.02.2905).

6. Mori T, O'Day SJ, Umetani N, Martinez SR, Kitago M, Koyanagi K, Kuo C, Takeshima T-L, Milford R, Wang H-J, Vu VD, Nguyen SL, Hoon DSB: **Predictive utility of circulating methylated DNA in serum of melanoma patients receiving biochemotherapy.** *J Clin Oncol* 2005, **23:**9351. published online EpubDecember 20, 2005 (10.1200/jco.2005.02.9876).

7. Ruteshouser EC, Robinson SM, Huff V: **Wilms tumor genetics: mutations in WT1, WTX, and CTNNB1 account for only about one-third of tumors.** *Genes Chromosomes Cancer* 2008, **47:**461. published online EpubJun (10.1002/gcc.20553).

8. Scott RH, Murray A, Baskcomb L, Turnbull C, Loveday C, Al-Saadi R, Williams R, Breatnach F, Gerrard M, Hale J, Kohler J, Lapunzina P, Levitt GA, Picton S, Pizer B, Ronghe MD, Traunecker H, Williams D, Kelsey A, Vujanic GM, Sebire NJ, Grundy P, Stiller CA, Pritchard-Jones K, Douglas J, Rahman N: **Stratification of Wilms tumor by genetic and epigenetic analysis.** *Oncotarget* 2012, **3:**327. published online EpubMar (.

9. Chilukamarri L, Hancock AL, Malik S, Zabkiewicz J, Baker JA, Greenhough A, Dallosso AR, Huang TH, Royer-Pokora B, Brown KW, Malik K: **Hypomethylation and aberrant expression of the glioma pathogenesis-related 1 gene in Wilms tumors.** *Neoplasia* 2007, **9:**970. published online EpubNov (.

10. Hubertus J, Zitzmann F, Trippel F, Muller-Hocker J, Stehr M, von Schweinitz D, Kappler R: **Selective methylation of CpGs at regulatory binding sites controls NNAT expression in Wilms tumors.** *PLoS One* 2013, **8:**e67605. 10.1371/journal.pone.0067605.

11. Malik K, Salpekar A, Hancock A, Moorwood K, Jackson S, Charles A, Brown KW: **Identification of differential methylation of the WT1 antisense regulatory region and relaxation of imprinting in Wilms' tumor.** *Cancer Res* 2000, **60:**2356. published online EpubMay 5, 2000 (.

12. Ehrlich M, Hopkins NE, Jiang G, Dome JS, Yu MC, Woods CB, Tomlinson GE, Chintagumpala M, Champagne M, Dillerg L, Parham DM, Sawyer J: **Satellite DNA hypomethylation in karyotyped Wilms tumors.** *Cancer Genet Cytogenet* 2003, **141:**97. published online EpubMar (.

13. Ludgate JL, Le Mee G, Fukuzawa R, Rodger EJ, Weeks RJ, Reeve AE, Morison IM: **Global demethylation in loss of imprinting subtype of wilms tumor.** *Genes Chromosomes Cancer* 2012. published online EpubOct 17 (10.1002/gcc.22017).

14. Zhang L, Anglesio MS, O'Sullivan M, Zhang F, Yang G, Sarao R, Mai PN, Cronin S, Hara H, Melnyk N, Li L, Wada T, Liu PP, Farrar J, Arceci RJ, Sorensen PH, Penninger JM: **The E3 ligase HACE1 is a critical chromosome 6q21 tumor suppressor involved in multiple cancers.** *Nat Med* 2007, **13:**1060. published online EpubSep (10.1038/nm1621).

15. Wagner KJ, Cooper WN, Grundy RG, Caldwell G, Jones C, Wadey RB, Morton D, Schofield PN, Reik W, Latif F, Maher ER: **Frequent RASSF1A tumour suppressor gene promoter methylation in Wilms' tumour and colorectal cancer.** *Oncogene* 2002, **21:**7277. published online EpubOct 17 (10.1038/sj.onc.1205922).

16. Dallosso AR, Hancock AL, Szemes M, Moorwood K, Chilukamarri L, Tsai HH, Sarkar A, Barasch J, Vuononvirta R, Jones C, Pritchard-Jones K, Royer-Pokora B, Lee SB, Owen C, Malik S, Feng Y, Frank M, Ward A, Brown KW, Malik K: **Frequent long-range epigenetic silencing of protocadherin gene clusters on chromosome 5q31 in Wilms' tumor.** *PLoS Genet* 2009, **5:**e1000745. published online EpubNov (10.1371/journal.pgen.1000745).

17. Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, Daunay A, Busato F, Mein CA, Manfras B, Dias KR, Bell CG, Tost J, Boehm BO, Beck S, Leslie RD: **Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis.** *PLoS Genet* 2011, **7:**e1002300. published online EpubSep (10.1371/journal.pgen.1002300).

18. Morris T, Butcher L, Feber A, Teschendorff A, Chakravarthy A, Wojdacz T, Beck S: **450k Chip Analysis Methylation Pipeline (ChAMP).** *Bioinformatics* 2013, published online EpubDecember 12, 2013 (10.1093/bioinformatics/btt684).

19. Houseman EA, Molitor J, Marsit CJ: **Reference-free cell mixture adjustments in analysis of DNA methylation data.** *Bioinformatics* 2014, **30:**1431. published online EpubMay 15 (10.1093/bioinformatics/btu029).

20. Vujanic GM, Sandstedt B, Harms D, Kelsey A, Leuschner I, de Kraker J: **Revised International Society of Paediatric Oncology (SIOP) working classification of renal tumors of childhood.** *Med Pediatr Oncol* 2002, **38:**79. published online EpubFeb.

21. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC Jr, Wright MW, Wain HM, Trowsdale J, Ziegler A, Beck S: **Gene map of the extended human MHC.** *Nat Rev Genet* 2004, **5:**889. published online EpubDec (10.1038/nrg1489).

22. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J,

Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD: **A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC.** *Nat Genet* 2006, **38:**1166. published online EpubOct (10.1038/ng1885).

23. Garrido C, Paco L, Romero I, Berruguilla E, Stefansky J, Collado A, Algarra I, Garrido F, Garcia-Lora AM: **MHC class I molecules act as tumor suppressor genes regulating the cell cycle gene expression, invasion, and intrinsic tumorigenicity of melanoma cells.** *Carcinogenesis* 2012. published online Epub January 4, 2012 (10.1093/carcin/bgr318).

24. Doyle A, Martin WJ, Funa K, Gazdar A, Carney D, Martin SE, Linnoila I, Cuttitta F, Mulshine J, Bunn P, Minna J: **Markedly decreased expression of class I histocompatibility antigens, protein, and mRNA in human small-cell lung cancer.** *J Exp Med* 1985, **161:**1135. published online EpubMay 1 (.

25. Jäger E, Ringhoffer M, Altmannsberger M, Arand M, Karbach J, Jäger D, Oesch F, Knuth A: **Immunoselection in vivo: Independent loss of MHC class I and melanocyte differentiation antigen expression in metastatic melanoma.** *Int J Cancer* 1997, **71:**142. 10.1002/(sici)1097-0215(19970410)71:2<142::aid-ijc3>3.0.co;2-0.

26. Degenhardt Y, Huang J, Greshock J, Horiates G, Nathanson K, Yang X, Herlyn M, Weber B: **Distinct MHC gene expression patterns during progression of melanoma.** *Gene Chromosome Cancer* 2010, **49:**144. published online EpubFeb (10.1002/gcc.20728).

27. Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, Smith AV, Gudnason V: **Heterogeneity in white blood cells Has potential to confound DNA methylation measurements.** *PLoS One* 2012, **7:**e46705. 10.1371/journal.pone.0046705).

28. Koestler DC, Marsit CJ, Christensen BC, Accomando W, Langevin SM, Houseman EA, Nelson HH, Karagas MR, Wiencke JK, Kelsey KT: **Peripheral blood immune cell methylation profiles are associated with nonhematopoietic cancers.** *Cancer Epidemiol Biomarkers Prev* 2012, **21:**1293. published online EpubAugust 1, 2012 (10.1158/1055-9965.epi-12-0361).

29. Lowe R, Rakyan VK: **Marmal-aid--a database for Infinium HumanMethylation450.** *BMC Bioinformatics* 2013, **14:**359. 10.1186/1471-2105-14-359.

30. Rumore PM, Steinman CR: **Endogenous circulating DNA in systemic lupus erythematosus. Occurrence as multimeric complexes bound to histone.** *J Clin Investig* 1990, **86:**69. published online EpubJul (10.1172/JCI114716).

31. **Marmal-aid website.** http://marmal-aid.org/.

32. Thirlwell C, Eymard M, Feber A, Teschendorff A, Pearce K, Lechner M, Widschwendter M, Beck S: **Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip.** *Methods* 2010, **52:**248. published online EpubNov (10.1016/j.ymeth.2010.04.012).

33. **R: A language and environment for statistical computing.** www.r-project.org/.

34. **Bioconductor website.** www.bioconductor.org/.

35. **Bioconductor ChAMP package website.**
http://www.bioconductor.org/packages/release/bioc/html/ChAMP.html.

36. Smyth GK: **Limma**. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Volume 23. Edited by Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S. New York: Springer; 2005:397–420.

37. **Bioconductor Limma package website.**
http://www.bioconductor.org/packages/release/bioc/html/limma.html.

38. McCarthy DJ, Smyth GK: **Testing significance relative to a fold-change threshold is a TREAT.** *Bioinformatics* 2009, **25:**765. published online EpubMarch 15, 2009 (10.1093/bioinformatics/btp053).

39. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Roy Stat Soc B (Methodological)* 1995, **57:**589. 10.2307/2346101.

40. **e1071 package website.** http://cran.r-project.org/web/packages/e1071/index.html.

41. Koressaar T, Remm M: **Enhancements and modifications of primer design program Primer3.** *Bioinformatics* 2007, **23:**1289–1291.

42. Li LC, Dahiya R: **MethPrimer: designing primers for methylation PCRs.** *Bioinformatics* 2002, **18:**1427–1431.

43. **Bismark website.** http://www.bioinformatics.babraham.ac.uk/projects/bismark/.

44. **Bowtie 2 website.** http://bowtie-bio.sourceforge.net/bowtie2/.

45. Robinson JT, Thorvaldsdottir K, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer.** *Nat Biotechnol* 2011, **29:**24–26.

46. **SAMtools website.** http://samtools.sourceforge.net/.

47. **ANGSD website.** http://popgen.dk/wiki/index.php/ANGSD.

48. **Gene Expression Omnibus (GEO) website.** http://www.ncbi.nlm.nih.gov/geo/.

Additional files

Additional_file_1 as XLSX

Additional file 1: Table S1 Clinical information for the discovery cohort (n=22), validation set 1 (n=12) and validation set 2 (n=86).

Additional_file_2 as DOCX

Additional file 2 Figure S1: Quantification of cell proportions in each micro-dissected

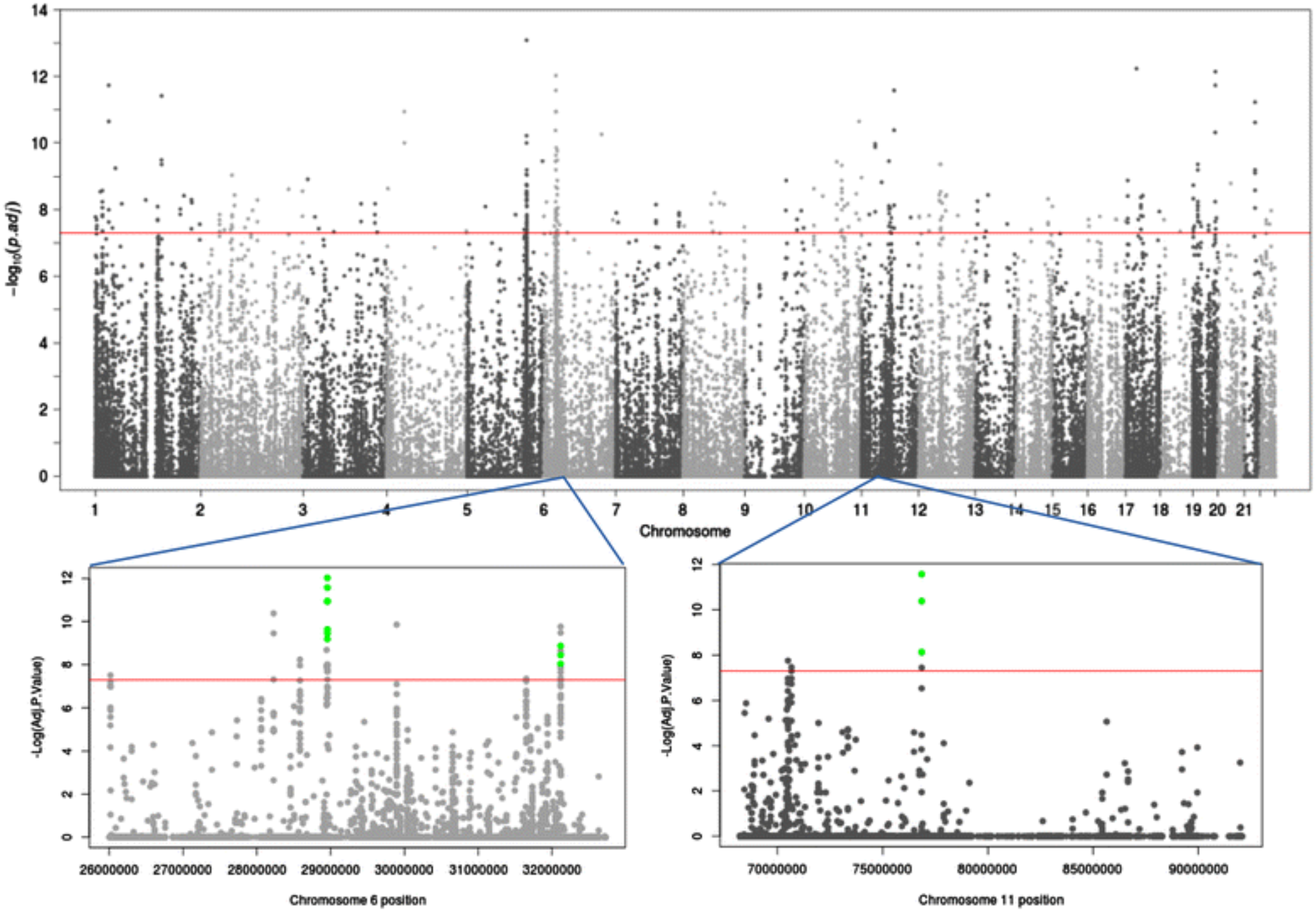
Wilms tumour section used for DNA extraction in the discovery cohort. **Figure S2:** WT precursor lesions show intermediate methylation at significant DMRs. **Figure S3:** Comparison of methylation values assessed by 450k array and bisulfite sequencing. **Table S4:** Fresh frozen WT (n = 86) classified by overall tumour histology with average methylation β-values across all significant DMR CpGs. **Table S5:** Validation of 450k methylation signal by bisulfite-sequencing. **Table S6:** Clinical information on patients from which cfDNA was isolated. **Table S7:** List of primers used.

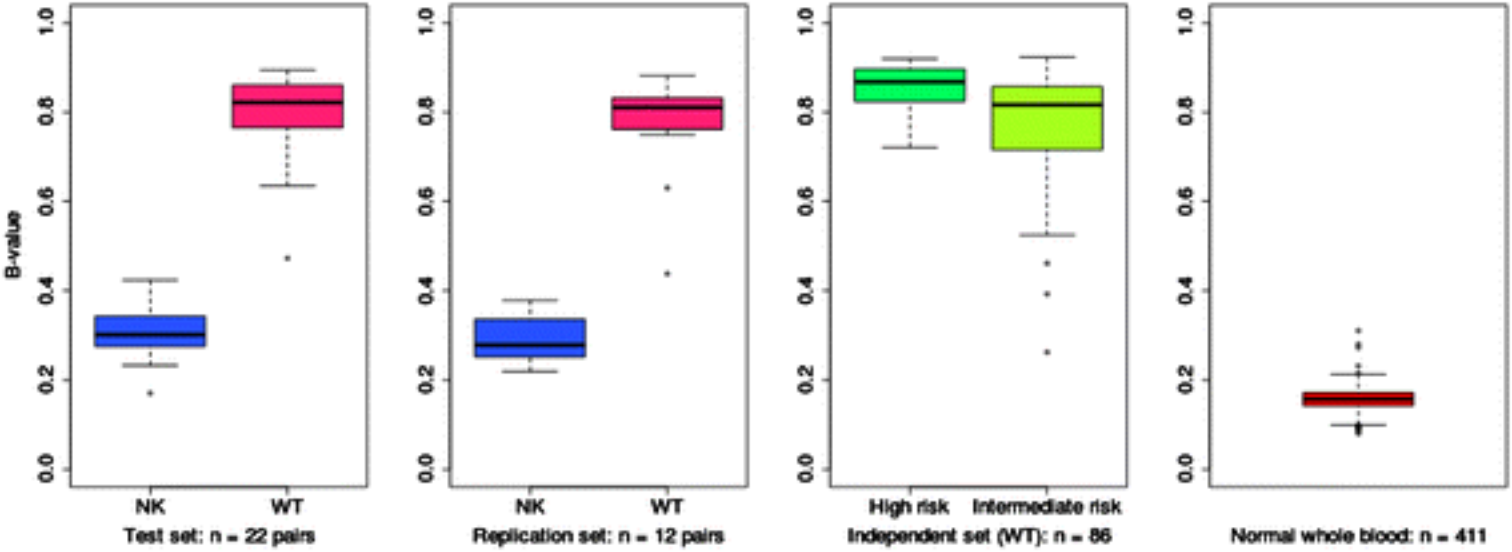
Additional_file_3 as XLSX

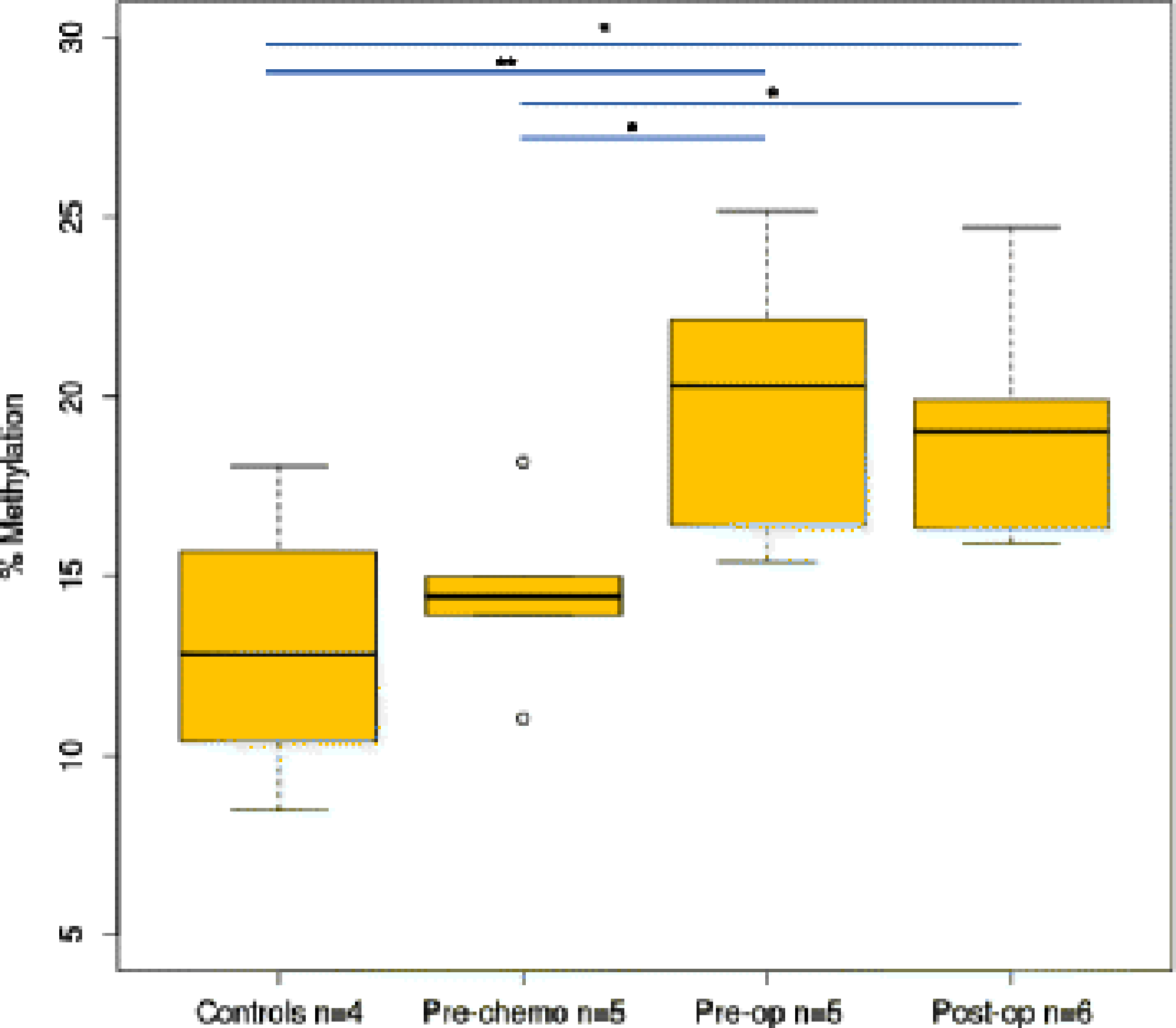
Additional file 3: Table S2 A list of all cell-corrected methylation variable positions (n=937).

Additional_file_4 as XLSX

Additional file 4: Table S3 A list of the CpG islands (n=483) that contained significantly hypermethylated or hypomethylated cell-corrected MVPs, with the respective number of CpGs per island.


−log10(p.adj)
Chromosome
-Log(Adj.P.Value)
Chromosome 6 position
Chromosome 11 position


B-value
0.0
0.2
0.4
0.6
0.8
1.0
NK
WT
Test set: n = 22 pairs
NK
WT
Replication set: n = 12 pairs
High risk
Intermediate risk
Independent set (WT): n = 86
Normal whole blood: n = 411


% Methylation
30
25
20
15
10
5
*
**
*
*
Controls n=4
Pre-chemo n=5
Pre-op n=6
Post-op n=6