

Northumbria Research Link

Citation: Iliasova, Alice (2020) Investigation of the molecular and clinical heterogeneity of medulloblastoma subgroups. Doctoral thesis, Northumbria University.

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/49626/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

Investigation of the molecular and clinical heterogeneity of medulloblastoma subgroups

Alice Iliasova

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

Research undertaken in the Faculty of Health and Life Sciences
Northumbria University
in collaboration with
Northern Institute for Cancer Research, Newcastle University

July 2020

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others. The work was done part-time over 5 years in collaboration with Newcastle University Brain Tumour Research Group, Translational and Clinical Research Institute, Newcastle University (former Northern Institute for Cancer Research, Newcastle University). Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the Newcastle / North Tyneside Research Ethics Committee study reference 07/Q0905/71. I declare that the Word Count of this Thesis is 49, 817 words.

Alice Iliasova

July 2020

Acknowledgements

I would like to thank my principal supervisor Dr Ed Schwalbe for his support and endless patience in guiding me through this project, and my second supervisor Professor Stephen Todryk. Also I would like to thank all members of Newcastle University Brain Tumour Research Group, collaboration with whom made this project possible and in particular Dr Debbie Hicks, Dr Dan Williamson and Professor Simon Bailey for valuable advice and Dr Louise Pease for proofreading the early drafts of this thesis. Special thanks go to Dr Sirintra Nakjang of Newcastle University Bioinformatics Support Unit who provided matching genotyping array profiles for validation of the methods developed in this project, proofreading and advice. I am grateful to Northumbria University Research and Development Fund for providing me with the studentship that made this project possible.

Dedication

This work is dedicated to inspirational little warrior Grace Elizabeth Frazer who beat medulloblastoma not once but twice and to all sufferers and survivors of childhood cancers; to my parents and friends who made sure I finished this journey relatively sane, and my daughter Sophie, my most loyal and patient supporter.

Abstract

Background Medulloblastoma is the most common paediatric malignant brain tumour, with heterogeneous clinico-molecular characteristics and survival outcomes. Current WHO classification distinguishes 3 molecular subgroups: WNT, SHH (named after characteristic activation of the WNT/wingless and Sonic Hedgehog signalling pathways) and non-WNT/non-SHH medulloblastoma. Current risk stratification incorporates molecular and clinico-pathological disease correlates. Survival is associated with high-risk factors such as metastasis, large cell/anaplastic histology, *MYC/MYCN* amplification and subtotal resection; the presence of one or more of these factors defines high-risk disease. High-risk patients receive more intensive therapies at the cost of severe late effects as survivors. Moreover, 20% of standard-risk patients (defined by absence of all high-risk features) will die of their disease. Tumour profiling with genome-wide DNA methylation arrays is the current gold standard for molecular classification of brain tumours. Methylation arrays are also suitable for identifying DNA copy number (CN) changes, enabling simultaneous genomic and epigenomic characterisation.

Hypothesis It was hypothesised that genome-wide Illumina HumanMethylation arrays provide a robust alternative to gold-standard SNP arrays for DNA CN detection and allow for single-platform, integrated genetic and epigenetic assessment, suitable for application to DNA derived from fresh-frozen and formalin-fixed, paraffin embedded tumour materials

Aims

- Confirm usability of methylation arrays and develop methods to detect genomic alterations (aneuploidy and focal oncogene amplifications)
- Showcase application of methylation arrays as a cost-effective single-platform, integrated approach for improved prognostication within medulloblastoma patients.

Methods In this project, methods to detect genomic alterations (aneuploidy and focal oncogene amplifications) using Illumina 450k methylation arrays were developed and validated. These methods were implemented to assess previously published cytogenetic prognostication schemes in medulloblastoma. Next, the GLMnet algorithm was used to identify prognostic methylation loci. These markers were assessed in non-WNT/non-SHH high-risk medulloblastomas and validated in an independent, mixed-risk non-WNT/non-SHH cohort. The previously published cytogenetic prognostic signature for standard-risk, non-WNT/non-SHH medulloblastoma, identified in the PNET4 clinical trial, and its potential for prognostication was assessed in high-risk, non-WNT/non-SHH disease, alone and in conjunction with methylation markers

Results Easy to use methods with a low barrier to entry were developed to robustly identify genomic copy number and oncogene amplification. These methods were applied to validate two independent, previously published cytogenetic prognostication schemes within medulloblastoma. Two DNA methylation loci, mapping to *MYO7A* and *TRIM72* genes, were identified as independently prognostic markers. A novel prognostication scheme, that combined DNA methylation markers with the PNET4 cytogenetic signature, was devised for non-WNT/non-SHH medulloblastoma. This scheme outperformed the PNET4 signature in the high-risk cohort, reclassifying 21% of high-risk patients to a favourable-risk category.

Conclusion These results demonstrate the potential for routine cytogenetic assessment concurrent with molecular sub-classification using DNA methylation microarrays. Additionally, the integrated genetic and epigenetic stratification from a single platform enabled a more refined prognostication and the identification of a subset of patients, currently classified as high risk, who demonstrate improved outcomes and who may be eligible for reduced intensity treatments that would offer a better quality of life as brain tumour survivors.

List of abbreviations

3' UTR	Three prime untranslated region
5' UTR	Five prime untranslated region
A	Adenine
AATK	Apopotosis-associated tyrosine kinase
APC	Adenomatous polyposis coli
ASCL2	Achaete-scute complex homologue 2 (Drosophila)
AT/RT	Atypical teratoid/rhabdoid tumour
ATP	Adenosine triphosphate
B2M	Beta-2-microglobulin
BCHE	Butyrylcholinesterase
BCL2	B-cell CLL / lymphoma 2
BCR-ABL	Breakpoint cluster region c-abl oncogene 1, non-receptor tyrosine kinase
BMI1	BMI polycomb ring finger oncogene
BP	Base pair
BRAF	v-raf murine sarcoma viral oncogene homologue B1
BRCA1	Breast cancer 1, early onset
BRCA2	Breast cancer 2, early onset
BLAST	Basic Local Alignment Search Tool
C	Cytosine
CDH1	Cadherin 1, type 1, E-cadherin (epithelial)
CDH13	Cadherin 13, H cadherin (heart)
CDKN2A	Cyclin dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)
cDNA	Complementary DNA
CGH	Comparative genomic hybridisation
CI	Confidence interval
CIMP	CpG island methylator phenotype
CLAS	Classic type medulloblastoma
CNGA3	Cyclic nucleotide gated channel alpha 3
CNN	Copy number neutral
ChIP	Chromatin immunoprecipitation
CNS	Central nervous system
CpG	Cytosine-guanine dinucleotide

cRNA	Complementary RNA
CRUK	Cancer Research UK
CSF	Cerebrospinal fluid
CT	Computed tomography
CTCF	CCCTC-binding factor (zinc finger protein)
CTTNB1	Catenin (cadherin-associated protein), beta 1
DAPK1	Death-associated protein kinase 1
DCC	Deleted in colorectal carcinoma
DHH	Desert hedgehog
DKK1	Dickkopf homologue 1 (<i>Xenopus laevis</i>)
DKK2	Dickkopf homologue 2 (<i>Xenopus laevis</i>)
DN	Desmoplastic/nodular medulloblastoma
DNA	Deoxyribose nucleic acid
DNMT	DNA methyltransferase
dsDNA	Double-stranded DNA
ECM	Extracellular matrix
EFS	Event free survival
EGL	External granule layer
ERBB2	v-erb-b2 erythroblastic leukaemia viral oncogene homologue2, neuroma / glioblastoma
ESC	Embryonic stem cell
ESR1	Oestrogen receptor 1
EZH2	Enhancer of Zeste Homolog 2
FC	Fold change
FDA	Food and Drug Administration
FDR	False discovery rate
FF	Fresh Frozen
FFPE	Formalin-Fixed Paraffin Embedded
FISH	Fluorescent in situ hybridisation
FP	False positive
G	Guanine
GAB1	GRB2-associated binding protein 1
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase

gDNA	Genomic DNA
GLI1	GLI family zinc finger 1
GLI2	GLI family zinc finger 2
GLI3	GLI family zinc finger 3
GNP	Granule neuron precursor
GRM1	Glutamate receptor, metabotropic 1
GRM8	Glutamate receptor, metabotropic 8
GSTP1	Glutathione-S-transferase pi 1
GTP	Guanine triphosphate
HAT	Histone acetyl transferase
HDAC	Histone deacetylase
HFE	Hemochromatosis
HGF	Hepatocyte growth factor (hepapoietin A; scatter factor)
Hh	Hedgehog
HIC1	Hypermethylated in cancer 1
HL	Hodgkin's lymphoma
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene homologue
IDH1	Isocitrate dehydrogenase 1
IHH	Indian hedgehog
IL8	Interleukin 8
ITIH2	Inter-alpha (globulin) inhibitor H2
KCDT11	Potassium channel tetramerisation domain containing 11
KCNA1	Potassium voltage-gated channel, shaker-related subfamily, member 1
KRAS	v-ki-ras2 Kirsten rat sarcoma viral oncogene homologue
LCA	Large cell / anaplastic medulloblastoma
LDHB	Lactate dehydrogenase B
LEF	Lymphoid enhancer factor
LINE-1	Long interspersed nuclear element 1
MAD	MAX dimerisation protein
MATH1	Atonal homologue 1 (Drosophila)
MAX	MYC associated factor X
MB	Medulloblastoma

MBEN	Medulloblastoma with extensive nodularity
MBP	Methyl-cytosine binding protein
MDS	Myelodysplastic syndrome
MeDIP	Methylated DNA immunoprecipitation
MET	Met proto-oncogene (hepatocyte growth factor receptor)
MGMT	O-6-methylguanine-DNA methyltransferase
MICAL1	Microtubule associated monooxygenase, calponin and LIM domain containing 1
MLH1	MutL homologue 1, colon cancer, nonpolyposis type 2 (E. coli)
MLL2	Myeloid / lymphoid or mixed –lineage leukaemia 2
MLL3	Myeloid / lymphoid or mixed –lineage leukaemia 3
MRI	Magnetic Resonance Imaging
mRNA	Messenger RNA
MSH2	MutS homologue 2, colon cancer, nonpolyposis type 1 (E. coli)
MSP	Methylation specific PCR
MXI1	MAX interactor 1
MYC	v-myc myelocytomatosis viral oncogene homologue (avian)
MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)
MYO7A	unconventional myosin family VII member A
NC	Not classifiable
NGS	Next generation sequencing
NHL	Non-Hodgkin's lymphoma
NMB	Newcastle medulloblastoma
NMF	Non-negative matrix factorisation
NOS2A	Nitric oxide synthase 2, inducible
NPR3	Natriuretic peptide receptor C
OS	Overall survival
OTX2	Orthodenticle homeobox 2
PALB2	Partner and localiser of BRCA2
PCA	Principal component analysis
PcG	Polycomb group proteins
PCR	Polymerase chain reaction
PFS	Progression free survival

PNET3	Primitive-neuroectodermal tumour 3 trial
PNET4	Primitive-neuroectodermal tumour 4 trial
PNET5	Primitive-neuroectodermal tumour 5 trial
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PTCH1	Patched 1
PTCH2	Patched 2
PTEN	Phosphatase and tensin homologue
QC	Quality control
RAB33A	RAB33A, member RAS oncogene family
RB1	Retinoblastoma 1
REN	Renin
RNA	Ribonucleic acid
ROC	Receiver Operating Characteristic
rRNA	Ribosomal RNA
RT	Reverse transcriptase
SD	Standard Deviation
SFRP1	secreted frizzled-related protein 1
SHH	Sonic hedgehog
SIOP	International Society of Paediatric Oncology
SMARCA1	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 1
SMARCA4	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4
SMO	Smoothened homologue (Drosophila)
SNHL	Sensorineural hearing loss
SNP	Single nucleotide polymorphism
SOCS1	Suppressor of cytokine signalling 1
SPDEF	SAM pointed domain containing ets transcription factor
SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homologue (avian)
SUFU	Suppressor of fused homologue (Drosophila)
T	Thymine
TCF	T cell factor

TET	Ten-eleven translocation
TF	Transcription factor
TFPI2	Tissue factor pathway inhibitor 2
TIMP3	TIMP metallopeptidase inhibitor 3
TNC	Tenascin C
TP	True positive
TP53	Tumour protein 53
TRIM72	Tripartite motif containing 72
TSG	Tumour suppressor gene
TSP-1	Thrombospondin-1
TSS	Transcriptional start site
U	Uracil
UV	Ultra-violet
VEGF	Vascular endothelial growth factor
VEGFA	Vascular endothelial growth factor A
WGSBS	Whole genome shotgun bisulfite sequencing
WHO	World Health Organisation
WIF1	WNT inhibitory factor 1
WNT	Wingless
WT1	Wilms' tumour 1
ZIC2	Zic family member 2 (odd-pairing homologue, Drosophila)

Contents

Declaration	i
Acknowledgements	ii
Dedication	iii
Abstract	iv
List of abbreviations	vi
List of Figures	xviii
List of Tables	xx
Chapter 1. Introduction	1
1.1. Cancer	1
1.1.1. Cancer incidence	2
1.1.2. Childhood cancer	2
1.1.3. Genetic model of cancer	5
1.1.4. Cancer development	5
1.1.4.1. Sustaining Proliferative Signalling	6
1.1.4.2. Evading Growth Suppressors	7
1.1.4.3. Resisting Cell Death	7
1.1.4.4. Enabling replicative immortality	7
1.1.4.5. Inducing angiogenesis	8
1.1.4.6. Activating invasion and metastasis	8
1.1.4.7. Genome Instability and Mutation	9
1.1.4.8. Tumour-Promoting Inflammation	9
1.1.4.9. Reprogramming Energy Metabolism	10
1.1.4.10. Evading Immune Destruction	10
1.1.5. Genetic origin of cancer	10
1.1.6. Oncogenes	11
1.1.6.1. Activation of proto-oncogenes	11
1.1.7. Tumour suppressor genes	12
1.1.7.1. Inactivation of tumour suppressor genes	13
1.1.8. Copy number variations and cancer	13
1.1.9. Epigenetics and cancer	14
1.1.9.1. DNA methylation in cancer	15
1.1.9.2. Histone Modifications and chromatin remodelling in cancer cells	16
1.1.9.3. MicroRNA and Long Non-Coding RNAs dysregulation	18
1.1.10. Relationship between genetic and epigenetic factors in cancer development	18
1.2. Paediatric tumours of the central nervous system	20
1.2.1. Embryonal tumours of the Central Nervous System	20
1.3. Medulloblastoma	22
1.3.1. Histology	22
1.3.1.1. Classic medulloblastoma	22

1.3.1.2.	Desmoplastic/nodular medulloblastoma and medulloblastomas with extensive nodularity	22
1.3.1.3.	Large cell/anaplastic medulloblastoma	23
1.3.2.	Diagnosis and clinical presentation of medulloblastoma	23
1.3.3.	Medulloblastoma staging	25
1.3.4.	Genetics and molecular classification of medulloblastoma	26
1.3.4.1.	Familial syndromes	26
1.3.5.	Embryonic signaling pathways and their role in cerebellar development	28
1.3.5.1.	WNT pathway: role in cerebellar development and medulloblastoma	28
1.3.5.2.	SHH pathway: role in cerebellar development and medulloblastoma	29
1.3.6.	Subgroups of medulloblastoma: Current consensus	31
1.3.7.	Heterogeneity within medulloblastoma subgroups	35
1.3.8.	Origins of medulloblastoma	38
1.3.9.	Prognostic factors and current risk stratification in clinical settings	39
1.3.9.1.	High-risk factors and standard risk disease	40
1.3.10.	Risk Stratification within subgroups	41
1.3.11.	Epigenetic drivers in medulloblastoma	42
1.3.11.1.	DNA methylation in medulloblastoma	42
1.3.11.2.	Histone modifications in medulloblastoma	43
1.3.11.3.	Chromatin remodelling in medulloblastoma	46
1.3.11.4.	MicroRNAs and Long Non-Coding RNAs in medulloblastoma	47
1.3.12.	Treatment of medulloblastoma	47
1.3.13.	Treatment <i>sequelae</i>	48
1.3.14.	Survival and quality of life dilemma: potential of improving patients outcome	50
1.3.15.	Hypothesis and aims	52
Chapter 2.	Materials and methods	53
2.1.	Cohorts	53
2.2.	Microarrays	55
2.2.1.	Affymetrix SNP6.0 arrays	55
2.2.2.	Methylation arrays	56
2.3.	Methylation microarrays data pre-processing and quality control	59
2.3.1.	Initial quality control: identification of unusual samples and technical artefacts	59
2.3.1.1.	Filtering	60
2.3.1.2.	Normalisation and batch correction	61
2.3.2.	Minfi	61
2.4.	Sample quality control: Derivative Log Ratio Spread	63
2.5.	Circular binary segmentation	64
2.6.	Copy Number analysis	65
2.7.	Conumee	66
2.8.	High-throughput sequencing: RNA-seq	68
2.9.	Survival analysis in medulloblastoma	68
2.9.1.	Univariable analysis: Kaplan-Meier plots and log-rank tests	70
2.9.2.	Multivariable analysis: Cox proportional hazards models	71
2.9.3.	Median follow-up time	71
2.10.	Confusion matrices	72
2.11.	Over-fitting	73
2.12.	Cross-validation	73
2.13.	Selection of testable, potentially prognostic methylation probes	74
2.14.	Bimodality	74

2.15.	GLMnet	74
2.15.0.1.	Ridge regression	74
2.15.0.2.	LASSO	75
2.15.0.3.	Elastic Net	75
2.16.	Concordance index	75
2.17.	Statistical Analysis	77
Chapter 3.	Development and implementation of the workflow and methods for copy number analysis using Illumina Human Methylation arrays	78
3.1.	Introduction	78
3.2.	Aims	81
3.3.	Materials and methods	81
3.3.1.	Cohorts	81
3.3.2.	Overview of copy number analysis workflow	82
3.3.3.	Input data preparation	82
3.3.4.	Pipeline specific sample quality control	85
3.3.5.	CN analysis methods	85
3.3.5.1.	Methods for automatic assessment of CNA on chromosomal arm-level	86
3.3.5.2.	Methods for assessing focal amplification	86
3.3.6.	CNA profiles assembly methods	87
3.3.6.1.	Large-scale CN profiles	88
3.3.6.2.	Locus-specific CN profiles	89
3.3.6.3.	Profile type IV: full combined CN profiles	89
3.3.7.	CNA analysis pipelines developed in this project	90
3.3.7.1.	Pipeline for automated calling of large scale chromosomal alterations	90
3.3.7.2.	Pipeline for automated calling of focal, locus-specific alterations	92
3.3.8.	Performance and suitability evaluation metrics	93
3.4.	Results	93
3.4.1.	Pipeline specific sample quality control	93
3.4.2.	Evaluation of suitability and performance of methodology for automated calling of large scale chromosomal alterations	94
3.5.	Evaluation of semi-automated calling of focal, locus-specific alterations	99
3.6.	Discussion	102
3.6.1.	Development of the bioinformatics workflow and methods: rationale and implementation	102
3.6.1.1.	Bioinformatics pipeline for large-scale copy number detection confidently identifies arm-level CNAs	104
3.6.1.2.	The focal, locus-specific method is able to detect high-level amplifications	105
3.6.2.	Illumina Human Methylation array is a robust and accurate platform for DNA copy number estimation	106
3.6.3.	Future work	106
3.6.4.	Summary	108
Chapter 4.	Application of the developed methodology to validation of existing prognostication schemes and determination of the utility of PNET4 stratification scheme for prognostication of patients with unavailable clinical risk factors	109
4.1.	Introduction	109
4.1.1.	Shih risk stratification scheme	111
4.1.2.	PNET4 risk stratification scheme	111
4.2.	Aims	113
4.3.	Materials and methods	114
4.3.1.	Shih scheme biomarkers validation	114

4.3.1.1.	Discovery cohort	114
4.3.1.2.	Validation cohort.	114
4.3.1.3.	Copy-number profiling	114
4.3.2.	PNET4 scheme biomarkers validation and application to high-risk disease cohort	117
4.3.2.1.	Discovery cohort	117
4.3.2.2.	Validation cohort	117
4.3.2.3.	High risk cohort	117
4.3.2.4.	Procedures	119
4.3.2.5.	Copy-number profiling	119
4.3.3.	Assessment of missing information patterns within Newcastle Medulloblastoma (NMB) cohort	119
4.3.4.	Statistical analyses	120
4.4.	Results	121
4.4.1.	Assessment of missing information pattern within NMB cohort	121
4.4.2.	Shih biomarkers validation	121
4.4.2.1.	Cohort characteristics	121
4.4.2.2.	Stratification of patients with SHH tumours and their survival characteristics	122
4.4.2.3.	Identification of high-risk patients with Group 3 medulloblastoma was confirmed by metastatic status, iso17q and <i>MYC</i> amplification	123
4.4.2.4.	Low-risk group of patients with metastatic Group 4 medulloblastoma was identified	123
4.4.3.	Validation of biomarkers discovered in PNET4 cohort in Newcastle Medulloblastoma cohort	128
4.4.3.1.	Cohort characteristics	128
4.4.3.2.	Validation of previously reported high-risk clinico-pathological markers	128
4.4.3.3.	Validation of PNET4 prognostication scheme: standard-risk disease	129
4.4.3.4.	Application of PNET4 prognostication scheme to high-risk disease	130
4.4.4.	Comparison of performance of PNET4 risk stratification scheme: standard-risk and high-risk cohorts	131
4.5.	Discussion	136
4.5.1.	Copy number analysis of cohorts processed using methylation arrays validates previously reported survival markers and classification schemes	136
4.5.2.	Application of CN analysis workflow and methods to standard-risk Group 3 and Group 4 medulloblastoma cohort validates previously reported risk biomarkers and prognostication scheme	137
4.5.3.	Application of the PNET4 prognostication scheme to high-risk medulloblastoma cohort shows its prognostic relevance to risk-stratification for disease with known risk factors	137
4.5.4.	Clinical relevance of PNET4 prognostication scheme within high-risk medulloblastoma	138
4.5.5.	Future work	138
4.5.6.	Summary	139
Chapter 5.	Standard risk medulloblastoma: methylomic analysis	140
5.1.	Introduction	140
5.2.	Aims	141
5.3.	Materials and methods	142
5.3.1.	NMB Standard Risk non-WNT/non-SHH cohort	142

5.3.2.	NMB High Risk non-WNT/non-SHH cohort and available clinico-pathological correlates	142
5.3.3.	Age-matched, independent non-WNT/non-SHH cohort with limited clinical risk information	142
5.3.3.1.	Validation of known clinico-pathological covariates	142
5.3.4.	Identification of potentially prognostic methylation markers	144
5.3.5.	GLM-net to identify significantly differentially methylated CpG sites within standard risk non-WNT/non-SHH medulloblastoma	144
5.3.6.	Additional prognostic covariates and definition of improved standard risk non-WNT/non-SHH medulloblastoma stratification scheme	144
5.3.7.	Investigation of the prognostic potential of the novel risk stratification methylation markers in high risk non-WNT/non-SHH medulloblastoma	145
5.3.8.	Additional prognostic covariates in high risk non-WNT/non-SHH medulloblastoma	145
5.3.9.	Validation of the identified methylation markers in an independent cohort with only some clinical risk factors available	145
5.4.	Results	146
5.4.1.	Validation of the previously reported clinico-pathological correlates and cytogenetic markers in the high-risk non-WNT/non-SHH medulloblastoma cohort	146
5.4.2.	Validation of the previously reported clinico-pathological correlates and cytogenetic markers in the age-matched external cohort with uncertain risk status	146
5.4.3.	Identification of significant methylation markers in the standard-risk non-WNT/non-SHH medulloblastoma cohort	146
5.4.4.	Assessment of prognostic potential of identified methylation probes and selection of the cut-off value for classification of methylated and unmethylated loci	147
5.4.5.	Methylation pattern of the selected probes	149
5.4.6.	Prognostic methylation biomarkers identified lower-risk patients within standard risk non-WNT/non-SHH medulloblastoma	149
5.4.7.	Integration of previously reported survival markers with identified methylomic correlates	152
5.4.8.	Investigation of prognostic potential of identified methylomic markers in high-risk non-WNT/non-SHH medulloblastoma and integration of the additional prognostic covariates	152
5.4.8.1.	Metastatic stage as a prognostic covariate	153
5.4.9.	Validation of the identified methylomic biomarkers in the external cohort with limited risk status information	155
5.4.10.	Formulation of extended testable survival model for risk stratification of patients with non-WNT/non-SHH medulloblastoma and its application to cohorts of variable risk status	157
5.4.11.	Relationship between methylation of identified loci and their gene expression	163
5.5.	Summary and discussion	166
5.5.1.	Mixed-risk survival cohort with some unavailable risk-factors to mimic the real world clinical settings	166
5.5.2.	The methylomic signature, consisting of both prognostic markers, identified favourable risk group in standard risk non-WNT/non-SHH medulloblastoma and is prognostic in conjunction with the PNET4 signature scheme	166

5.5.3. Prognostic potential of the methylomic markers in non-metastatic high-risk and uncertain risk non-WNT/non-SHH medulloblastoma	167
5.5.4. Gene function of identified survival methylation biomarkers	168
5.5.4.1. <i>MYO7A</i> is associated with cancerous growth and metastasis	168
5.5.4.2. <i>TRIM72</i>	169
5.5.4.3. Potential function of methylation of identified loci	170
5.5.5. Future work	170
5.5.6. Summary	171
Chapter 6. Chapter 6: Discussion and conclusion	172
6.1. Introduction	172
6.2. CN analysis workflow has potential for integrated array analysis	175
6.2.0.1. Advantages and limitations	175
6.3. Whole chromosomal aberration signature has utility for rapid disease risk stratification in clinical settings	176
6.4. Methylomic signature has prognostic potential in non-WNT/non-SHH medulloblastoma with limited availability of known clinical risk markers	177
6.5. Risk stratification in non-WNT/non-SHH disease	178
6.6. Future work	180
6.6.1. Conclusion	182
Appendix A. Cohorts	184
Appendix B. Publications	234
Appendix. References	259

List of Figures

1.1	Childhood cancers by cancer type	3
1.2	Twenty most common cancers	3
1.3	Childhood Cancers by Cancer Type	5
1.4	Hallmarks of cancer	6
1.5	Epigenetic mechanisms	15
1.6	Summary of the major changes in the WHO 2016 classification for the diagnosis of central nervous system (CNS) embryonal tumours	21
1.7	Histopathological subtypes of medulloblastoma	24
1.8	Anatomy of the brain	25
1.9	Medulloblastoma, T1-weighted sagittal MRI scan	26
1.10	Scheme of active WNT pathway	28
1.11	Scheme of active SHH pathway	31
1.12	Comparison of medulloblastoma subgroups - consensus 2012	32
1.13	Summary of molecular subtypes of Group 3/4 medulloblastoma	38
1.14	Epigenetic regulators	44
1.15	Role of EZH2 and KDM6A/UTX in maintaining stem cell state	45
2.1	Single nucleotide polymorphism arrays	56
2.2	Infinium Methylation Assay scheme	58
2.3	Minfi preprocessing flow chart	63
2.4	Conumee flow chart	67
2.5	Kaplan-Meier plot	70
2.6	Concordance statistic	76
3.1	Summary of copy number analysis workflow	83
3.2	Summary of automatic assessment of copy number changes	87
3.3	Summary of copy number analysis workflow	88
3.4	Copy number analysis pipelines	91
3.5	Pipeline specific quality control	94
3.6	Results of grid search for gain and loss cut-offs	95
3.7	Balanced chromosome 17	96
3.8	Isochromosome 17	97
3.9	Chromosome 17 p-arm loss	97
3.10	Chromosome 17 q-arm gain	98
3.11	Whole chromosome 17 gain	98
3.12	Conumee chromosomal arm-level calls	99
3.13	Focal high-level amplification of <i>MYCN</i> oncogene	101
4.1	Shih prognostication scheme	111
4.2	Shih prognostication scheme	112
4.3	Missing data pattern within the initial Standard Risk cohort	122
4.4	Shih prognostication scheme validation: SHH group	124
4.5	Shih prognostication scheme validation: Group 3	125
4.6	Shih prognostication scheme validation: Group 4	127
4.7	Continues on next page	130

4.7	Event-free survival in the high-risk cohort of non-WNT/non-SHH medulloblastoma by clinical and disease-associated molecular features	131
4.8	Original PNET4 prognostication scheme survival	132
4.9	PNET4 prognostication scheme validation: standard-risk disease	133
4.10	PNET4 prognostication scheme application to high-risk disease	134
5.1	Kaplan-Meier plots of PNET4 signature and metastatic status in Cavalli dataset	147
5.2	Methylation pattern of the identified prognostic probes	150
5.3	Distribution of β -values of the identified prognostic probes	150
5.4	Kaplan-Meier plots of prognostic methylation markers in standard risk non-WNT/non-SHH medulloblastoma	151
5.5	The number of prognostic methylation markers in relationship to survival in the standard risk non-WNT/non-SHH cohort	151
5.6	The number of prognostic methylation markers combined with PNET4 scheme signature in relationship to survival in the standard risk non-WNT/non-SHH cohort	152
5.7	Kaplan-Meier plots of identified prognostic methylation markers tested in high risk non-WNT/non-SHH cohort	154
5.8	The number of prognostic methylation markers in relationship to survival in the high risk non-WNT/non-SHH cohort	154
5.9	The number of prognostic methylation markers in relationship to survival in the high risk non-WNT/non-SHH cohort with metastatic status as a covariate	155
5.10	Kaplan-Meier plots of identified prognostic methylation markers tested in in mixed-risk external cohort	156
5.11	The number of prognostic methylation markers in relationship to survival in mixed-risk external cohort	157
5.12	The number of prognostic methylation markers in relationship to survival in the uncertain risk non-WNT/non-SHH cohort with metastatic status as a covariate	158
5.13	Decision tree for classification of non-WNT/non-SHH survival cohorts	160
5.14	DNA methylation markers add prognostic value to PNET4 signature prognostication scheme in the high risk non-WNT/non-SHH cohort	161
5.15	DNA methylation markers add prognostic value to PNET4 signature prognostication scheme in the mixed risk external cohort	163
5.16	Relationship between methylation of identified loci and their gene expression	165

List of Tables

1.1	Classification of the oncogenes	12
1.2	Chang’s medulloblastoma staging	26
1.3	Familial cancer syndromes	27
1.4	Clinical and genomic characteristics of medulloblastoma subgroups	35
1.5	Cells of origin of medulloblastoma	39
1.6	Currently accepted prognostic factors in medulloblastoma	40
1.7	Clinical risk stratification of medulloblastoma	41
1.8	Proposed risk stratification in childhood non-infant medulloblastoma	42
1.9	Summary of genes encoding epigenetic regulators in medulloblastoma	42
2.1	Genomic probe distribution	55
2.2	Illumina HumanMethylation450k control probes	60
2.3	An example Cox model	72
2.4	Table of confusion	73
3.1	Rules for whole chromosome status assessment from statuses of individual chromosomal arm	89
3.2	Optimal cut-offs for median intensities	95
3.3	Confusion matrix describing performance of CNA analysis workflow for Illumina Methylation arrays	95
3.4	Performance of the pipeline for automated calling of large scale chromosomal abnormalities	96
3.5	Table of confusion describing performance of CNA analysis locus-specific pipeline	100
3.6	Performance of the pipeline for automated calling of large scale chromosomal abnormalities	100
3.7	Updated minfi functions supporting multiple versions of Illumina HumanMethylation arrays (Fortin et al., 2017).	108
4.1	Patient characteristics of the discovery (Shih et al. (2014)) and validation (NMB) cohorts	116
4.2	Clinical and molecular characteristics of non-WNT/non-SHH HIT-SIOP PNET4 and NMB cohorts cohorts	118
4.3	Relationship between Group 3 and 4 subtypes described by Northcott et al. and cytogenetic risk scheme.	132
4.4	Relationship between Group3/4 subtypes described by Schwalbe et al. and cytogenetic risk scheme	133
4.5	PNET4 prognostication scheme standard risk <i>vs</i> high risk disease comparison table	135
5.1	Comparison of clinical and molecular characteristics of non-WNT/non-SHH (Cavalli et al., 2017) and NMB cohorts. SR - s	143
5.2	Assessment of prognostic methylation markers as continuous variables in standard risk cohort	148
5.3	Assessment of prognostic methylation markers as binary variables in standard risk cohort	149

5.4	Assessment of prognostic methylation markers as binary variables in standard risk cohort	153
5.5	Relationship between Group 3 and 4 subtypes described by Northcott et al. and combined methylomic/PNET4 scheme scheme	162
5.6	Relationship between Group 3 and 4 subtypes described by Schwalbe et al. (2017b) and combined methylomic/PNET4 scheme scheme	162
5.7	Combined methylomic/PNET4 prognostication scheme comparison table for 5 years survival in non-WNT/non-SHH medulloblastoma	162
A.1	Clinical demographic of the arm-level pipeline test cohort	189
A.2	Clinical demographic of high-level amplification test cohort	196
A.3	Shih biomarkers validation cohort	217
A.4	Clinical demographics of standard-risk NMB cohort	220
A.5	Clinical demographics of high-risk NMB cohort	224
A.6	Clinical demographics of the Cavalli dataset used for chapter 4. Cohort consisted of 244 primary medulloblastoma tumours from Cavalli et al. (2017) study. Molecular subgroup, gender, age, metastatic stage and pathology are shown. Histological subtype is coded as follows: LCA- large cell/anaplastic; Desmoplastic- desmoplastic/nodular; MBEN- MB with extensive nodularity	233

CHAPTER 1

Introduction

1.1. Cancer

Cancer is a group of diseases characterised by abnormal cell growth with a potential for invasion of nearby tissue and spread to other parts of the body. Cancer can originate from any cell type, anywhere in the body. Many cancers form solid masses called tumours but some, like blood cancers do not form tumours. Abnormal growth of body tissue defines a tumour. Tumours can be classified as benign, pre-malignant and malignant.

Benign tumours lack the ability to invade nearby tissue or spread and therefore are classed as not cancerous. When removed, benign tumours do not usually regrow and are therefore not considered dangerous. However, they can be quite large and, in some cases, for example benign brain tumours, can be life threatening.

Pre-malignant tumours tend to eventually turn malignant, meaning they can invade into nearby tissue. Malignant tumours are known as cancers.

Tumour formed where the cancer started is called a primary tumour. Sometimes cancer cells can detach and travel through the blood or lymph system into distant parts of the body, where they form secondary tumours, called metastases. Metastatic staging is an important prognostic factor in the majority of cancers.

Cancer cells lose their specialisation and the ability of natural programmed cell death, called apoptosis and cancer can be produced by a disruption in the homoeostasis between cell proliferation and apoptosis in favour of proliferation. It can be driven by mutations or chromosomal rearrangement occurring over a period of time. Therefore cancer is a genetic disease (CRUK, 2018).

Each cell type has its own specialisation and various cancers are grouped by the type their progenitor. Currently, more than 200 different cancers are described and they mainly fall into five major types:

- Carcinoma - cancer that originates in the skin or organ lining tissue;

- Sarcoma - originates from the connective or supportive tissue such as muscle, fat, bone;
- Leukaemia - blood cancers that originate in blood forming tissue such as bone marrow
- Lymphoma and myeloma - cancers that originate from the immune system cells
- Brain and spinal cord cancers - also known as central nervous system (CNS) cancers, originate from brain or spinal cord cells.

Cancer is the leading cause of deaths worldwide and was responsible for 28% of all registered deaths in the UK in 2015 (CRUK, 2018).

1.1.1. Cancer incidence. Cancer incidence in the developed world is on the rise despite increasing awareness of the causes and lifestyle choices that can reduce the risk of developing cancer. There were around 360,000 new cancer cases in the UK in 2015, which currently is the most recent year with available statistics; this is roughly 980 cases diagnosed every day and with someone diagnosed with cancer every 2 minutes (CRUK, 2018).

Cancer is mostly a disease of the elderly population (see Fig. 1.1), as nearly two-thirds of cancers (63%) are diagnosed in people over 65 years of age, with the incidence peaking at the age of 85 and only 1% of all cancer cases are children (here defined as aged under 16).

Therefore, despite complex reasons for the increase of cancer incidence in the general population, this increase should be considered in the light of an increased overall longevity in the population due to improvements in the treatment of other diseases.

Social and life-style factors play an important role in cancer incidence and mortality, which is reflected in increased numbers of cancer incidence in males compared to females (see Figure 1.2). Apart from gender-specific cancers such as breast, ovarian and uterine, the incidence and mortality in males are higher almost in every type of cancer due to higher rates of smoking and alcohol intake and a greater reluctance to seek medical attention in males (Peate, 2011).

1.1.2. Childhood cancer. Childhood cancer is generally defined as a cancer in children aged 0-14, however in some cases, the definition includes children under 16 and, less often, adolescents between 15-19 years old (Bahadur, 2000).

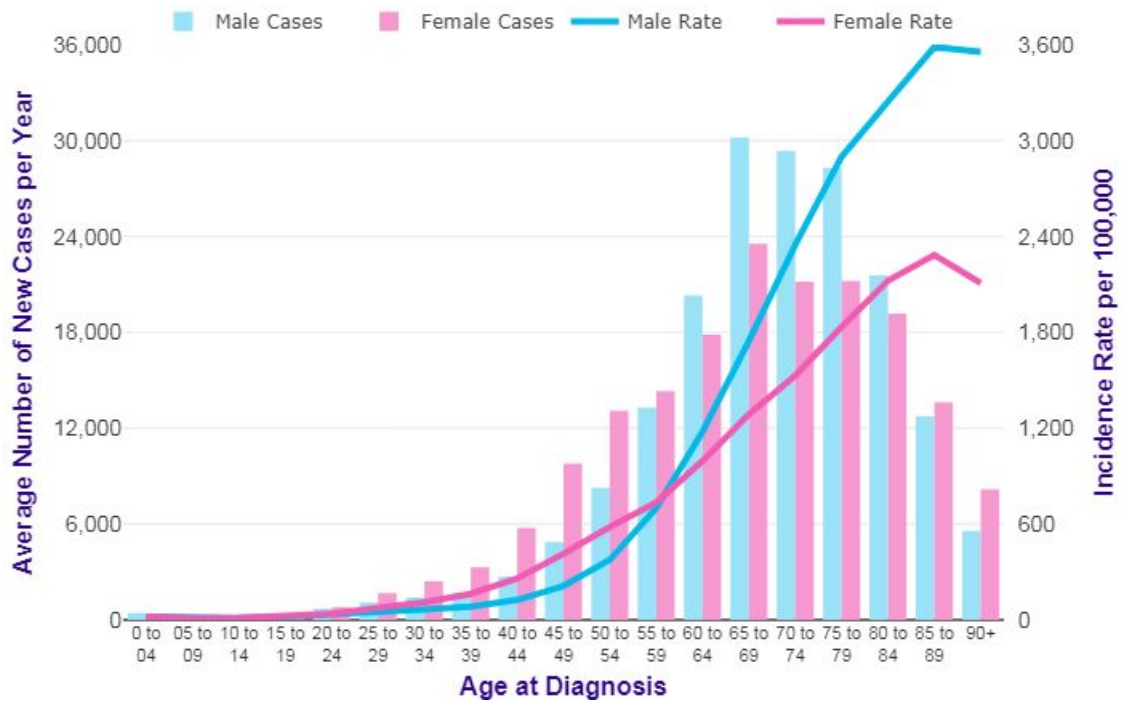


FIGURE 1.1. All cancers, average number of new cases per year and age-specific incidence rates per 100,000 Population, UK, 2013-2015 (CRUK, 2018)

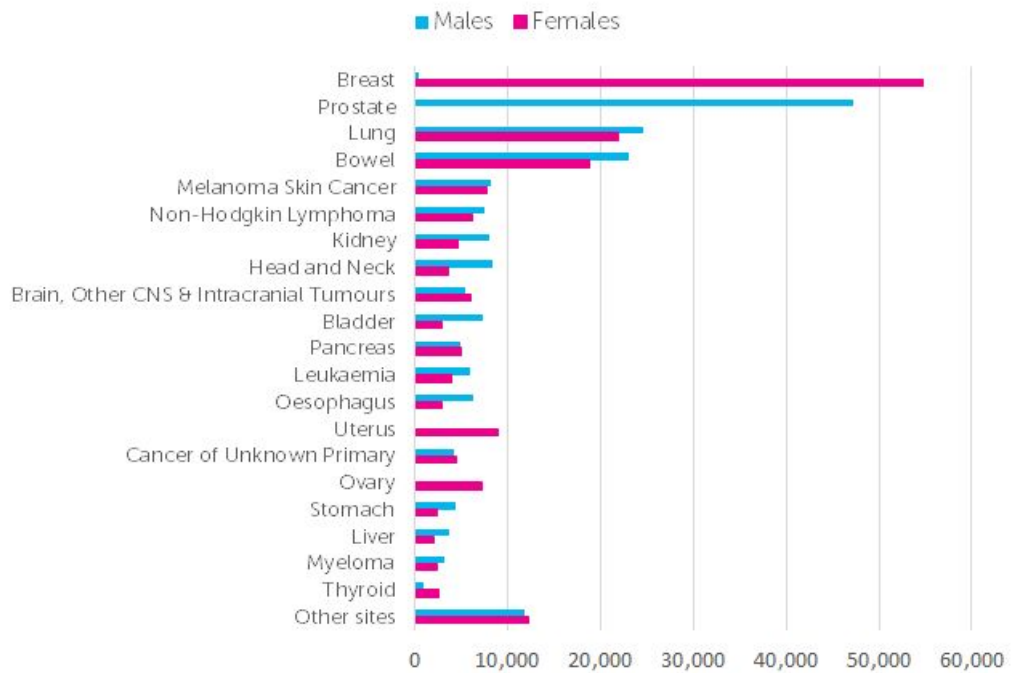


FIGURE 1.2. The 20 most common cancers (by gender), UK, 2015 (CRUK, 2018)

Childhood cancers typically arise from cells derived from embryonic tissue, while adult cancers are mostly epithelial in origin.

There were around 1,800 cases of childhood cancer diagnosed in the UK each year between 2012-2014 (CRUK, 2018), which amounts to around 5 children diagnosed

every day and accounts for about 1% of all cancer cases; there were about 140 more boys diagnosed with cancer than girls.

More than two-thirds of all childhood cancer cases diagnosed in the UK are leukaemia, brain, other CNS and intracranial tumours and lymphomas (see Figure 1.3).

Lifestyle choices and environmental risk factors will have less impact on the incidence of childhood cancer than in adult cases, since children had less time to be exposed for these factors to become significant. However, risk factors for leukaemia may include parental smoking, parental exposure to paint or high-level residential exposure to magnetic fields (CRUK, 2018).

The major risk factors for brain tumours, as well as other CNS and intracranial tumours, are thought to relate to genetic syndromes and congenital disorders. For lymphomas, risk factors include problems with the immune system and certain infections such as Epstein-Barr virus (EBV), the bacterium *Helicobacter pylori* and malaria (Shannon-Lowe et al., 2017). About 90% of the world population is infected with the Epstein-Barr virus that causes little harm in most cases. However EBV can cause cancers such as lymphomas, carcinomas and childhood cancers. Almost all stomach cancer cases occur in people who have had a long-term infection with *Helicobacter pylori*. The malaria parasite is thought to play a role in the development of one of the most frequent cancers among African children called Burkitt's lymphoma (Robbiani et al., 2015). However evidence for the childhood cancer risk factors are unclear (CRUK, 2018).

Cancer survival rates for children are improving and are more than double the survival rates of 40 years ago, amounting to around three-quarters (76%) of children diagnosed with cancer surviving their disease for ten years or more (2001-2005) (CRUK, 2018).

However, the therapy responsible for this improved survival often produces adverse long-term health-related outcomes, referred to as late effects, which manifest months to years after completion of cancer treatment (Howlader et al., 2016).

This problem is characteristic of childhood cancers: late effects (see 1.3.13) are commonly experienced by adults who have survived childhood cancer; the prevalence

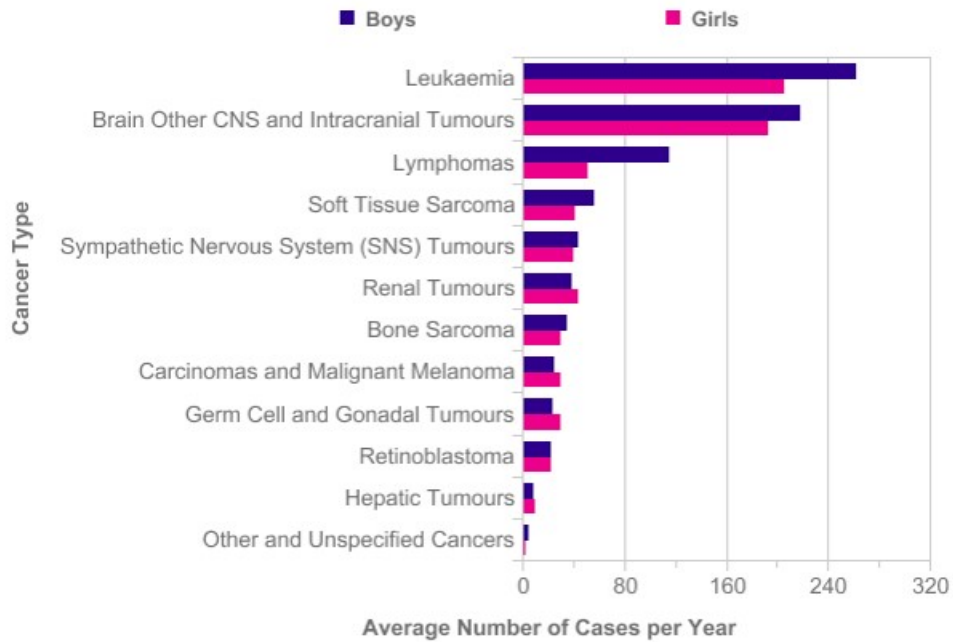


FIGURE 1.3. Childhood Cancers by Cancer Type (in concordance with International Classification of Childhood Cancer (ICCC)), Average Number of New Cases per Year, Ages 0-14, Great Britain, 2006-2008, (CRUK, 2018)

of late effects increases with the time elapsed from cancer diagnosis (Rebholz et al., 2011; Nathan et al., 2008).

1.1.3. Genetic model of cancer. A series of mutations are normally required in order for a normal cell to transform to a cancerous cell.

In 1976, Peter Nowell presented a perspective on cancer as an evolutionary process, driven by a multistage process of somatic cell mutations with sequential, sub-clonal selection. He drew a parallel to Darwinian natural selection with cancer equivalent to an asexually reproducing, unicellular, quasi-species (Nowell, 1976; Greaves and Maley, 2012).

In 1990, Fearon and Vogelstein presented a model of clonal evolution of colorectal carcinoma, thus formalising the model for the clonal evolution of cancer. Under this model, a mutation in a single cell provides growth advantage over surrounding cells. This cell proliferates, increasing the likelihood of further mutations which will provide further survival advantage (Fearon and Vogelstein, 1990).

1.1.4. Cancer development. Cancer is a complex multistage process during which a normal cell progenitor must acquire various capabilities to avoid the natural defense mechanisms that have evolved to keep the organism protected from cancer.

There are at least six processes, the so-called hallmarks of cancer, that a cell must acquire in order to become malignant (Hanahan and Weinberg, 2000, 2011).

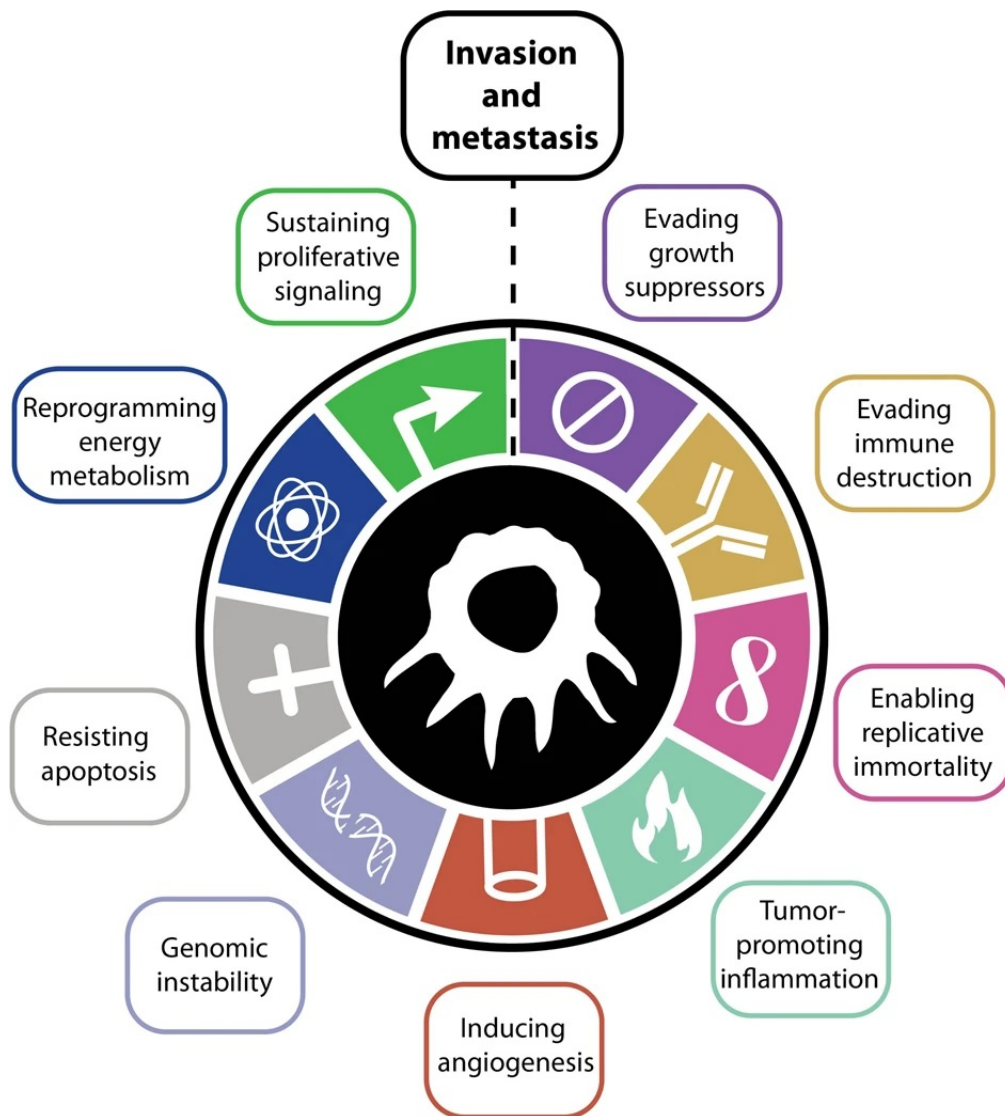


FIGURE 1.4. The hallmark capabilities of cancer, necessary for tumour growth. Figure courtesy (Hanahan and Weinberg, 2011; Meirson et al., 2020)

Despite these hallmarks being common to most cancers, the mechanism of their acquisition will differ within specific cancers, as well as across all cancers.

Each specific process is discussed below.

1.1.4.1. *Sustaining Proliferative Signalling.* One of the most fundamental distinctive traits of cancer cells is their ability to sustain constant proliferation. Normal tissue maintains homoeostasis of cell number and thus normal tissue function and architecture by careful control of the production and release of growth signals that control the cycle of cellular growth and division.

Normal cells require external mitogenic growth signals (GS) before they can enter an active proliferative state. These signals are transmitted to the cell by transmembrane

receptors that bind different classes of signalling molecules. There are no known types of normal cells that can proliferate in the absence of these stimulatory signals. However, many oncogenes are capable of mimicking normal growth signaling, allowing cancer cells to acquire GS autonomy. This liberation from dependence on externally derived signals disrupts the homeostatic mechanism that maintains normal tissue (Hanahan and Weinberg, 2000, 2011).

1.1.4.2. *Evading Growth Suppressors.* Another important trait of normal tissue is that in addition to GS autonomy, tissue homeostasis is maintained by multiple anti-proliferative signals. These growth-inhibitory signals, as well as growth signals, are received by trans-membrane cell surface receptors. Anti-growth signals can block cell proliferation by two different mechanisms. A cell may be forced out of an active proliferative state into a quiescent state until another extra-cellular signal returns it to the active state; alternatively, the cell may be induced to enter a post-mitotic state, usually associated with differentiation-associated traits.

Many tumour-suppressor genes are known to control anti-proliferative signaling and loss or mutation of these genes is common in cancers (Hanahan and Weinberg, 2000, 2011).

1.1.4.3. *Resisting Cell Death.* The ability of a tumour cell population to grow uncontrollably is determined not only by an increased cell proliferation rate, but also by ability of cancer cells to avoid programmed cell death - apoptosis.

An apoptotic program is present in virtually all cell types throughout the body.

Cancer cells acquire resistance to apoptosis in a variety of ways. The most common mechanism is loss or mutation of the tumour suppressor gene *TP53*, which acts as proapoptotic regulator, resulting in functional inactivation of p53 protein. This mechanism is observed in more than 50% of all human cancers (Hanahan and Weinberg, 2000, 2011).

1.1.4.4. *Enabling replicative immortality.* These three important acquired capabilities of cancer cells (growth signal autonomy, insensitivity to anti-proliferative signals and resisting of apoptosis) lead to disconnecting of a cell's growth program from external signals in its environment, promoting autonomy. However, research over the last 30 years shows that this acquired disruption of cell to cell signaling on its own is not sufficient to ensure expansive tumour growth (Hanahan and Weinberg, 2000).

All types of mammalian cells have cell-autonomous programs that limit their replicative potential and which is independent of the cell-to-cell signaling pathway. In normal cells, there is only a limited number of cell divisions possible before a cell enters senescence - a non-proliferative but viable state. This counting mechanism for cell generations is the telomeres at the end of chromosomes. In normal cells, telomeres get progressively shortened with every replication until they are massively eroded and lose their protective function and trigger senescence. Telomerase, an enzyme that controls erosion of telomeres by adding species-specific telomere repeat sequence to the 3' end of telomeres, is actively expressed in cancer cells and its expression correlates with senescence and apoptosis resistance (Hanahan and Weinberg, 2011).

1.1.4.5. *Inducing angiogenesis.* During embryogenesis, development of vasculature involves formation of new blood vessels from epithelial cells (vasculogenesis) as well as sprouting of new vessels from existing ones (angiogenesis). In adults, normal vasculature becomes mainly dormant and is only turned on temporarily in particular physiological processes like wound healing. Just as normal tissue, tumours require nutrients and oxygen, as well as the ability to free themselves of metabolic waste. This is addressed by angiogenesis, and in contrast to a normal tissue a so-called "angiogenic switch" is almost always on, causing usually dormant vasculature to continue sprouting new blood vessels in order to sustain growing neoplasms (Hanahan and Weinberg, 2011). Tumours appear to activate the angiogenic switch by altering the balance between angiogenic inducers and inhibitors (Hanahan and Folkman, 1996).

1.1.4.6. *Activating invasion and metastasis.* Most human cancers are known to spread by separating cells from primary tumour that invade nearby tissue and also move out to remote sites of the body by entering the bloodstream or lymphatic vessels from where cancer cells escape to surrounding parenchyma and form small masses of cancer cells (micro-metastases) which, in turn, grow into new tumours in a process called colonisation. These distant colonies of tumour cells, metastases, are responsible for 90% of cancer deaths (Sporn, 1996). Cancer cells typically develop changes in their shape as well as in their attachment to adherent cells and to extra-cellular matrix (ECM). Commonly, expression of genes encoding cell-to-cell and cell-to-ECM adhesion molecules is significantly altered (Hanahan and Weinberg, 2011).

1.1.4.7. *Genome Instability and Mutation.* Acquisition of the multiple hallmarks depends on a succession of alterations in the genomes of neoplastic cells. Certain mutant genotypes lead to selective advantage on sub-clones of cells, enabling their outgrowth and eventual dominance in a local tissue environment (Hanahan and Weinberg, 2011). The ability of genome maintenance systems to detect and repair defects in the DNA ensures that rates of spontaneous mutation are very low in each cell generation. In the course of acquiring the set of mutant genes needed to trigger tumorigenesis, cancer cells often increase the rates of mutation (Negrini et al., 2010; Salk et al., 2010). This is achieved by increased sensitivity to mutagenic agents, or a breakdown in one or several components of the genomic maintenance machinery. The accumulation of mutations can be accelerated by compromising the surveillance systems that normally monitor genomic integrity and force cells with damaged DNA into either senescence or apoptosis ((Kastan, 2008; Sigal and Rotter, 2000) The defects in genome maintenance and repair are selectively advantageous and therefore important for tumour progression, as they accelerate the rate at which evolving pre-malignant cells can accumulate favourable genotypes. This makes genome instability an enabling characteristic that is associated with the acquisition of hallmark capabilities (Hanahan and Weinberg, 2011).

1.1.4.8. *Tumour-Promoting Inflammation.* It has long been recognised that some tumours are highly infiltrated by cells of both the innate and adaptive immune systems and therefore, mimic inflammatory conditions arising in non-neoplastic tissue (Dvorak et al., 1986). Inflammation can contribute to multiple hallmark capabilities by supplying bioactive molecules, such as growth factors, survival factors that limit cell death, pro-angiogenic factors, extracellular matrix-modifying enzymes that facilitate angiogenesis, invasion, and metastasis, and others, to the tumour microenvironment (DeNardo et al., 2010; Grivennikov et al., 2010; Qian and Pollard, 2010).

In some cases, inflammation is present at the earliest stages of neoplastic progression and facilitates the development of neoplasias into full-blown cancers (Qian and Pollard, 2010; De Visser et al., 2006). Additionally, inflammatory cells are capable of releasing chemicals, such as reactive oxygen species, that are actively mutagenic for nearby cancer cells, accelerating their genetic evolution toward states of increased malignancy (Grivennikov et al., 2010). Therefore, inflammation can be considered an enabling

characteristic as it contributes to the acquisition of core hallmark capabilities (Hanahan and Weinberg, 2011).

1.1.4.9. *Reprogramming Energy Metabolism.* The uncontrolled cell proliferation in neoplastic disease, alongside deregulated cell proliferation control, includes changes of energy metabolism in order to sustain cell growth and division. Under aerobic conditions, normal cells process glucose, first to pyruvate via glycolysis in the cytosol and then to carbon dioxide in the mitochondria; under anaerobic conditions, glycolysis prevails and relatively little pyruvate is available to the oxygen-consuming mitochondria. Cancer cells can reprogram their glucose metabolism and their energy production even in the presence of oxygen by limiting their energy metabolism largely to glycolysis, leading to a state that has been termed "aerobic glycolysis" (Hanahan and Weinberg, 2011). Increased uptake and utilisation of glucose have been documented in many human tumour types and has been shown to be associated with activated oncogenes such as *RAS*, *MYC* and mutant tumour suppressors such as *TP53* (DeBerardinis et al., 2008), which are also confer the hallmark capabilities of cell proliferation, avoidance of cytostatic controls, and reduction of apoptosis (Hanahan and Weinberg, 2011).

1.1.4.10. *Evading Immune Destruction.* Cancer cells adapt in order to evade detection and destruction by the host's immune system (Hanahan and Weinberg, 2000, 2011). They do it by hijacking normal mechanisms of immune checkpoint control and modulation of the innate immune response via stimulator of interferon genes (STING). STING is a key mediator of innate immunity, and the STING pathway has been shown to be involved in the induction of an anti-tumour immune response (Barber, 2015).

Immune checkpoints refer to the built-in control mechanisms of the immune system that maintain self-tolerance and help to avoid damage to own tissue during a physiological immune response. It is now evident that tumours engineer microenvironments to evade immune surveillance and attack, particularly by modulating certain immune-checkpoint pathways (Sharma and Allison, 2015)).

1.1.5. Genetic origin of cancer. As discussed in section 1.1, cancer is a genetic disease and more than one mutation is necessary for tumorigenesis. The majority of these mutations occur under the influence of environmental factors such as chemical mutagens, ionising irradiation and many others.

Also there is a number of familiar cancer syndromes, where predisposition to various cancers is inherited. Examples of such syndromes will be discussed later in this text (see section 1.3.4.1) and are presented in table 1.3.

There are two types of genes that are essential for cancer development - oncogenes and tumour suppressor genes, or anti-oncogenes.

1.1.6. Oncogenes. One of the driving forces of cancer is oncogenes - genes that have the potential to cause cancer when mutated or expressed at abnormally high levels. Oncogenes are capable of causing cells otherwise destined for apoptosis to survive and proliferate, or promote development of other cancer hallmarks described in section 1.1.4.

The theory of oncogenes was first anticipated by the German biologist Theodor Boveri in his 1914 book *Zur Frage der Entstehung Maligner Tumoren* ('The Origin of Malignant Tumours', (Boveri, 1914)); and later rediscovered in 1969 by National Cancer Institute scientists George Todaro and Robert Heubner (Mukherjee, 2010).

Most oncogenes are derived from proto-oncogenes, normal genes that are involved in the regulation of cell proliferation, differentiation and apoptosis, through dominant gain-of-function mutations or by upregulated expression. Multiple oncogenes have been identified across human cancers and many cancer drugs target the proteins encoded by oncogenes (Croce, 1995; Yokota, 2000).

The main categories of well-studied oncogenes are shown in Table 1.1.

1.1.6.1. *Activation of proto-oncogenes.* A proto-oncogene can become an oncogene by a relatively minor modification of its original function. There are three main ways in which a proto-oncogene can become activated.

First: a mutation within a proto-oncogene, or within a regulatory region (e.g. its promoter region), can alter protein structure, causing either an increase in protein activity or a loss of regulation.

Second: amplification of a gene, that may increase its transcription, resulting in an increased amount of protein in the cell.

Third: a chromosomal translocation - chromosome abnormality caused by rearrangement of parts between nonhomologous chromosomes. There are two different types of chromosomal translocations that can occur: translocation events which relocate a

Category	Example	Normal gene function	Cancer Association
Growth factors	c-Sis	induces cell proliferation	glioblastomas, fibrosarcomas, osteosarcomas, breast carcinomas, and melanomas
Transcription factors	<i>MYC</i>	regulate transcription of genes that induce cell proliferation	malignant T-cell lymphomas and acute myeloid leukemias, breast cancer, pancreatic cancer, retinoblastoma, and small cell lung cancer
Receptor tyrosine kinases	epidermal growth factor receptor (EGFR), platelet-derived growth factor receptor (PDGFR), and vascular endothelial growth factor receptor (VEGFR), HER2/neu	transduce signals for cell growth and differentiation	breast cancer, gastrointestinal stromal tumours, non-small-cell lung cancer and pancreatic cancer
Cytoplasmic tyrosine kinases	Src-family, Syk-ZAP-70 family, and BTK family of tyrosine kinases, the Abl gene in CML - Philadelphia chromosome	mediate the responses to, and the activation receptors of cell proliferation, migration, differentiation, and survival	colorectal and breast cancers, melanomas, ovarian cancers, gastric cancers, head and neck cancers, pancreatic cancer, lung cancer, brain cancers, and blood cancers
Cytoplasmic Serine/threonine kinases and their regulatory subunits	Raf kinase, and cyclin-dependent kinases (through overexpression)	involved in organism development, cell cycle regulation, cell proliferation, differentiation, cells survival, and apoptosis	malignant melanoma, papillary thyroid cancer, colorectal cancer, and ovarian cancer
Regulatory GTPases	Ras protein	involved in signalling a major pathway leading to cell proliferation	adenocarcinomas of the pancreas and colon, thyroid tumours, and myeloid leukemia

TABLE 1.1. Classification of the oncogenes, their normal function and association with the cancers in which they become activated. Table based on data from (Chial et al., 2008; Gschwind et al., 2004; Hilgenfeld, 1995).

proto-oncogene to a new chromosomal site that leads to higher expression; translocation events that lead to a fusion between a proto-oncogene and a second gene, creating a fusion protein with increased oncogenic activity (Chial et al., 2008; Lodish et al., 2013).

1.1.7. Tumour suppressor genes. Tumor-suppressor genes (TSG) or antioncogenes, are generally proteins that promote inhibition of cell proliferation and apoptosis. Loss or mutation of one or more of these genes contributes to the development of many cancers. Five broad classes of proteins are generally recognized as being encoded by tumor-suppressor genes (Lodish et al., 2013):

- Intracellular proteins, that regulate or inhibit progression through a specific stage of the cell cycle, such as the p16 cyclin-kinase inhibitor.

- Receptors for secreted hormones such as tumour-derived growth factor β , that function to inhibit cell proliferation
- Checkpoint-control proteins that arrest the cell cycle if DNA is damaged or chromosomes are abnormal
- Proteins that promote apoptosis, such as TP53
- Enzymes that participate in DNA repair

1.1.7.1. *Inactivation of tumour suppressor genes.* Generally tumor suppressor genes inactivation follows the "two-hit hypothesis". The two-hit hypothesis of TSG inactivation was first proposed by A.G. Knudson in 1971. After studying cases of retinoblastoma, Knudson observed that the age of onset of retinoblastoma followed second order kinetics which implies that two separate genetic effects were necessary.

According to this theory both alleles coding for a particular protein must be affected before tumours occurred. This is because if only one allele for the gene is damaged, the second one can still produce the correct protein, unlike in oncogene activation. Thus, mutations in tumor suppressors' alleles are usually recessive, whereas mutations in oncogene alleles typically involve gain-of-function and are typically dominant (Knudson, 1971).

However, there are important exceptions to the "two-hit" rule for tumor suppressors, such as those that exhibit haplo-insufficiency, including PTCH1 in medulloblastoma and NF1 in neurofibroma. Further example of this is the p27Kip1 cell-cycle inhibitor, in which mutation of a single allele causes increased carcinogen susceptibility (Kemp et al., 1998).

Another notable exception is specific mutations in the *TP53* gene. These mutations are "dominant negative", and a mutated p53 protein can counteract the function of normal protein from the non-mutated allele (Baker et al., 1990).

1.1.8. Copy number variations and cancer. DNA copy number variation (CNV) is a type of structural variation of DNA in which sections of the genome are repeated, and the number of repeats in the genome varies between individuals in the human population due to a duplication or deletion event that affects a considerable number of base pairs (Sharp et al., 2005; McCarroll and Altshuler, 2007). CNVs are an

important component of genetic variation of the population, affecting a greater proportion of the genome than single nucleotide polymorphisms (SNPs) (Shlien and Malkin, 2009).

Somatic copy number variations that can lead to abnormal cell functioning and eventually, to cancer, termed copy number aberrations or alterations (CNAs), are gains and losses of large segments of the genome that occur during the lifetime of an individual and affect a larger proportion of the genome in cancer cells than do any other type of somatic genetic alteration (Beroukhim et al., 2010; Kim et al., 2013). They range in size from a few kilobases to whole chromosomes, and may play a critical role in inactivation of tumour-suppressor genes or activation of oncogenes (Stratton et al., 2009; Xue et al., 2012) and, therefore, are major contributors to oncogenesis, particularly for solid tumors (Zack et al., 2013). An understanding of the biological effects of somatic CNAs on the cellular phenotype has led to substantial advances in cancer diagnostics and therapeutics (Tsao et al., 2005; Lowe et al., 1994; Cheang et al., 2009). An illustrative example of CNA with a phenotypic effect in cancer is the mitochondrial tumor suppressor gene (*Mtus1*): a small deletion in *Mtus1* is associated with a increased risk of familial and high-risk breast cancer (Frank et al., 2007).

Recently, more advanced technologies have been introduced to measure CNAs. Cytogenetic techniques such as comparative genomic hybridization were superseded by more advanced array-comparative genomic hybridization (aCGH), SNP genotyping arrays and high-throughput sequencing platforms. In line with the technological developments, numerous computational methods have been developed to identify CNAs in single samples using these newer platforms (Chiang et al., 2009; Hupé et al., 2004; Olshen et al., 2004).

1.1.9. Epigenetics and cancer. Epigenetics, which literally means upon genetics, is the study of various mechanisms that regulate both gene expression and genome stability without modifying the DNA sequence itself (Baylin and Jones, 2011). The most widely studied epigenetic changes are DNA methylation of cytosines within CpG dinucleotides, histone modification including acetylation, methylation, phosphorylation, and ubiquitination (Esteller, 2007; Bernstein et al., 2007). Other mechanisms associated with gene regulation and chromatin structure, are non-coding RNAs and micro RNA (miRNA) gene silencing (Chen et al., 2007).

The normal function of these epigenetic processes is to provide a framework for development and differentiation, including tissue-specific gene expression, inactivation of the X-chromosome, and genomic imprinting (Ferguson-Smith, 2011; Lau et al., 2004). Epigenetics is also involved with ageing and response to environmental factors (Benayoun et al., 2015; Cortessis et al., 2012). Changes in epigenetic mechanisms are associated with a wide variety of malignant and non-malignant diseases (Jones and Baylin, 2007a) and, acting together with genetic mechanisms, provide cancers with the necessary hallmarks described in section 1.1.4 (Hanahan and Weinberg, 2000, 2011). The main epigenetic modification mechanisms are shown on Figure 1.5.

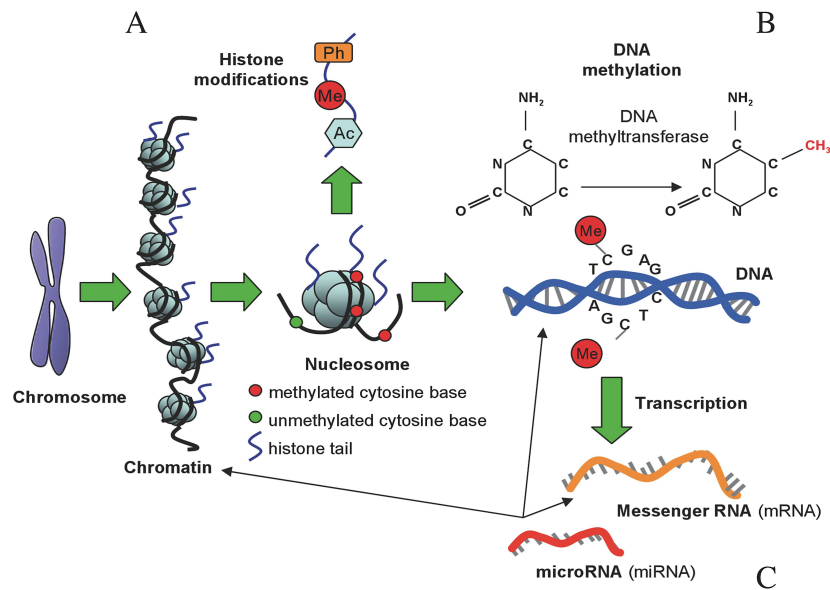


FIGURE 1.5. Epigenetic mechanisms. (A) Histones can undergo, among other chemical modifications, methylation (Me), acetylation (Ac) and phosphorylation (Ph). These modifications are involved in chromatin remodeling and transcriptional regulation. (B) DNA molecules are methylated by the addition of a methyl group to carbon 5 on cytosine bases, a reaction catalysed by DNA methyltransferase enzymes, which maintains repressed gene activity. (C) mRNA is translated into a protein. This process can be repressed by binding of microRNAs (miRNA), a class of non-coding RNA (ncRNA). Figure adapted from (Relton and Davey Smith, 2010).

1.1.9.1. *DNA methylation in cancer.* In mammals, all cells descended from a zygote share the same DNA sequence. The cells are differentiated during embryogenesis by means of epigenetic changes, such as methylation of CpG islands and chromatin remodelling. DNA methylation is a process of addition of a methyl CH_3 group to carbon position 5 on cytosine bases within DNA, that is positioned adjacent to a guanine base (typically named a CpG dinucleotide) within DNA, which regulates gene expression activity without changing the sequence. Hypermethylated gene promoters are associated with gene silencing, whereas hypomethylation of these regions is associated with active

expression of genes (Seisenberger et al., 2012). Once tissue differentiation is complete, methylation patterns of silencing or activation of particular genes are inherited during cell division (Seisenberger et al., 2012).

DNA methylation affects transcription of the genes in two ways: directly, by physically impeding transcriptional proteins from binding to DNA and indirectly, by recruiting proteins known as methyl-CpG-binding domain proteins (MBDs). MBD proteins then involve additional chromatin remodelling proteins, such as histone deacetylases, that can modify histones and form compact, inactive heterochromatin and therefore prevent transcription (see 1.1.9.2).

DNA methylation differs between normal cells and tumour cells. Cancer is typically characterised by genome-wide DNA hypomethylation and promoter hypermethylation (Howard et al., 2008; Esteller, 2007). The hypomethylation observed in cancer often occurs at satellite DNA, the main component of functional centromeres, and at other repeating sequences that do not function as transcriptional units. The CpG methylation profile associated with normal cells is often inverted in cells that become cancerous (Esteller, 2007). In normal cells, CpG islands located before gene promoters are generally unmethylated, and tend to be transcriptionally active, while other individual CpG dinucleotides throughout the genome tend to be methylated. However, in cancer cells, CpG islands upstream from tumour suppressor gene promoters are often hypermethylated, while CpG methylation of oncogene promoter regions and repeat sequences is often decreased (Banister, 2012). Hypermethylation of tumour suppressor gene promoter regions can result in silencing of those genes. This type of epigenetic aberration allows cells to grow and reproduce uncontrollably, leading to tumorigenesis (Esteller, 2007).

1.1.9.2. *Histone Modifications and chromatin remodelling in cancer cells* . Another important epigenetic factor in cancer, histone modification patterns are also highly disrupted in cancer. Histones are alkaline proteins found in eukaryotic cell nuclei, that package DNA to form a DNA/protein complex called chromatin. The basic unit of chromatin, the nucleosome, is composed of DNA wrapped around four pairs of histones, called core histones (H2A, H2B, H3 and H4), forming the nucleosomes (Esteller, 2007).

The functions of histones are regulated by post-translational modification of their amino-terminal tail, which includes methylation, acetylation, phosphorylation and

ubiquitination, thus regulating gene expression by modifying chromatin into inactive heterochromatin regions, where DNA is compact and less accessible for transcription, or euchromatin, active regions accessible for transcription (Sharma et al., 2009).

Addition and removal of histone modifications (marks) is carried out by a number of different histone modifying enzymes. The enzymes involved in addition of these reversible marks, so-called "writers", and removal ("erasers") include histone lysine methyltransferases (HMTs, "writers"), demethylases (HDM, "erasers"), acetyltransferases (HAT, "writers"), deacetylases (HDACs, "erasers").

Acetylation of histone tails is typically associated with transcriptional activation of genes, whereas histone methylation can either activate or repress transcription, depending on location and number of methyl groups (Chervona and Costa, 2012). For example, histone 3 lysine 27 di- and trimethylation (H3K27me2 and H3K27me3) and histone 3 lysine 9 di- and trimethylation (H3K9me2 and H3K9me3) are associated with inactive chromatin and repression of gene expression, while histone 3 lysine 4 di- and trimethylation (H3K4me2 and H3K4me3) and histone 3 lysine 9 monomethylation (H3K9me1) are associated with open chromatin and active gene expression (Schneider and Grosschedl, 2007)

Hypermethylation of CpG islands in the promoter regions of tumour-suppressor genes in cancer cells is associated with a particular combination of histone marks: Deacetylation of histones H3 and H4, and methylation changes at various amino acid residues of H3 (Jones and Baylin, 2007b), including loss of tri-methylation at lysine 4, and gain of methylation at lysine 9 and trimethylation at lysine 27 (Martín-Subero and Esteller, 2017). The presence of the hypoacetylated and hypermethylated histones H3 and H4 silences certain genes with tumor-suppressor-like properties, such as *p21WAF1*, despite the absence of hypermethylation of the CpG island (Richon et al., 2000). On the other hand, silencing of *SIRT1*, a histone deacetylase, leads to increased H3 and H4 acetylation of cancer genes that become reactivated despite full retention of DNA hypermethylation (Pruitt et al., 2006). These are examples of complex interplay between DNA methylation and histone modifications, which determines gene activation or repression.

Knowledge of the methylation status of certain genes can be very important for targeted cancer therapy. For example, epigenetic silencing of the *MGMT* (O6-methylguanine

- DNA methyltransferase) DNA-repair gene by promoter methylation compromises DNA repair and has been associated with longer survival in patients with glioblastoma who receive alkylating agents, such as temozolomide. Patients who did not have methylated *MGMT* promoter, did not benefit from the treatment (Hegi et al., 2005).

1.1.9.3. *MicroRNA and Long Non-Coding RNAs dysregulation.* MicroRNAs (miRNA) are small noncoding RNAs made of short stretches of 22 nucleotides that regulate post-transcriptional gene expression by binding to complementary sequences in the 3' untranslated region (UTR) of messenger RNA (mRNA). Target genes of miRNAs are involved in diverse cellular functions, including cell differentiation, proliferation and apoptosis (Esquela-Kerscher and Slack, 2006; Bruce et al., 2015).

The expression of miRNA is similar to that of protein-coding genes as they are regulated by both genetic and epigenetic mechanisms (Biswas and Rao, 2017) and have been found to be dysregulated in cancer development (Lu et al., 2005; Calin and Croce, 2006). Recent research has mapped the presence of miRNA genes in the common breakpoint regions of oncogenes and tumour suppressor genes, in addition to fragile regions of the genome which are preferential sites for deletion, translocation or amplification in cancer, suggesting their involvement in driving the behaviour of tumour growth (Biswas and Rao, 2017). Moreover, there is evidence that miRNAs play a role in epigenetic regulation, by modulating the activity of epigenetic modifying enzymes associated with carcinogenesis (Guil and Esteller, 2009). miRNA serves as a part of the complex regulatory network which takes part in silencing gene expression by methylation and modifying the structure of chromatin (Baer et al., 2013).

Long Non-Coding RNAs (LncRNAs) are considered as non-protein coding transcripts that are transcribed in an anti-sense manner. They regulate integrity of the nuclear structure, regulation of gene expression, and post-transcriptional processing. LncRNAs are involved in epigenetic gene silencing and in tumour development by promoting expression of genes involved in metastasis (Huarte, 2015).

1.1.10. Relationship between genetic and epigenetic factors in cancer development. Experimental evidence support both the epigenetic and genetic origin of cancer (Feinberg et al., 2006; Fearon, 1997) and shows that genetic and epigenetic mechanisms closely interact in carcinogenesis (Brena and Costello, 2007). This interaction

between genetic and epigenetic changes can be clearly observed in the case of tumor-suppressor gene inactivation, which can be caused by genetic (mutation, deletion) or epigenetic (DNA hypermethylation) means (Martín-Subero and Esteller, 2017). For example, the VHL gene is mutated in 60 % of the renal carcinomas and hypermethylated in 20 % of the remaining cases (Jones and Baylin, 2002).

The close interaction between genetic and epigenetic changes is also supported by the discovery of epigenetic changes that affect the stability of the genome, e.g. hypomethylation of DNA repeats leads to chromosomal changes by inducing chromosomal instability. Also, genes targeted by DNA hypermethylation are involved in DNA repair pathways like *BRCA1*, *MGMT*, and *WRN* (Esteller, 2007). In these cases, silencing of the DNA-repair gene blocks the repair of mutations, thus triggering neoplastic transformation of the cell.

In summary:

- (1) genetic and epigenetic changes represent alternative mechanisms targeting the same genes in cancer
- (2) genetic changes of epigenetic genes can lead to epigenetic modifications (and *vice versa*), and
- (3) epigenetic changes of DNA repair genes can lead to genetic alterations.

(Martín-Subero and Esteller, 2017).

1.2. Paediatric tumours of the central nervous system

Tumours of the Central Nervous System (CNS), including brain and intracranial tumours, are the second most common group of cancers in children, responsible for more than a quarter (26%) of all childhood cancers (CRUK, 2018).

The largest subgroup of CNS cancers is astrocytoma; this accounts for over two-fifths (43%) of all brain, other CNS and intracranial tumours in children (CRUK, 2018). These tumours have no distinct pattern regarding sex or age of children and are low grade. Grading is a measure of the degree of abnormality of the tumour compared to the normal tissue from which the tumour is thought to have originated. Some cancers are graded using specific grading system, but in general tumours are graded 1 (I) to 4(IV), with lower grade cancers cells appearing close to normal, with slow growth and spreading rates, whereas high grade tumours are aggressive and rapidly progressing (Louis et al., 2016; NCI, 2018).

The second most frequent subgroup is the intra-cranial and intra-spinal embryonal tumours, which account for around a fifth to a quarter (20-25%) of all childhood brain, other CNS and intracranial tumours. Most of this subgroup are embryonal tumours of CNS, with nearly three-quarters (73%) being medulloblastoma. These tumours most commonly present in young children (CRUK, 2018).

The third largest group, accounting for about 10% of childhood brain, other CNS and intracranial tumours, comprises ependymoma and choroid plexus tumours; incidence is highest in one-year-olds in this subgroup (CRUK, 2018).

Most of these CNS tumours occur sporadically, with fewer than 5% of all brain, other CNS and intracranial tumours occurring in children with familial genetic predisposition (Frühwald and Rutkowski, 2011).

1.2.1. Embryonal tumours of the Central Nervous System. Embryonal tumours (ET) of the CNS are tumours formed from undifferentiated embryonal cells that remained in the brain after birth (NCI, 2018). About 20-25% of childhood brain tumours are embryonal tumours. Embryonal tumours occur most frequently in younger children (aged 0-10 years), and incidence decreases with age, with more than half being diagnosed in children less than 10 years old. These tumours were formerly known as Primitive Neuro-Ectodermal Tumours (PNETs) and categorised by their common round cell morphology (Ellison, 2002). As understanding of distinctive

molecular and clinico-pathological characteristics progressed, these tumours were classified into five distinct types: medulloblastoma, atypical teratoid/rhabdoid tumours (AT/RT), ependymoblastoma, medulloepithelioma and supratentorial PNETs (Kleihues et al., 2002). This classification was further refined in 2007 to include three distinct types of embryonal tumours: medulloblastoma, AT/RT and primitive neuroectodermal tumours (CNS-PNETs); five histological variants of medulloblastoma were described, which have been associated with disease outcome (Louis et al., 2007). The

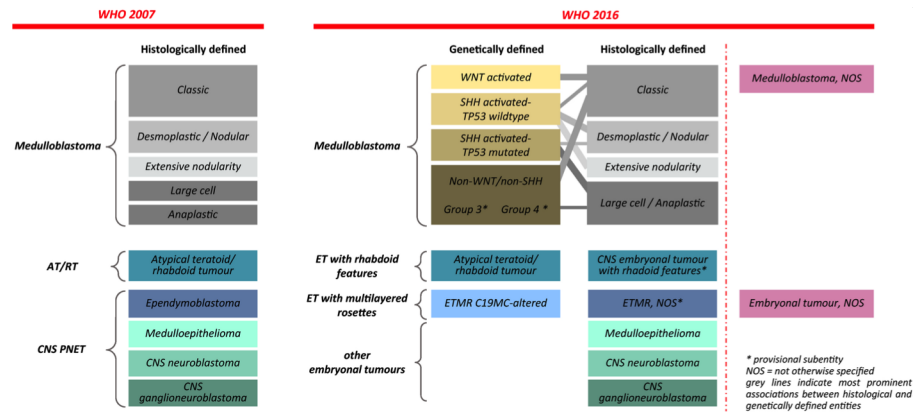


FIGURE 1.6. Summary of the major changes in the WHO 2016 classification for the diagnosis of central nervous system (CNS) embryonal tumours. In the 2007 WHO classification tumours were histologically defined. In the current edition, an integrated diagnosis combines histology with genetically defined tumours. For embryonal tumours, this has meant four new genetic subgroups of medulloblastoma. The CNS-PNET entity is no longer recognised, instead embryonal tumour with multilayered rosettes form their own embryonal entries, whereas any remaining tumours are currently classified based on histology alone and fall under ‘other embryonal tumours’. This group of genetically undefined tumours contains medulloepithelioma, CNS neuroblastoma, CNS ganglioneuroblastoma and CNS embryonal tumour, NOS. *Provisional subentity; NOS, not otherwise specified. Figure taken from (Pickles et al., 2018)

WHO 2007 classification relied only on histological features, which has been refined in the 2016 classification to additionally integrate genetic information (Fig.1.6). This classification divides ETs into medulloblastoma, embryonal tumours with rhabdoid features (AT/RT), embryonal tumours with multilayered rosettes (ETMR) and other ET that include medulloepithelioma, CNS neuroblastoma and CNS ganglioneuroblastoma. Medulloblastoma was further subclassified into four genetically defined subgroups and four histologically defined variants (Louis et al., 2016; Pickles et al., 2018).

Embryonal tumours of CNS are WHO grade IV and are associated with high mortality and long-term morbidity for survivors (Pfister et al., 2010; Louis et al., 2016).

1.3. Medulloblastoma

Medulloblastoma is the most common highly malignant, undifferentiated childhood brain tumour that accounts for 15-20% of all paediatric brain tumours, with incidence of approximately 90 cases per year in the UK, and responsible for nearly 10% of all childhood cancer deaths (Pizer and Clifford, 2009).

Medulloblastoma mainly occurs within the first decade of life with a peak incidence between 4 and 7 years of age. However, there are well described cases of adolescent and adult medulloblastoma. There is very clear gender imbalance with an approximately 1.7:1 male to female ratio (Pizer and Clifford, 2009).

Medulloblastoma was first identified and described by Bailey and Cushing in 1925 as a novel undifferentiated tumour derived from embryonic cells (Bailey and Cushing, 1925). Subsequently, medulloblastoma has been well described and both diagnosis and treatment strategies have been developing and improving. Recent advances in treatment have extended 5-year survival rates from 3% in 50s to 70% during 2001-2005 (Massimino et al., 2016; Newton, 2001) and currently reaching 85% (Ramaswamy et al., 2016a).

Usually, medulloblastomas arise in the posterior fossa, commonly from the cerebellar vermis in the roof of the 4th ventricle, presenting as a mid-line tumour, although in a minority of cases, particularly in older children, medulloblastoma can arise in a cerebellar hemisphere (Taylor et al., 2012).

1.3.1. Histology. The current histological classification of medulloblastoma, according to 2016 WHO classification (Louis et al., 2016), defines four main long-established variants: classic medulloblastoma (72-80%), desmoplastic/nodular and MBEN (Medulloblastoma with Extensive Nodularity) medulloblastomas (together amounting to 15%) and large cell /anaplastic medulloblastomas (19%) (Pizer and Clifford, 2009; Louis et al., 2016; McManamy et al., 2007).

1.3.1.1. *Classic medulloblastoma.* The most common subtype of medulloblastoma, comprising approximately 80% of all medulloblastomas, is described as having a classic phenotype, composed of sheets of generally small round cells with densely packed undifferentiated cytoplasm and frequent mitoses.

1.3.1.2. *Desmoplastic/nodular medulloblastoma and medulloblastomas with extensive nodularity.* Desmoplastic medulloblastomas incorporate the desmoplastic/nodular

MB variant and the MB with extensive nodularity (MBEN). These tumours have nodular architecture with rich reticulin fibres around the pale nodules, giving a desmoplastic appearance. Desmoplasia is most notably seen in very young children, where it occurs in up to 50% of tumours in infants of under 3 years of age, and has been associated with an improved outcome (Pizer and Clifford, 2009). Also desmoplasia appears to have high rates in adult cases, and just 5% of cases among children 3-15 years of age (McManamy et al., 2007), however any improved prognosis associated with desmoplasia in older patients is unclear (Pizer and Clifford, 2009). The histological subgroup MBEN is a variant closely related to DN, with a reduced inter-nodular component. Round cells of MBEN have uniform nuclei, a high level of neuronal differentiation and low proliferative index, and are gathered into nodules. This variant comprises 1-2% of all medulloblastomas, is mostly present in infants under 3 years of age and, similarly to the infant DN, is associated with favourable prognosis (Pietsch and Haberler, 2016).

1.3.1.3. *Large cell/anaplastic medulloblastoma.* Medulloblastoma with severe anaplasia, characterised by cytologic pleomorphism and a high mitotic count, is seen in about 15% of cases and is associated with a poorer prognosis as compared the classic phenotype (Pizer and Clifford, 2009; Ellison, 2002).

The rarer large-cell medulloblastoma characterised by large cells, with large pleomorphic nuclei and prominent nucleoli accounts for fewer than 5% of cases and has clearly been associated with a poor prognosis (Pizer and Clifford, 2009). This subtype has a higher apoptotic and mitotic rate than other medulloblastoma subtypes and, as a consequence, necrotic regions are often present. As both large cell and anaplastic medulloblastomas share a similarly poor prognosis, in most studies of medulloblastoma, they are usually grouped into a single large cell/anaplasia (LCA) category (Gilbertson and Ellison, 2008).

1.3.2. Diagnosis and clinical presentation of medulloblastoma. As medulloblastomas mostly arise in the posterior fossa, usually from the cerebellar vermis in the roof of the 4th ventricle (see Fig.1.8 and 1.9), most patients present with the symptoms of intracranial pressure caused by blockage of the cerebrospinal fluid (CSF) pathways at the forth ventricle (Pizer and Clifford, 2009).

This effect is increased by the volume of the tumour mass and the peritumoral edema. The primary symptoms are headache, nausea, and vomiting in the morning

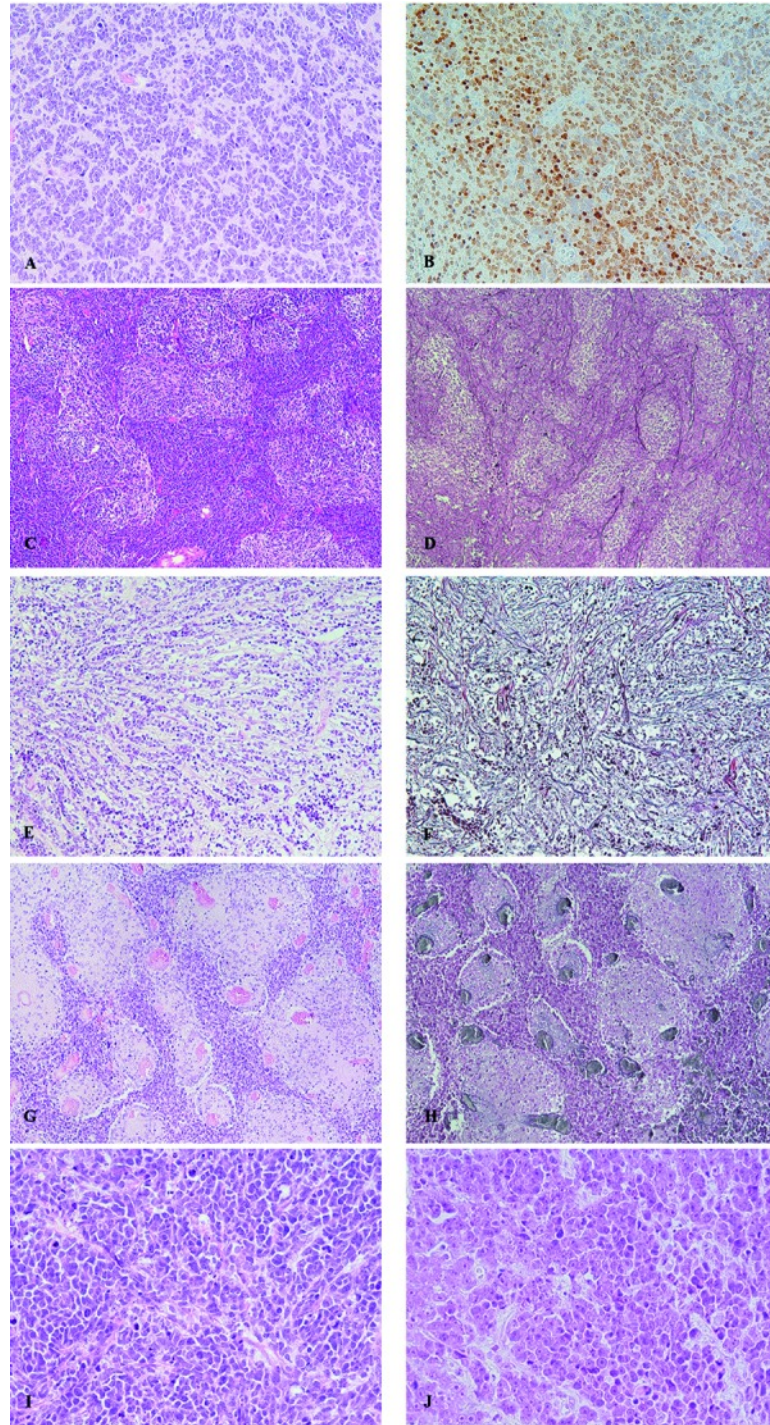


FIGURE 1.7. Histopathological subtypes of medulloblastoma. A: Classic medulloblastoma with (B) strong NeuN expression in preexisting granule cells and weaker expression in tumor cells (NeuN- neuron specific nuclear protein); C: desmoplastic/nodular medulloblastoma with (D) reticulin fibers in internodular areas (reticulin stain); E: classic medulloblastoma without pale nodular areas but with (F) desmoplastic reaction due to leptomeningeal invasion (reticulin stain); G: classic medulloblastoma with pale nodules but (H) without desmoplasia (reticulin stain); I, J: large/cell anaplastic medulloblastoma with (I) severely anaplastic nuclei with nuclear moulding/wrapping and frequent mitotic and apoptotic figures; J: large round cells with prominent nucleoli. Figure taken from Pietsch and Haberler, 2016.

before intake of food. The nausea often improves during the course of the day, and the headache may become less severe after vomiting (Frühwald and Rutkowski, 2011).

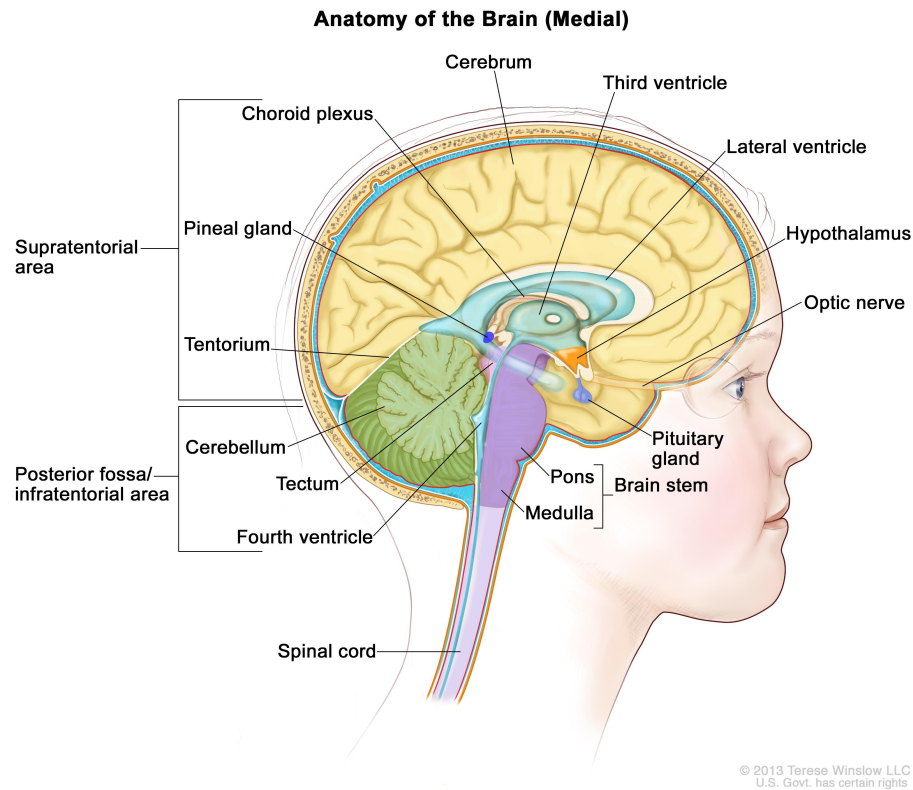


FIGURE 1.8. Anatomy of the inside of the brain, showing the cerebrum, cerebellum, brain stem, spinal cord, optic nerve, hypothalamus, and other parts of the brain. Figure taken from (NCI, 2018)

In infants under one year old, the symptoms are often nonspecific, such as lethargy or irritability, which often engender a delay in diagnosis, although a raised intracranial pressure often causes an increase in head size (Pizer and Clifford, 2009). In older children, behavioural changes and poorer academic performance are common, as are ataxia and cranial nerve palsies (Pizer and Clifford, 2009).

Magnetic resonance imaging (MRI) is preferable to computed tomography (CT) for evaluation of the anatomy due to higher resolution. For medulloblastoma, spinal MRI is indicated to detect metastases and should ideally be performed before surgery, otherwise together with the early postoperative follow-up imaging within 48 to 72 h (Frühwald and Rutkowski, 2011; Packer et al., 1999).

1.3.3. Medulloblastoma staging. Tumour staging and metastatic status are important factors in prognostication and are crucial for selecting appropriate treatment.

Staging of the tumour is usually determined by cranial and spinal MRI scans, ideally prior to the surgery, followed by cytological and immunohistochemical examination of CSF a minimum of 10-14 days after surgery (Frühwald and Rutkowski, 2011).

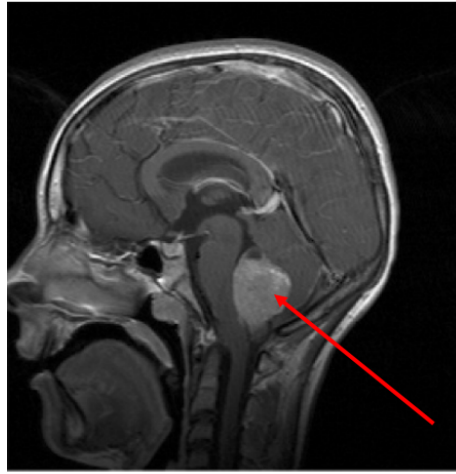


FIGURE 1.9. Medulloblastoma (indicated with red arrow). T1-weighted Sagittal MRI scan. Image kindly provided by Prof. Simon Bailey

Chang's classification (Chang et al., 1969) for the staging of medulloblastoma has been used with adaptations since 1969 and is shown in the table 1.2.

T Stage	Description	M Stage	Description
T1	Tumour <3 cm in diameter	M0	No evidence of subarachnoid or hematogenous metastasis
T2	Tumour >3 cm in diameter	M1	Tumour cells found in cerebrospinal fluid
T3a	Tumour >3 cm in diameter with extension	M2	Intracranial tumour beyond primary site
T3b	Tumour >3 cm in diameter with unequivocal extension into the brainstem	M3	Gross nodular seeding in spinal subarachnoid space
T4	Tumour >3 cm in diameter with extension up past the aqueduct of Sylvius and/or down past the foramen magnum (ie. Beyond the posterior fossa)	M4	Metastasis outside the cerebrospinal axis (especially to bone marrow, bone)

TABLE 1.2. Chang's medulloblastoma staging based on data from Pediatric Radiation Oncology, Fifth Edition (Marcus and Haas-Kogan, 2009);

1.3.4. Genetics and molecular classification of medulloblastoma.

1.3.4.1. *Familial syndromes* . The vast majority of medulloblastoma (about 95%) cases are sporadic (non-inherited). However, a small proportion of medulloblastoma cases co-occur with hereditary syndromes which have an increased risk of cancer. The syndromes known to be associated with medulloblastoma are shown in Table 1.3:

Basal Cell Nevus Syndrome, also known as Gorlin Syndrome,(Gorlin and Goltz, 1960), is an autosomal dominant condition characterised by the occurrence of basal cell carcinomas, together with skeletal abnormalities, odontogenic keratocysts and increased risk of medulloblastoma (occurring in about 5% of children with the syndrome).

Familial Cancer Syndrome	Gene	Locus	Cancer Association
Li-Fraumeni Syndrome	<i>P53</i>	17p13.1	medulloblastoma, other brain tumors, sarcomas, leukemia, breast cancer
Gorlin Syndrome: Nevoid basal cell carcinoma syndrome (NBCCS)	<i>PTCH</i> , protein: patched	9q22.3	medulloblastoma, basal cell skin carcinoma
Fanconi Anaemia	<i>BRCA2 FANCD2 PALB2 FANCC FANCA</i>	13q12.3 (type D1), 16p12.1 (type D2)	medulloblastoma, Wilms' tumour, neuroblastoma
Rubinstein-Taybi Syndrome	<i>CREBBP EP300</i>	16p13.3	medulloblastoma, other CNS tumours, neural crest tumour
Turcot Syndrome	<i>PMS2 MLH1 APC MSH2 MSH6</i>	5q21-22	medulloblastoma, multiple colorectal adenoma

TABLE 1.3. Familial cancer syndromes caused by loss of function of a TSG based on data from (Gajjar et al., 2006).

It is associated with germline mutation of the tumor suppressor gene *PTCH1*, which encodes a protein involved in SHH pathway (Cambruzzi, 2018).

Turcot Syndrome is characterised by malignant tumors of the central nervous system (mostly astrocytomas and medulloblastoma) associated with familial polyposis of the colon. There are different sub-types characterised by various mutations (Paraf et al., 1997). The most important one, germline mutation of the TSG *APC* is involved in WNT pathway activation. APC protein binds β -catenin, acting as antagonist of the WNT signalling pathway (Rubinfeld et al., 1993). About 79% of children with the syndrome develop medulloblastoma.

Li-Fraumeni syndrome is a rare inherited autosomal dominant disorder that increases the risk of developing several types of cancer, particularly in children and young adults, including medulloblastoma (in about 2% of patients) and is characterised by mutations in the *TP53* gene. The most frequent types of cancer associated with Li-Fraumeni syndrome are breast cancer, osteosarcoma, and soft tissue sarcomas. *TP53* is a transcription factor that activates in response to DNA damage and regulates cell cycle and apoptosis and inactivation of this gene leads to immortalisation of cancer cell (Malkin, 2011).

The mutations identified during the studies of familial syndromes cases associated with WNT and SHH pathways provided an insight into the way these pathways drive tumorigenesis in medulloblastoma by aberrant activation of the WNT and SHH signalling pathways.

1.3.5. Embryonic signaling pathways and their role in cerebellar development.

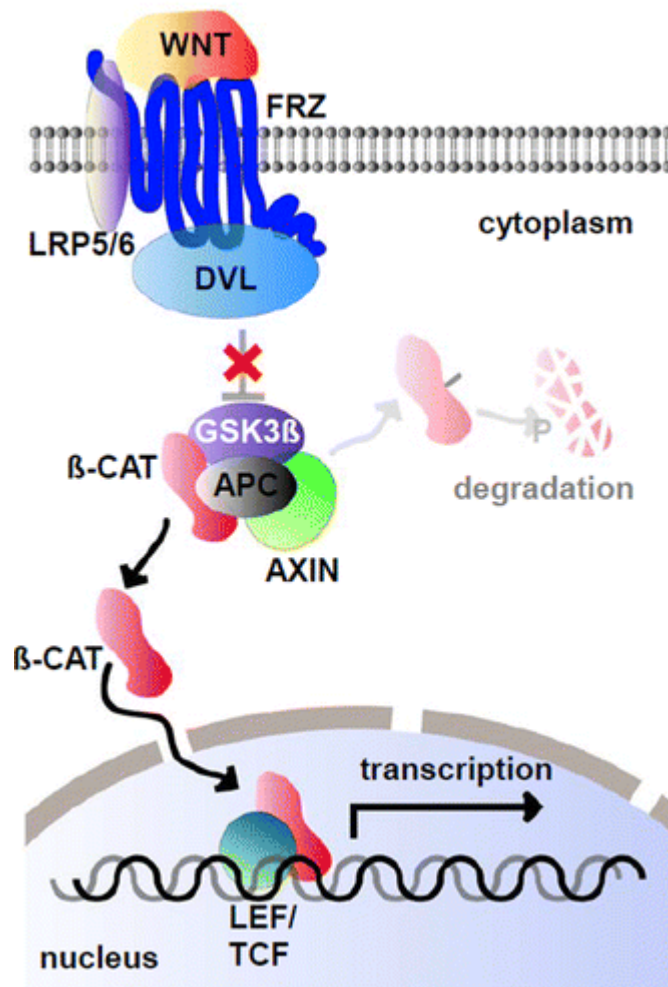


FIGURE 1.10. Scheme of active WNT pathway taken from Neumann et al. (2017)

1.3.5.1. *WNT pathway: role in cerebellar development and medulloblastoma*. The Wnt signalling pathway (Fig.1.10) is involved in numerous fundamental processes essential for embryonic development and normal adult homeostasis (Neumann et al., 2017). The WNT/wingless gene, one of the main regulators of *Drosophila melanogaster* segment polarity was discovered in a mutagenesis screen for visual phenotypes, affecting various developmental patterning processes in *Drosophila*. The first member of the Wnt family in mammals was initially discovered as the proto-oncogene *Int-1* in mice

and later found to be the homologue of the wingless gene of *Drosophila melanogaster*, which had already been described (Nusse and Varmus, 1992).

The pathway is generally divided into three sub-pathways:

- (1) canonical,
- (2) non-canonical planar cell polarity (PCP) pathway, and,
- (3) non-canonical Wnt/calcium pathway.

All three pathways are activated by binding a Wnt-protein ligand to a Frizzled family receptor, which passes the biological signal to the Dishevelled protein inside the cell. The canonical Wnt pathway leads to regulation of gene transcription, and is thought to be negatively regulated in part by the *SPATS1* gene (Zhang et al., 2010). The non-canonical planar cell polarity pathway regulates the cytoskeleton that is responsible for the shape of the cell and non-canonical Wnt/calcium pathway regulates calcium inside the cell (Nusse and Nusse, 2005).

The most widely studied functions of WNT signalling is in the establishment of the midbrain-hindbrain boundary that gives rise to the cerebellum (McMahon and Bradley, 1990).

1.3.5.2. *SHH pathway: role in cerebellar development and medulloblastoma* . The hedgehog pathway is a signalling pathway that plays an important role in embryonic development, regulating cell differentiation, proliferation and polarity (De Luca et al., 2016). It was initially discovered in 1980 by Nusslein-Volhard and Weischaus during their large-scale mutation screening of *Drosophila* fruit fly (Nüsslein-volhard and Wieschaus, 1980). The vertebrate homologues of the *Drosophila* hedgehog gene are Sonic Hedgehog (*SHH*), Desert (*DHH*) and Indian (*IHH*). The most studied, SHH, is found to be expressed from early embryogenesis and has been established as one of the key molecules responsible for the regulation of central nervous system (CNS) patterning (Machold and Fishell, 2002). During CNS organogenesis, SHH protein plays key roles as a morphogen, mitogen, and guidance molecule (Fuccillo et al., 2006; Yam and Charron, 2013).

The SHH pathway is activated in two major ways:

- (1) canonical signaling: by ligand-dependent interaction or through receptor-induced signaling

- (2) (2) non-canonical signalling, when SHH signalling pathway is controlled by Suppressor of Fused (SUFU) in the absence of SHH ligand.

In the canonical activation of SHH pathway, activation occurs by ligand-dependent interaction when SHH binds to PTCH1 at the cell membrane. In response to this binding, PTCH1 no longer inhibits SMO, which accumulates at the primary cilium (PC) and initiates the downstream signalling pathway cascade. SMO regulates GLI1 processing and activation at the PC. When GLI1 is activated, it translocates to the nucleus, where it activates SHH target genes, see Figure 1.11 (Carlotti et al., 2008; Neumann et al., 2017).

In the absence of ligand, SUFU negatively regulates the pathway by directly binding to the GLI1 transcription factor, inhibiting its translocation to the nucleus, and preventing pathway activation (Kogerman et al., 1999).

Non-canonical SHH signalling usually occurs through GLI1-independent mechanisms of the following types: Type I is downstream of SMO, which modulates Ca²⁺ ions and the actin cytoskeleton; type II is independent of SMO and increases cell proliferation and survival (Robbins et al., 2012). The non-canonical SHH signalling pathway can regulate chemotaxis and cell migration through actin rearrangement. Mutations in PTCH1 and SUFU, which are negative regulators of SHH signalling, are linked to tumorigenesis, with important evidence being a link between Gorlin's syndrome (see section 1.3)and basal cell carcinoma (Archer et al., 2012), which is dependent upon aberrant SHH pathway activation. Non-canonical Shh signalling has been linked to the induction of chromosomal instability (Szczepny et al., 2017).

During embryonal developmental stages, SHH signalling is involved in the formation of several CNS regions, including the cerebellum, where it regulates cerebellar progenitor proliferation and maturation and is critical during initial phases of territorial determination (De Luca et al., 2016). Also, SHH signalling affects Bergmann glial differentiation and promotes cerebellar foliation during development (Sudarov and Joyner, 2007).

Given its prominent role during development, alterations of its physiological functions are implicated in many human cerebellar pathologies, such as ataxias (Lim et al., 2006), Joubert syndrome, and medulloblastoma (Vaillant and Monard, 2009).

Medulloblastomas with activated SHH signalling pathway represent between one quarter and one third of all medulloblastoma cases (Crawford et al., 2007; Northcott et al., 2012a) with the majority being infant (*i.e.* under 3 years old) and adult cases, often demonstrating a desmoplastic phenotype (Northcott et al., 2011; Remke and Hielscher, 2011b). Medulloblastomas with activated SHH pathway mostly harbour mutations or aberrations in SHH pathway components such as *PTCH1*, *SUFU*, *SMO* and *GLI*, or *TP53* mutations (Sengupta et al., 2017).

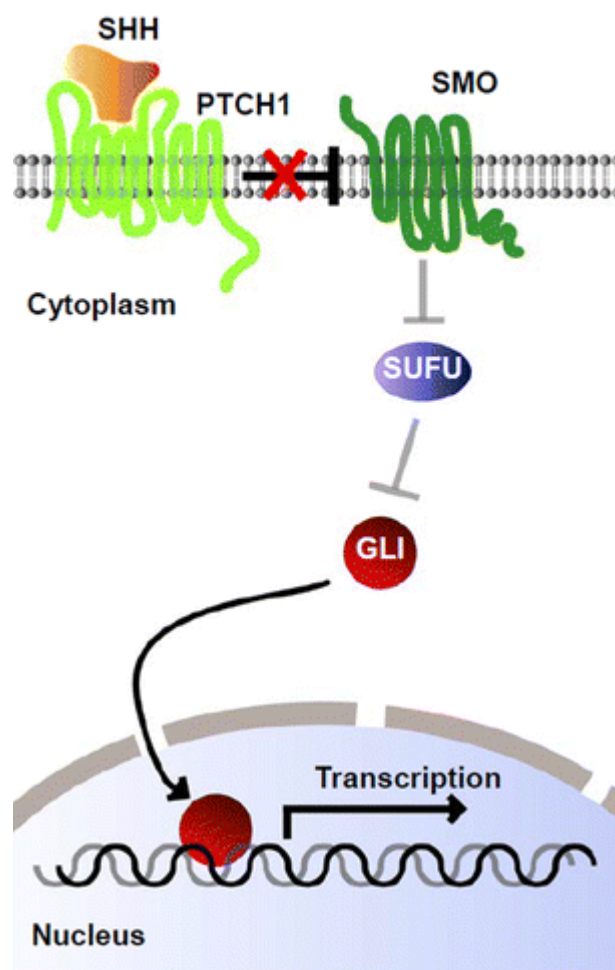


FIGURE 1.11. Scheme of active SHH pathway taken from (Neumann 2017)

1.3.6. Subgroups of medulloblastoma: Current consensus. The 2016 WHO CNS tumour classification (Louis et al., 2016) defines main molecular subgroups of medulloblastoma as follows:

- WNT-activated medulloblastoma – medulloblastoma characterised by activation of the Wnt/Wingless embryonic signalling pathway;

- SHH-activated medulloblastoma (with subtypes *TP53*-mutant and *TP53*-wild type) – medulloblastoma with activation of the Sonic Hedgehog embryonic signalling pathway;
- Non-WNT/non-SHH, group 3 medulloblastoma
- Non-WNT/non-SHH, group 4 medulloblastoma

Consensus review of several transcriptional profiling studies of medulloblastoma cohorts from various research groups suggested the existence of at least four molecular subgroups named WNT, SHH, Group 3 and Group 4, distinct in their demographics, transcriptomes, somatic genetic events, and clinical outcomes (Taylor et al., 2012). These are summarised in Figure 1.12. Additionally, it was accepted that there was preliminary evidence for the existence of subtypes within the subgroups, particularly for Group 3. These subtypes of subgroups were not well characterised at the time and the consensus was to name them using Greek letters (α, β, γ , etc.) until additional characterisation was available using larger cohorts.









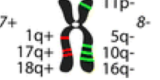
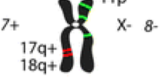
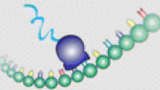
Molecular Subgroups of Medulloblastoma				
CONSENSUS	WNT	SHH	Group 3	Group 4
Cho (2010)	C6	C3	C1/C5	C2/C4
Northcott (2010)	WNT	SHH	Group C	Group D
Kool (2008)	A	B	E	C/D
Thompson (2006)	B	C, D	E, A	A, C
DEMOGRAPHICS				
Age Group: 				
Gender: ♀ ♂	♂♂ : ♀♀	♂♂ : ♀♀	♂♂ : ♀	♂♂ : ♀
CLINICAL FEATURES				
Histology	classic, rarely LCA	desmoplastic/nodular, classic, LCA	classic, LCA	classic, LCA
Metastasis	rarely M+	uncommonly M+	very frequently M+	frequently M+
Prognosis	very good	infants good, others intermediate	poor	intermediate
GENETICS				
	 CTNNB1 mutation	 PTCH1/SMO/SUFU mutation GLI2 amplification MYCN amplification	 i17q MYC amplification	 i17q CDK6 amplification MYCN amplification
GENE EXPRESSION				
	WNT signaling MYC +	SHH signaling MYCN +	Photoreceptor/GABAergic MYC +++	Neuronal/Glutamatergic minimal MYC / MYCN

FIGURE 1.12. Comparison of the various subgroups of medulloblastoma including affiliations with previously published papers on medulloblastoma molecular subgrouping (Taylor et al., 2012)

WNT subgroup

The WNT medulloblastoma subgroup is named for the Wingless (WNT) signalling pathway activated in this subgroup of tumours. These tumours are the least common of the medulloblastoma subgroups, accounting for 11% of cases, with a gender ratio of about 1:1. The peak age of diagnosis for WNT tumours is between 6-10 years of age (Northcott et al., 2011). WNT patients have a very favourable prognosis. Over 95% of patients survive for more than 5 years (Ellison et al., 2011) which could be related to increased sensitivity to radiotherapy (Salaroli et al., 2015); a 2016 study suggested that WNT-activated medulloblastomas have a disrupted blood-brain barrier, which facilitates greater access to the tumour for chemotherapy. (Phoenix et al., 2016). The majority of WNT medulloblastoma have classic histology, with very rare documented cases of large cell/anaplasia. WNT subgroup tumours are rarely metastatic. Germline mutations of the WNT pathway inhibitor *APC* are seen in Turcot syndrome, whose symptoms include a predisposition to medulloblastoma. Fewer than 5% of medulloblastoma patients are carriers of this mutation (Huang et al., 2009), however studying Turcot syndrome patients affected by medulloblastoma established the importance of WNT pathway activation in medulloblastoma. Somatic missense mutations in exon 3 of *CTNNB1* (which encodes β -Catenin) are present in more than 90% of WNT medulloblastomas (Northcott et al., 2012a). These mutations result in nuclear accumulation of β -catenin and subsequent dysregulation of WNT target genes. WNT pathway tumours are typically characterised by deletion of one copy of chromosome 6 (monosomy 6) in the majority of patients (79%) (Shih et al., 2014). Other than monosomy 6, the genome of WNT medulloblastoma has relatively few chromosomal gains and/or losses across the genome (Gilbertson and Ellison, 2008).

Copy number variation (CNV) and/or single nucleotide variants (SNV) include (in addition to monosomy 6) mutations in the DEAD-box RNA helicase gene, *SMARCB4* (26%), *TP53* mutation (16%), tetraploidy (14%), and *KMT2D* (also known as *MLL2* or *MLL4*) mutations (2%) (Northcott et al., 2012a; Jones et al., 2012).

Sonic hedgehog subgroup

The Sonic hedgehog (SHH) subgroup is also named for the signalling pathway that is disrupted in these tumours. The SHH subgroup of medulloblastoma comprises 25% of all cases (Northcott et al., 2012a). SHH tumours have a bimodal age distribution and they affect mostly the infants and adult patients with medulloblastoma, but less

frequently children of 3-16 years old (Northcott et al., 2011; Taylor et al., 2012; Gibson et al., 2010). The gender ratio is about 1:1, although there is a slight male predominance among infants (Gajjar and Robinson, 2014). Overall survival for the group is 70% (Cho et al., 2011). Histology is typically of the nodular/desmoplastic type with MBEN being exclusively classified into this group. Remaining tumours are either of classic or LC/A histology. SHH medulloblastomas can be identified using tumour IHC expression for GAB1, SFRP1, and GLI1 proteins (Ellison et al., 2011). Metastatic disease at diagnosis occurs in the SHH subgroup mostly in infants (17%) and children (22%) but is uncommon in adults (Kool et al., 2012). About 3-5% of patients with Gorlin Syndrome with a germline *PTCH1* mutation exclusively develop nodular/desmoplastic medulloblastoma.

Group 3

Group 3 medulloblastoma accounts for 25-28% of cases and it is almost exclusively found in infants and younger children, with extremely rare cases recorded in adults (Remke and Hielscher, 2011a) with a male predominance of about 2:1. These tumours present with high incidence of metastatic disease at diagnosis (Ramaswamy and Taylor, 2017), and frequently have a LC/A histology (Northcott et al., 2012a). This subgroup has the worst outcome with less than 50% five years survival.

Somatic nucleotide variants are rare in this group, and Group 3 medulloblastomas are considered copy-number driven with highly unstable genomes (Northcott et al., 2012a). Focal high-level amplification of *MYC* (12-16%) and *OTX2* (7%) proto-oncogenes and frequent gains of 1q, 7, and 17q (i17q) are observed along with 10q, 11, 16q, and 17p deletions (Northcott et al., 2012a). Chromothripsis (in the absence of *TP53* mutations) is frequently seen, likely as result of an attempt at ineffective DNA repair (Schroeder and Gururangan, 2014). Tetraploidy is seen in 54% of group 3 tumours and probably occurs as an early event in tumorigenesis (Jones and Jallepalli, 2012).

Group 4

This type of medulloblastoma is the most common (about 35% of all cases), with significantly more males affected ($\sim 3:1$ ratio) (Northcott et al., 2012a). This type of medulloblastoma is predominant in the 3-16 year old age group, and rarely observed in children under 3. About 30% of Group 4 medulloblastomas are metastatic on presentation. The majority of Group 4 cases are of classic histology, although cases of

large-cell/anaplastic tumours have been observed, with an intermediate prognosis (\sim 75% overall 5 years survival) (Northcott et al., 2012a). This group is also considered to be a copy number-driven tumour, with rare SNVs. The most common chromosomal aberrations are i17q (isochromosome, simultaneous loss of p arm and gain of q arm, in almost 80% of tumours), and, less commonly, 1q, 7, 17q, 18q gain and 8, 10q, 11, 16p, 17p loss. Chromosome X loss occurs in 80% of affected females. Similarly to Group 3, tetraploidy occurs in 40% of cases.

	WNT	SHH	Group 3	Group 4
Histology	Classic, Rarely LCA	Desmoplastic, Classic, LCA	Classic, LCA	Classic, LCA
Metastatic rate at diagnosis	Very low	Low	High	High
Prognosis	Excellent	Intermediate	Poor	Intermediate
Focal copy number aberrations		<i>MYCN</i> (12%) <i>GLI2</i> (8%)	<i>MYC</i> (17%) <i>PVT1</i> (12%) <i>OTX2</i> (8%)	<i>SNCAIP</i> (10%) <i>MYCN</i> (6%) <i>CDK6</i> (5%)
Broad copy number aberrations	6 Loss	3q Gain 9q, 10q, 14q Loss	1q, 7, 17q, 18q Gain 8, 10q, 11, 16p, 17p Loss	7, 17q, 18q Gain 8, 11p, X Loss
SNVs	<i>CTNNB1</i> (91%) <i>DDX3X</i> (50%) <i>SMARCA4</i> (26%) <i>MLL2</i> (13%) <i>TP53</i> (13%)	<i>TERT</i> (60%) <i>PTCH1</i> (46%) <i>SUFU</i> (24%) <i>MLL2</i> (16%) <i>SMO</i> (14%) <i>TP53</i> (13%)	<i>SMARCA4</i> (11%) <i>MLL2</i> (4%)	<i>KDM6A</i> (13%) <i>MLL3</i> (5%)
Expression	WNT signaling	SHH signaling	<i>MYC</i> /Retinal Signature	Neuronal Signature
Pattern of relapse	Local and metastatic	Predominantly local	Metastatic	Metastatic

TABLE 1.4. **Clinical and genomic characteristics of medulloblastoma subgroups, adapted from (Ramaswamy and Taylor, 2017)**

1.3.7. Heterogeneity within medulloblastoma subgroups. Current international consensus (Taylor et al., 2012) recognises four molecular subgroups of medulloblastoma: WNT, SHH, G3 and G4 - in this consensus, further heterogeneity within these subgroups was recognised. WNT and SHH subgroups are defined by activation of the WNT and SHH signalling pathways, often in association with specific mutations and copy number aberrations. In contrast, Group3 and Group4 medulloblastomas have fewer mutations and multiple DNA copy number aberrations (Jones and

Jallepalli, 2012; Northcott et al., 2012a), and an overlap between these two groups is a limitation of current molecular subtyping of MB with either DNA methylation profiling or gene expression analysis alone. This is manifested in the 2016 WHO classification of CNS tumors, which defines Group 3 and Group 4 subgroups as a single entity, "non-WNT/non-SHH MB" (Louis et al., 2016).

Further studies have demonstrated that there is a wide variation in patient outcomes within each subgroup (Ramaswamy et al., 2016a; Zhukova et al., 2013) and also there is a need for new subgrouping and risk stratification schemes to accommodate the fact that patients from the same molecular subgroup can fall into different risk groups, based on modulation by additional clinical or molecular features.

In 2017, three independent studies were published, which investigated clinical and molecular heterogeneity of medulloblastoma subgroups using larger cohorts of patients.

A study, conducted by Schwalbe and colleagues (Schwalbe et al., 2017b), identified seven robust and reproducible primary molecular subgroups of childhood medulloblastoma. WNT-activated medulloblastoma remained unchanged, whereas each remaining consensus subgroup was split in two. SHH medulloblastoma was split into age-dependent subgroups corresponding to infant and childhood patients, Group3 and Group4 were each split into high-risk and low-risk subgroups. These seven subgroups reveal biological overlap between Group3 and Group4, where the high/low-risk group is defined by a common biological signature. It was shown that these novel subgroups can be further subdivided according to secondary clinico-pathological and molecular features. These seven groups along with the secondary features and known disease risk-factors were used to build survival models, that stratified patients into four clinical risk groups for 5-year progression-free survival: favourable risk (91% 5-year PFS), standard risk (81% 5-year PFS), high-risk (42% 5-year PFS) and very high-risk (28% 5-year PFS). This new stratification scheme outperformed existing disease risk-stratification schemes (Schwalbe et al., 2017b).

Another study by Cavalli and colleagues used Similarity Network Fusion (SNF) to analyse genome-wide gene expression and DNA methylation data of 763 medulloblastomas and identified 12 medulloblastoma subtypes: two within the WNT subgroup, four in SHH subgroup and three in each Group3 and Group4 (Cavalli et al., 2017). In contrast to the study by Schwalbe et al., this study reported a clear distinction

between Group3 and Group4 tumours, and showed that the set of overlapping Group 3 and Group 4 tumours identified by DNA methylation profiling was different than that identified using gene expression analysis, suggesting that this overlap was due to testing modality, rather than biology.

The third study, by Northcott and colleagues, analysed 740 Group3/4 medulloblastomas using methylation arrays and identified subtypes using t-Distributed Stochastic Neighbor Embedding (t-SNE), a dimension reduction technique, coupled with the DBSCAN (density-based clustering of applications with noise) algorithm to identify clusters. They reported eight subtypes (I-VIII), some comprising pure Group3/4 tumours, others with mixed subgroups. (Northcott et al., 2017).

In order to resolve the inconsistencies in nature of the subtypes and their numbers between these studies, another study was conducted with the aim to characterise the number and nature of Group 3 and Group 4 medulloblastoma subtypes in an unbiased way. The same analytical techniques and approaches of the component studies were applied to a combined dataset from the cohorts from all three publications and an additional 153 tumours. This analysis most strongly supported a definition of eight robust subtypes of Group 3/Group 4 (subtypes I-VIII) with additional evidence that there are two variants of subtype VII, VII-A and VII-B (Sharma et al., 2019).

Subtype-specific survival was assessed using all available PFS (n=550) and OS (n=837) data. However, as progression free survival data was unavailable for the Cavalli et al. (2017) dataset, the eight Group 3 and Group4 subtypes were risk stratified into three distinct groups based on the overall survival.

The first, a very high-risk group (5-year OS 50%, 95% CI 43-58) encompasses subtypes II (5-year OS 50%, 95% CI 40-62), III (43%, 95% CI 32-59), and V (5-year OS 59%, 95% CI 46-75).

The second group consists solely of subtype VIII (5-year OS 81%, 95% CI 75- 87) and is associated with a feature unique to this subtype, late relapse/death (35% of death within this subtype occurred ≥ 5 years after diagnosis).

The third group is standard risk group (5-year OS 82%, 95% CI 78-87), comprises of subtypes I (5-year OS 77%, 95% CI 62-97), IV (80%, 95% CI 70-91), VI (81%, 95% CI 72-91), and VII (85%, 95% CI 79-91).

Equivalent consensus studies that define subtypes of subgroups for SHH and WNT disease have not yet been reported.

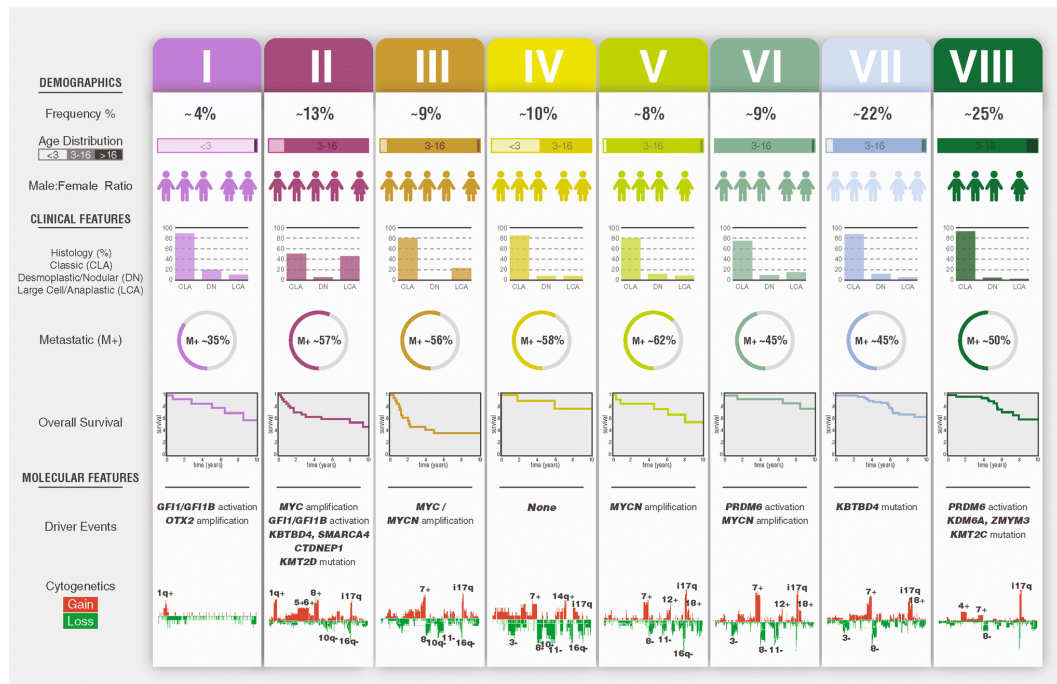


FIGURE 1.13. Summary of molecular subtypes of Group 3/4 medulloblastoma. The major demographic, clinico-pathological, and molecular features of the concordant subtypes are summarised. CLAS = classic, DN = desmoplastic nodular, LCA = large-cell anaplastic, M+ = metastatic. Overall survival shows subtype-specific survival in years. Copy number gains are shown in red, losses in green. Figure taken from Sharma et al. (2019).

Summary of molecular subtypes of Group 3 and 4 medulloblastomas is presented on Figure 1.13.

This novel subtyping is an important step to understanding the nature of medulloblastoma and personalised treatment, however, it is still too early for clinical application, since it is problematic to recruit sufficient patients for clinical trials to assess the prognostic meaning of the subtypes, due to the large number of the subtypes and the rarity of the disease.

1.3.8. Origins of medulloblastoma. Medulloblastomas are thought to arise from cerebellar stem cells that have been prevented from dividing and differentiating into their normal cell types (Colditz, 2014). This accounts for the varying histological variants seen on biopsy. Establishment of a series of genetically engineered mouse (GEM) models of medulloblastoma have provided important insights into the cellular origins of these tumours (Table 1.5)

Medulloblastoma type	GEM model	Cells of origin	References
WNT	Ctnnb1 ^{lox(ex3)} , Trp53 ^{flx/flx}	Dorsal brainstem neuronal progenitors	(Kawauchi et al., 2012)
SHH	Ptch loss or ND2-SmoM2	External granular layer, rhombic lip granular neuroprogenitors	(Yang et al., 2008; Gibson et al., 2010)22
Group 3	MycT58A, Trp53+/-	Ventricular zone neuronal progenitors, or external granular layer and rhombic lip granular neuroprogenitors	(Pei et al., 2012; Swartling et al., 2010)
Group 4	Glt1-MYCN	Unknown, possibly cerebellum progenitors or stem cells, or upper rhombic lip	(Swartling et al., 2012; Singh et al., 2004; Lin et al., 2016)

TABLE 1.5. Cells of origin of medulloblastomas (MBs) and a summary of the related genetically engineered mouse (GEM) model. Adapted from Huang et al. (2016)

Granule neuron progenitors in the external germinal layer on the surface of the developing cerebellum have long been proposed as putative cells of origin for MB, and it is now clear that SHH-driven MB arises from this group (Schüller et al., 2008; Yang et al., 2008). The WNT group is now thought to originate from progenitors in the dorsal brainstem (Gibson et al., 2010). Group 3, which is the most aggressive subgroup of MB and is characterised by a higher incidence of *MYC* amplification and/or *MYC* overexpression, appears to derive from cerebellar stem cells (Kawauchi et al., 2012; Pei et al., 2012). The cellular origin of Group 4, the most prevalent MB subgroup, remains unknown, however a recent study reports that upper rhombic lip (uRL) and its derivative precursors are putative cell-of-origin for this group (Lin et al., 2016).

1.3.9. Prognostic factors and current risk stratification in clinical settings. There are a number of factors that have been shown to be important for determining patient prognosis in medulloblastoma. Current risk stratification schemes are based on consideration of the age of the patient (in particular, infant patients under 3 years of age at diagnosis), metastatic spread and residual disease following surgical removal of the tumour. Current risk markers are summarised in Table 1.6

Risk factor marker	Reference
Favourable risk	
DN/MBEN histology in patients <3 years of age	(McManamy et al., 2007; Rutkowski et al., 2010)
WNT pathway activation (β -catenin nuclear accumulation)	(Ellison et al., 2005; Clifford et al., 2006)
Adverse risk	
Age <3 years of age	(Zeltzer et al., 1999; Rutkowski et al., 2010)
Metastatic disease	(Bailey et al., 1995; Zeltzer et al., 1999)
Post-surgical residual disease ≥ 1.5 cm ²	(Zeltzer et al., 1999; Tamayo et al., 2011)
Large cell / anaplastic medulloblastoma	(Eberhart et al., 2003; McManamy et al., 2003)
<i>MYC</i> family gene amplification	(Lamont et al., 2004; Eberhart et al., 2003)

TABLE 1.6. Currently accepted prognostic factors in medulloblastoma, validated in clinical trials PNET3, SJMB96, HIT-SKK'92, CNS9204 and HIT '91. Markers of favourable and adverse risk are listed, with supporting references included. Table adapted from Pizer and Clifford (2009);

The higher risk prognosis in infants can partially be attributed to a consequence of the avoidance of radiotherapy in these cases, because of the unacceptable neurocognitive consequences associated with cranial radiotherapy applied to a small, developing brain (Crawford et al., 2007; Robinson et al., 2018). Due to these serious *sequelae*, therapies in infant cases are usually designed to avoid or delay the use of cranio-spinal radiotherapy. A recent clinical trial (NCT00602667, ClinicalTrials.gov) identified a low-risk subtype of SHH-activated medulloblastoma that had improved 5 year progression-free survival (90.9%) in the absence of radiotherapy, intraventricular therapy or high-dose chemotherapy and was characterised by SMO mutations enrichment (Robinson et al., 2018).

1.3.9.1. *High-risk factors and standard risk disease.* For treatment purposes, patients with medulloblastoma are divided into two prognostic groups: children over 3 years of age with non-metastatic disease (Chang stage M0) and minimal residual disease ($< 1.5\text{cm}^2$) post-operatively comprise the standard-risk group. Other patients with risk-factors such as a sub-total resection *MYC* amplification or metastatic disease and younger patients below 3 years of age comprise the high-risk group (Table 1.7) (Parkes et al., 2015). Therefore, by definition, standard-risk medulloblastoma group comprises patients with an absence of risk-factors.

Standard Risk	High Risk
All of the following: >3 years of age <1.5 cm ² residual tumour after resection CSF negative for tumour cells on L.P. MRI spine negative for leptomeningeal spread Classic or desmoplastic subtypes on pathology Complete staging possible Absence of <i>MYC(N)</i> gene amplification	Any one of the following: <3 years of age Subtotal resection (>1.5 cm ² residual) CSF positive for tumour cells MRI shows leptomeningeal spread Large cell or anaplastic subtype Incomplete staging <i>MYC(N)</i> gene amplification

TABLE 1.7. Clinical risk stratification of medulloblastoma adapted from Parkes et al. (2015)

Several clinical trials from SIOP, COG and St. Jude’s have shown that outcomes for average / standard risk medulloblastoma exceed 80% five-year survival (Gajjar et al., 2006; Lannering et al., 2012; Packer et al., 2006). In standard-risk medulloblastoma, relapse is the major cause of death, occurring in around 20% of patients and the prognosis at relapse is poor, with generally less than 10% survival (Hill et al., 2015). Therefore, standard risk medulloblastoma is a heterogeneous disease in itself.

1.3.10. Risk Stratification within subgroups . The current clinical risk stratification divides patients over aged over three into standard-risk and high-risk. Here, high-risk is defined by positivity for subtotal-resection and/or metastatic disease in North America and Australia (Ellison et al., 2011). In Europe, in addition to the above, patients with large-cell and/or anaplastic pathology and/or *MYC* or *MYCN* gene amplification are excluded from standard risk trials (e.g. the PNET5 study), based on clinical, histopathological and biological studies from previous trials (e.g. PNET3) (Ellison et al., 2011). However, several studies over the past few years have identified molecular markers, which, together with previously used criteria, may provide additional information for improved risk stratification (Clifford et al., 2015; Pietsch et al., 2014; Schwalbe et al., 2013; Shih et al., 2014; Tamayo et al., 2011). Several studies were reviewed and a consensus was reached regarding a new proposed risk classification scheme (Table 1.8, Ramaswamy et al. (2016b)). A recent study subsequently identified novel risk subgroups of medulloblastoma, describing for the first time that non-WNT patients could be assigned to a favourable risk group (i.e. >90% survival) that could be potential candidates for treatment de-escalation (Goschzik et al., 2018).

Favourable Risk	Standard Risk	High Risk	Very High Risk
(>90% survival rate)	(75-90% survival rate)	(50-75% survival rate)	(<50% survival rate)
WNT Non-metastatic Group 4, with chr 11 loss or chr 17 gain	The rest of the patients in the absence of mentioned risk markers	Metastatic SHH or Group 4 <i>MYCN</i> -amplified SHH	Group 3 with metastases or SHH with <i>TP53</i> mutation

TABLE 1.8. Proposed risk stratification in childhood non-infant medulloblastoma (Ramaswamy et al., 2016b);

1.3.11. Epigenetic drivers in medulloblastoma . Epigenetic factors play important role in development of many cancers, including medulloblastoma, which exhibits high rates of alteration in epigenetic regulators across all four subgroups (Jones et al., 2016; Dubuc et al., 2013) (see Table 1.9). Over 30% of all medulloblastoma samples carry various mutations, deletions and amplifications and each subgroup displaying both common and distinct epigenetic properties (Jones and Jallepalli, 2012; Dubuc et al., 2013; Batora et al., 2014). The later suggests that epigenetic changes can be important drivers of medulloblastoma. Research suggests that the main, although not the only function of epigenetic factors, is transcriptional regulation and each medulloblastoma subtype may require different sets of histone modifiers and chromatin remodellers that regulate gene expression (Yi and Wu, 2018).

Genes	Epigenetic function of gene products	Recurrent abnormalities in MB
<i>Brg1/SMARCA4</i>	Chromatin remodeler	Mutation/WNT and Group 3
<i>BMI1</i> <i>EZH2</i> <i>MLL2/KMT2D</i> <i>MLL3/KMT2C</i> <i>G9A/EHMT2</i> <i>CREBBP</i> <i>PCAF/KAT2B</i>	Writer Subunit of PRC1 H2AK119 E3 ubiquitin ligase Subunit of PRC2 H3K27 methyltransferase H3K4 methyltransferase H3K4 methyltransferase H3K9 methyltransferase H3K27 acetyltransferase Histone acetyltransferase	Overexpression/across subgroups High expression/Groups 3 and 4 Mutation/across subgroups Mutation/across subgroups Not known Mutation/SHH Not known
<i>UTX/KDM6A</i> <i>JMJD3/KDM6B</i> <i>LSD1/KDM1A</i> <i>HDAC1/2</i> <i>SIRT1</i>	Eraser H3K27me3 demethylase H3K27me3 demethylase H3K9me1/2 demethylase Histone deacetylase NAD-dependent histone deacetylase	Mutation or deletion/Group 4 Not known Not known Not known High expression/multiple subgroups
<i>BRD2/3/4</i>	Reader Recognition of acetyl-lysine on histones	Not known

TABLE 1.9. Summary of genes encoding epigenetic regulators in medulloblastoma. Adapted from (Yi and Wu, 2018)

1.3.11.1. *DNA methylation in medulloblastoma.* Aberrant changes in DNA methylation is a common feature of tumorigenesis (Jones and Baylin, 2007b). Genome-wide studies of DNA methylation patterns in medulloblastoma have suggested that DNA

methylation plays a major role in its pathogenesis by repressing genes to avoid cell differentiation and cell death (Dubuc et al., 2013).

In 2013, four medulloblastoma subgroups were identified using DNA methylation microarrays; these subgroups were highly related to the previously described transcriptome subgroups. This suggested that each subtype of medulloblastoma has specific methylation patterns which can be used for robust subclassification (Schwalbe et al., 2013).

Bisulfite sequencing analysis combined with matched RNA-seq and ChIP-seq data provided a comprehensive view of DNA methylation and gene expression profiles in medulloblastoma and established a high incidence of regions of hypomethylation tens of kbp downstream of gene promoters, which correlated with gene overexpression (Hovestadt et al., 2014).

The above two studies both used bisulfite conversion to uncover cytosine methylation, during which unmethylated cytosines in CpG islands are converted to uracil and methylated cytosines are not converted. This technique allows methylation pattern to be read. Bisulfite conversion, followed by polymerase chain reaction (PCR) and bisulfite sequencing of specific regions, identified several tumour suppressor genes silenced in medulloblastoma by promoter hypermethylation. These included *CDKN2A*, *HIC1* and *RASSF1* (Lindsey et al., 2004; Shahi et al., 2010, 2011; Diede et al., 2010). Other techniques, such as array-based profiling of reference-independent methylation status (aPRIMES) identified *PTCH1*, the negative regulator of SHH signaling, transcriptional repressor *ZIC2*, *SFRP* family of the WNT signaling pathway inhibitors (Pfister et al., 2007).

1.3.11.2. *Histone modifications in medulloblastoma.* Extensive molecular analysis identified mutations and copy number aberrations (Northcott et al., 2009) in multiple epigenetic regulators that play role in changes chromatin state and gene expression in medulloblastoma (see Fig1.14) such as histone lysine methyl and acetyl transferases ("writers"), demethylases ("erasers") and polycomb group of transcriptional repressors (PRC1 and PRC2, "readers") (Roussel and Stripay, 2018).

Genome-wide sequencing of medulloblastoma identified disregulated chromatin regulation pathways involving H3K4 and H3K27 methylation (Fig. 1.15).

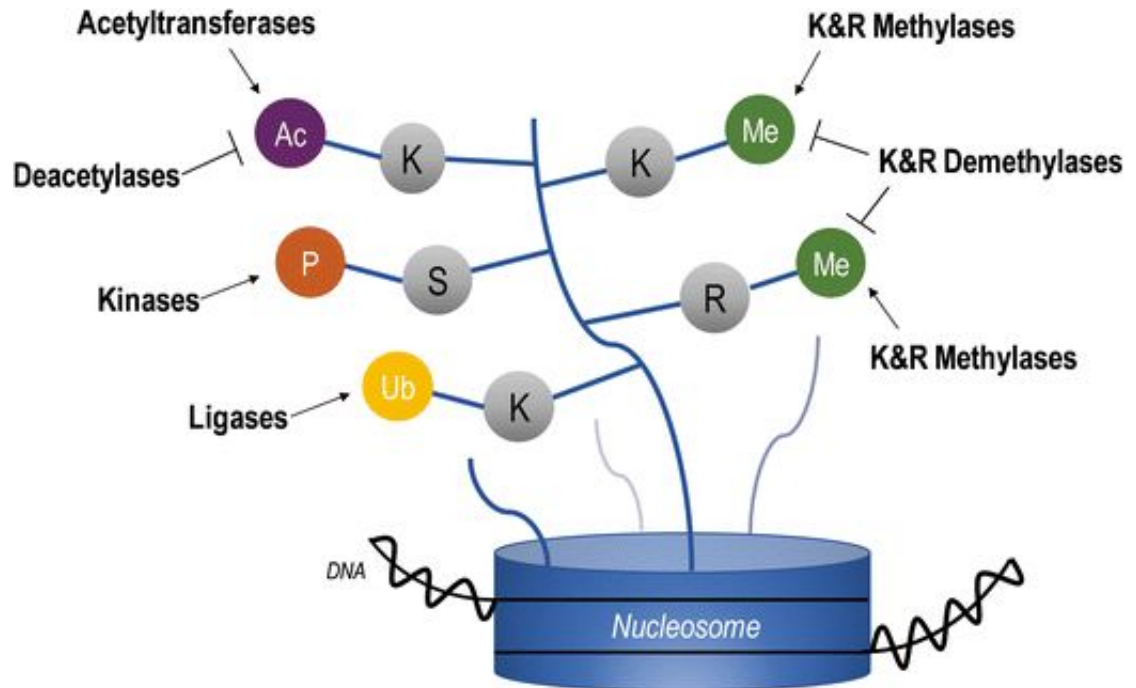


FIGURE 1.14. Epigenetic regulators. DNA is wrapped around the nucleosome that comprises two copies of each histone H2A, H2B, H3, and H4. Ac acetylation, P phosphorylation, Ub ubiquitination, Me methylation, K lysine, S serine, R arginine (Roussel and Stripay, 2018).

Genes that encode H3K4 methyltransferases *KMT2D* (also known as *MLL2* or *MLL4*) and *KMT2C* (also known as *MLL3*) are mutated in 8% and 3% of medulloblastoma subgroup-independently. Both, *KMT2D/C* complexes are required in order to maintain level of H3K4me1 which marks enhancers in cell-type-specific manner. These complexes may activate gene expression by changing enhancers activity during development and cancer (Rao and Dou, 2015; Sze and Shilatifard, 2016).

Alterations in histone demethylases consistent with aberrant histone methylation at H3K27 and H3K4 are common in Group3 and Group4 medulloblastoma (Roussel and Stripay, 2018).

Transcriptional effectors of lysine demethylase (KDM) regulation, including genes involved in cell cycle control and differentiation. Various mutations were identified across six KDM family members. Inactivating mutations in *KDM6A/UTX* is of a particular interest as it is the most common recurrent mutation in Group 4 medulloblastoma and (Robinson et al., 2012; Northcott et al., 2009) and generally mutations and homozygous deletions of this gene occur in 4% of medulloblastoma.

Gene *KDM6A/UTX* is located on the X-chromosome and Group 4 medulloblastoma patients are predominantly males and most of female patients have one of the X

chromosome lost (Dubuc et al., 2013). *KDM6A/UTX* activates genes by demethylating repressive H3K27 mark and interacts closely with and is considered a subunit of KMT2D/C complexes (Cho et al., 2007).

Together, *KMT2D/C* and *KDM6A/UTX* may be responsible for activation of genes by methylating H3K4 and H3K27me₃, and therefore, their mutations are mutually exclusive in medulloblastoma.

In addition to mutations of *KDM6A/UTX* in Group 4 medulloblastoma, EZH2, EED and SUZ12 subunits of H3K27 methyltransferase polycomb repressor complex PRC2, are highly expressed in medulloblastoma, often due to amplification of the genes encoding these subunits (Bautista et al., 2017; Dubuc et al., 2013).

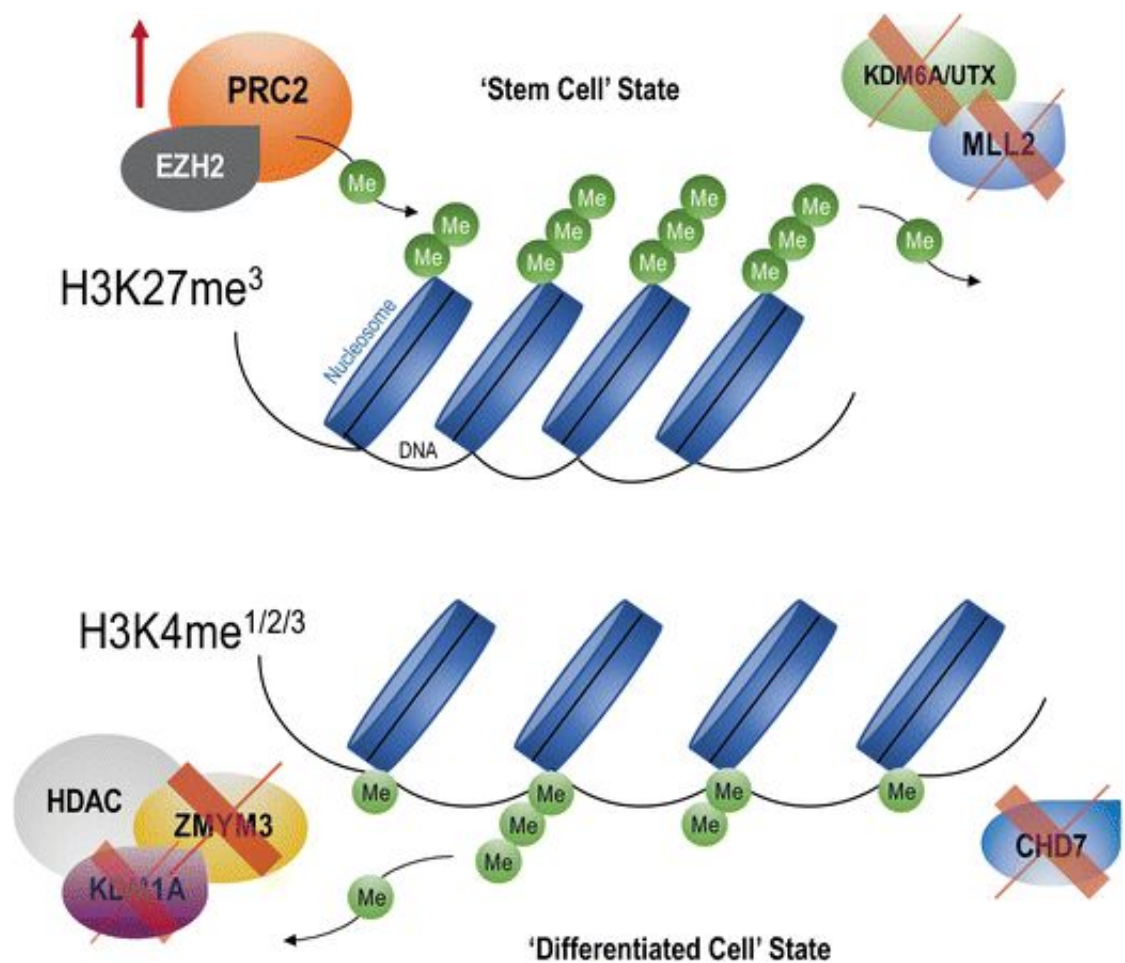


FIGURE 1.15. Schematic representation of bivalent chromatin domains, H3K4 and H3K27. Medulloblastoma retains a progenitor-like epigenetic profile by altering the balance between H3K27 and H3K4 methylation states. Aberrant “writing” or “erasing” of methyl groups of H3K27me₃ by EZH2 upregulation or *KDM6A/UTX* mutation, respectively, and inactivating mutations in *MLL2/KMT2D*, *CHD7*, and *ZMYM3* that disrupt H3K4me₃ transcription, maintains the stem cell state (Roussel and Stripay, 2018).

EZH2, catalytic partner of PRC2, is responsible for transferring methyl groups onto lysine 27 of histone 3 (H3K27me₃) and contributes to chromatin compaction and

transcriptional repression (Cao et al., 2002). EZH2 is critical for normal development as its deletion is lethal during embryonic stage (O’Carroll et al., 2001) and is over-expressed in many cancers, and in most cases plays a role of an oncogene or a tumour suppressor (Bachmann et al., 2006; Vo et al., 2017).

Mutations in *KDM6A/UTX* leading to a complete protein loss are mutually exclusive with EZH2 overexpression. This suggests a critical role of H3K27me3 in medulloblastoma tumorigenesis and potential clinical vulnerability (Dubuc et al., 2013).

1.3.11.3. *Chromatin remodelling in medulloblastoma.* ATP-dependent chromatin remodelling complexes represent another epigenetic mechanism for tumorigenesis in medulloblastoma, which regulate chromatin structure using energy from ATP hydrolysis. The prototypical complex, SWI/SNF family of proteins, also known as BAF in humans, contains 10-12 associated subunits and can function as transcription activator and repressor and plays important role during development (Hargreaves and Crabtree, 2011; Wu, 2012). Recurrent mutations in BAF family identified in medulloblastoma include *SMARCA4* and are typically restricted to WNT and Group 3 tumours. In addition to chromatin remodelling function, BAF complexes interact with other chromatin regulators, e.g. under specific conditions they functionally antagonise polycomb repressive complexes PRC1 and PRC2 (Kadoch et al., 2017; Stanton et al., 2017).

In addition to BAF complexes, mutations in another chromatin remodeler *CHD7* were identified in Group 3 and Group 4 medulloblastoma (Robinson et al., 2012) and inactivating mutations in a histone binding protein *ZMYM3*, that contributes to regulation of gene transcription at the H3K4me3 mark are found to be exclusive to Group 4 medulloblastomas (Leung et al., 2004). The latter often occur with sub-median EZH2 expression and *KDM6A/UTX* mutations, which indicates the importance of the role of H3K27/H3K4 changes in medulloblastoma (Robinson et al., 2017).

Medulloblastoma has been reported in people with CHARGE syndrome, a condition that affects the nervous system and which is caused by germline mutations of the *CDH7* gene (Martin, 2015; Gajjar et al., 2014). Mutations in BAF subunits cause Coffin-Siris syndrome, whilst mutations in *MLL2* and *UTX* cause Kabuki syndrome (Santen et al., 2012; Tsurusaki et al., 2012; Miyake et al., 2013). It is unclear if patients with these syndromes are predisposed to medulloblastoma (Yi and Wu, 2018).

Large-scale chromatin remodellers, e.g. NuRD complexes, contribute to enhancer dynamics as co-repressors or co-activators. Changes in enhancer landscape may contribute to aberrant activation of oncogenes and inactivation of tumour-suppressor genes. A recent study suggests that PRDM6, a putative target of enhancer hijacking and presumed histone methyl transferase, may be a driver of Group 4 medulloblastoma (Northcott et al., 2017).

1.3.11.4. *MicroRNAs and Long Non-Coding RNAs in medulloblastoma.* MiRNAs are generated from long transcripts initially processed in the nucleus as pre-miRNAs, which are then transported to cytoplasm where they are cleaved into mature miRNAs. Then miRNAs are incorporated into the RNA-induced silencing complex (RISC) in order to bind to their specific mRNAs (MacFarlane and R. Murphy, 2010).

The first miRNA reported as aberrantly expressed in medulloblastoma was *miR-124*. Its suppressed expression leads to overexpression of *CDK6* which is thought to be linked to poor prognosis in medulloblastoma (Pierson et al., 2008).

Another miRNA, *miR-17 92*, was found overexpressed in medulloblastoma. This miRNA is a target of MYCN, which is direct target of SHH signaling and is often overexpressed by amplification in SHH medulloblastomas (Northcott et al., 2009).

Suppression of *miR-17 92* was shown to inhibit SHH medulloblastoma development whereas enforced expression accelerates tumour formation, which is consistent with oncogenic activity (Murphy et al., 2013).

Recently, a LncRNA, Linc-NeD125, was identified as over-expressed in Group 4 medulloblastoma. This LncRNA isolates the RISC complex, containing three miRNAs: miR-19a, miR-19b and miR-106a, which are encoded by the *miR-17 92* cluster. This leads to de-repression of its targets: *CDK6*, *MYCN*, *KDM6A/UTX* and *SNCAIP*, major drivers of Group 4 medulloblastoma (Laneve et al., 2017). A better understanding of functions of miRNAs and LncRNAs in medulloblastoma may be useful for identifying novel pharmacological targets (Roussel and Stripay, 2018).

1.3.12. Treatment of medulloblastoma. Surgery is the primary treatment modality for medulloblastoma, as a tool for alleviation of the intra-cranial pressure, diagnosis and as a risk-stratification factor. Surgical resection is usually followed by radiotherapy where appropriate (e.g. in patients over 3 years of age), and chemotherapy. Radiation

avoiding strategies are often attempted in children under 3 years of age as infants patients are associated with particular risk of significant neurocognitive effects associated with radiation therapy (Von Bueren et al., 2011; Ashley et al., 2012).

Currently, the recommendation is to perform the maximal safe resection possible, prior to starting adjuvant treatments, such as chemotherapy and irradiation. This strategy may change in the near future, with improved understanding of the specific subgroups biology and with the advance in new treatments available.

In cases when gross total resection is possible, patients usually have longer relapse-free intervals than patients who have residual tumour at the end of surgery (Martin et al., 2014; Adamski et al., 2014). However, recent studies suggested that benefits of gross total resection vs subtotal resection are dependent on the subgroup, showing that no significant survival benefit existed for greater extent of resection for patients with WNT, SHH, or group 3 tumours. For patients with group 4 tumours, gross total resection conferred a benefit to progression-free survival compared with sub-total resection especially for those with metastatic disease, however no noticeable difference to overall survival (Thompson et al., 2016).

The choice and intensity of adjuvant therapy depends on the presence of risk factors at diagnosis such as age, metastatic stage, residual disease and histopathology. Adjuvant therapy of medulloblastoma typically includes craniospinal radiotherapy (CSI) and chemotherapy. Average-risk patients receive CSI 23.4 Gy in 30 fractions, followed by tumour bed boost to 54-56 Gy over 6 weeks, with or without vincristine, which is a mitotic inhibitor. Then children over 3 years old with non-metastatic disease receive 8 cycles of therapy that includes vincristine, cisplatin (a DNA cross-linker) and two alkylating agents - lomustine and cyclophosphamide for approximately a year. For high-risk disease, CSI is given at higher dose of 36-39.6 Gy in 30 fractions followed by a posterior fossa boost 54-56 Gy over 6 weeks and subsequent chemotherapy. Sometimes, stem cell transplants are used prior to the therapy (Packer and Vezina, 2008; Packer et al., 2012). Adult patients are treated in the similar way, however post-adjuvant therapy does not seem to improve survival and agents like vincristine and cisplatin cause significant toxicity (Sengupta et al., 2017).

1.3.13. Treatment *sequelae* . The improved survival observed for medulloblastoma patients over decades has come at the cost of considerable difficulties later in life,

known as late effects, as well as immediate side effects, that are a direct consequence of the treatment. The most significant consequences described below.

Endocrine system: Several endocrinological *sequelae* have been described in medulloblastoma survivors and are mainly, but not exclusively, consequences of direct radiation damage to the pituitary gland and hypothalamus (Fossati et al., 2009). Hypothyroidism has been reported in a percentage of patients varying from 8% to 60% (Chin et al., 1997; Paulino, 2002; Heikens et al., 1998). Adrenocorticotrophic hormone (ACTH) deficiency is rare but life threatening, it has been reported in 24% of pediatric medulloblastoma (Rose et al., 2005). Gonadotropin secretion abnormalities and early puberty have been observed in up to 50% medulloblastoma survivors (Ogilvy-Stuart et al., 1994; Heikens et al., 1998). Growth hormone (GH) deficiency is observed in 40–80% of the cases and is the most common endocrine abnormality in these patients (Adan et al., 2001).

Neurocognitive development: Neuropsychological toxicity has of course a strong impact on the whole life of these patients, resulting in a reduced probability of completing education, finding a job and marrying (Kiltie et al., 1997). These patients have deficits in IQ, memory, attention, language, mathematic ability and sleep-wake rhythm; they also have a lower likelihood of being able to drive a car (Yock et al., 2016).

Growth and bone development: Adult medulloblastoma survivors often suffer from reduced height. This phenomenon can be attributed mainly to the following three factors: GH deficiency, early puberty and impaired spinal growth due to direct RT toxicity on the growing vertebral bodies. Osteopenia (reduced bone mineral density) is common in medulloblastoma survivors. Quality of life, pain and physical activity are worse in osteopenic patients (Odame et al., 2006).

Ototoxicity: Sensorineural hearing loss (SNHL) is a well-documented side effect of cisplatin, a key drug in medulloblastoma chemotherapy, and full dose RT to the cochleae. It develops 6-12 months after RT and is irreversible (Miettinen et al., 1997).

Medulloblastoma survivors are at increased risk of developing new malignancies. Many of these second cancers arise in the previously irradiated area and may be radiation induced (Fossati et al., 2009).

Approximately 25% of patients undergoing resection of medulloblastoma experience cerebellar mutism, also known as posterior fossa syndrome. This is an immediate

complication of the surgery and usually manifests within 24 hours after resection of a posterior fossa tumour. Patients are unable to produce words, are extremely irritable, and develop hypotonia and ataxia. The duration of the syndrome varies from weeks to months, however the language difficulties may be lifelong (Robertson et al., 2006; Martin et al., 2014).

1.3.14. Survival and quality of life dilemma: potential of improving patients outcome. The advances in treatment and care over the last 50 years yielded improvement in 5 year survival rates of medulloblastoma patients from 3% to 70% (Newton, 2001; Crawford et al., 2007). This is very significant improvement, however there are still two major areas that need improvement:

- (1) improve effectiveness of treatment for those remaining 30% who still experience relapse;
- (2) improve quality of life of those who survive, who are impacted for the rest of their lives by debilitating late effects described in 1.3.13

The following approaches are currently being investigated:

Reducing intensity of the treatment low-risk patients. That is, reducing treatment for those patients that are expected to do well in order to minimise the treatment sequelae. These include clinical trials that focus on optimising combinations of chemotherapy and radiotherapy, and constant refinement of surgical techniques in order to minimise the damage to healthy brain tissue (Crawford et al., 2007; Pizer and Clifford, 2009).

Several prospective and retrospective studies, based on data from clinical trials such as SIOP-PNET3, PNET4 and SJMB96 have shown that non-metastatic patients with WNT-activated medulloblastoma who are under the age of 16 have an excellent survival independent of the treatment protocol used and are considered low-risk (Ramaswamy et al., 2016b; Cho et al., 2011; Clifford et al., 2015; Schwalbe et al., 2013; Remke and Hielscher, 2011b). Ongoing clinical trials (ClinicalTrials.gov identifiers NCT01878617 (SJMB12), NCT02066220 (SIOP-PNET5), NCT02212574 and NCT02724579) are currently evaluating de-escalation of therapy e.g. less chemotherapy and/or less radiation for this group of patients. There is evidence that patients treated with less radiation preserve more of their intellectual capabilities (Moxon-Emre et al., 2016).

Recently proposed stratification schemes outlined in sections 1.3.10 and 1.3.7 defined a subset of low-risk patients from non-WNT subgroups, who are potentially eligible for treatment de-escalation, such as Group4 patients with whole chromosome 11 loss, chromosome 17 gain or Group3/Group4 with chromosome 13 loss and non MYC-amplified as well as a subset of SHH-activated group (Goschzik et al., 2018; Cavalli et al., 2017; Shih et al., 2014; Ramaswamy et al., 2016b).

Targeting specific signalling pathways, such as WNT and SHH. Pharmacological treatments which specifically target these pathways may provide control of WNT/SHH activated medulloblastoma with significantly reduced side effects associated with current treatments. There is a lack of specific therapies targeting the WNT pathway in medulloblastoma, however inhibition of tankyrase (Huang et al., 2009), cyclo-oxygenase (Castellone et al., 2005), and Porcupine (Liu et al., 2013) to decrease beta-catenin signaling are under investigation in other tumour models. As WNT-activated medulloblastoma is almost universally driven by stabilizing mutations in *CTNNB1*, a rational approach would be targeting its interaction with another transcription factor, CREB binding protein (CREBBP), to inhibit transcription of their target genes. CREBBP:CTNNB1 interaction antagonist PRI-724 is being investigated in phase I clinical trials in pancreatic and liver cancers (ClinicalTrials.gov identifier NCT02413853) (Archer et al., 2017). However, inhibiting the WNT pathway in order to treat medulloblastoma is potentially problematic. Recent work has suggested that patients with WNT-activated tumours exhibit a disrupted blood-brain barrier which allows improved delivery of chemotherapy agents directly to the tumour cells (Phoenix et al., 2016). Inhibiting the WNT signalling pathway could improve the integrity of blood-brain barrier and increase resistance of the tumours to the chemotherapy (Archer et al., 2017).

In the SHH pathway, the patched protein (PTCH1) normally inhibits smoothed protein (SMO) and all the others elements of the SHH pathway are located downstream of SMO. When SMO is over-expressed, it targets downstream pro-proliferative transcription factors GLI 1-3 (Lin and Matsui, 2012). Small molecule inhibitors such as cyclopamine derivatives may be used to control SMO. Several of these molecules, such as Vismodegib, have shown activity against MB in a variety of clinical trials (Archer et al., 2017). However, mutations downstream of SMO are not affected by SMO inhibitors. Vismodegib is a synthetic SMO inhibitor based on cyclopamine. A recent

phase II clinical trial reported that patients who responded to the treatment with vismodegib had tumours with *PTCH1* mutations (Ransohoff et al., 2015; Robinson et al., 2015). However, patients with *SUFU* and *GLI1* mutations were not affected by the drug because SMO inhibitors is effective only on tumours with genetic mutations upstream of *SMO* in the pathway (Archer et.al., 2017). Also, vismodegib was found to cause premature irreversible growth plate fusion in children, therefore it is only suitable for use in adults (Robinson et al., 2017). SHH tumours are frequently prone to specific mutations that prevent binding to SMO or which amplify downstream effectors such as *GLI2* (Dijkgraaf et al., 2011). Other agents, that target downstream effectors (Kim et al., 2010; Beauchamp et al., 2011), such as arsenic trioxide, have been identified but have yet to be tested in clinical trials.

Improving current risk stratification scheme. Approximately 70% of newly diagnosed medulloblastomas will be stratified as "standard-risk". Under current treatment regimens about 80% will survive. However, the rest will relapse, with the vast majority of patients who relapse subsequently dying of their disease (Martin et al., 2014). Therefore, these patients will benefit from a more refined risk stratification. The novel risk stratification schemes discussed in 1.3.10 and 1.3.7 could be used to identify subsets within current standard risk, of low-risk patients eligible to treatment de-escalation or high-risk patients that could be subjected to novel treatment regimens in order to improve their outcome and quality of life balance; however, these schemes require validation in clinically-controlled trials cohorts.

1.3.15. Hypothesis and aims. This project investigates feasibility of integrated genetic and epigenetic prognostication using only methylation arrays in application to historic FFPE archived cohorts.

Aims

- Confirm usability of methylation arrays and develop methods to detect genomic alterations (aneuploidy and focal oncogene amplifications)
- Showcase application of methylation arrays as a cost-effective single-platform, integrated approach for improved prognostication within medulloblastoma patients

CHAPTER 2

Materials and methods

2.1. Cohorts

The cohorts of primary medulloblastoma samples used for research described in chapters 3, 4 and 5 are part of Newcastle Medulloblastoma Archive (NMB) and belong to the Paediatric Brain Tumour Research Group, Newcastle University. Tumour investigations were done with approval from Newcastle North Tyneside Research Ethics Committee (study reference 07/Q0905/71); all tumour material was collected in accordance with this approval.

Two representative cohorts from Newcastle Medulloblastoma Archive run on Illumina HumanMethylation450 arrays were used as the test cohort in chapter 3. A cohort of 135 medulloblastoma samples, for which Affymetrix SNP6.0 genotyping array calls were available, was a mixture of fresh-frozen (FF, n=133) and formalin-fixed paraffin-embedded (FFPE, n=2) samples of all molecular subgroups as detailed in Table A.1. Matching SNP6.0 array profiles were provided by Dr S. Nakjang of Newcastle University Bioinformatics Support Unit.

The test cohort of 203 mixed FF (n=188) and FFPE (n=15) samples representative of all molecular subgroups as outlined in Table A.2 was used for the evaluation of the pipeline-specific suitability and performance in chapter 3.

A non-WNT medulloblastoma cohort of 338 patients from Newcastle Medulloblastoma Archive (96 (28.4%) SHH, 99 (29.3%) Group 3 and 143 (42.3%) Group 4, as detailed in A.3) was used as a validation cohort in chapter 4. The cohort comprised only of infant and children cases of median age 2.83 years in SHH subgroup, 4.32 in Group 3 and 7.95 years in Group 4. Male to female ratio was 1.4:1 in SHH, 2.8:1 in Group 3 and 2.1:1 in Group 4. Median follow-up was 6.3, 5.8 and 4.6 years for SHH, Group 3 and Group 4 molecular subgroups respectively, calculated using the "reverse" Kaplan-Meier method as outlined in section 2.9.3 to match the original study by Shih et al. (2014) that was validated using this cohort in chapter 4.

The following two cohorts were used in both chapter 4 and 5.

The Newcastle Medulloblastoma Standard Risk non-WNT/non-SHH cohort (SR NMB) consisted of 75 samples which were negative for known high-risk features (see below) and were chosen according to the following criteria:

- age group 3-16 years old (children)
- Metastatic stage M0 (no metastasis), and M0/M1 (patients with metastatic stage estimated to be either M0 or M1, since imaging, which can distinguish between M0 and M1 disease, was unavailable at the time)
- Non LCA pathology
- Non *MYC(N)* amplified
- Gross total resection

This cohort comprised of 16 (21%) Group 3 and 59 (79%) samples as presented in Table A.4, 55 (73%) were male and 20 (27%) female samples with male to female ratio 2.75:1 with median age at diagnosis 7.46 years. Follow up time was calculated by simply finding a median of all the survival time to match the original study by Goschzik et al. (2018) that was validated using this cohort in chapter 4. Histological variants were mostly classic MB (n=68, 91%) with few desmoplastic/nodular samples (n=7, 9%) and all tumours gross-totally resected.

The Newcastle Medulloblastoma High Risk non-WNT/non-SHH cohort (HR NMB) consisted of 100 samples which were positive for at least one known high-risk feature, such as metastatic disease, large cell/anaplasia histopathology, sub-total resection or *MYC(N)* amplification. Subgroup composition was 33 (33%) Group 3 and 67 (67%) Group 4 samples as outlined in Table A.5, and 71 (71%) were male and 29 (29%) were female samples with male to female ratio 2.44:1. Histopathological composition was 24 (24%) LCA in addition to 75 (75%) samples with classic and one with DN pathology.

An external, independent mixed-risk cohort (n=244) from the study by Cavalli et al. (2017) was used as validation cohort in chapter 5. Subgroup composition was 64 (26%) Group 3 and 180 (74%) Group 4 samples as outlined in Table A.6. Male to female ratio was 2:1 with 165 (68%) males and 79 (32%) females, median age at diagnosis 8 years old, 94 (39%) samples were from patients with metastatic disease. More detailed outline and comparison with SR NMB and HR NMB cohorts is presented in the Table 5.1.

2.2. Microarrays

A DNA microarray (also known as DNA chip or biochip) is a collection of microscopic DNA spots attached to a solid surface that are used to measure the genome-wide transcriptome, methylome or for genome-wide assessment of genotypes. Two microarrays platforms were employed during this study: Affymetrix SNP 6.0 genotyping array (McCarroll et al., 2008) and Illumina Infinum HumanMethylation450k BeadChip array (Bibikova et al., 2011). Both platforms were used for development of array-based methods for detecting Copy-Number Variations in medulloblastoma (see chapter 3) and HumanMethylation450k arrays were used for validation of previously published cytogenetic prognostication in chapter 4 and for the identification of methylomic biomarkers in standard-risk disease in chapter 5. Affymetrix SNP 6.0 arrays are considered the gold standard for copy number variation interrogations due to a large number of probes (over 945,000 for CNA and around 1.8 million in total) that are relatively evenly distributed across genome in comparison with methylation arrays (see Table 2.1), which have their probes concentrated around gene promoters (Haraksingh 2011) :

Platform	Affymetrix SNP6.0	Illumina 450k Methylation array
Number of probes	945,806	485,577
Median intermarker distance (kb)	2.3	0.35
Mean intermarker distance (kb)	3.0	5.8

TABLE 2.1. Genomic probe distribution. Number of probes, mean and median intermarker distance interrogating copy number alterations from Affymetrix SNP 6.0 and Infinum HumanMethylation450 BeadChip. Table adapted from Feber et al. (2014);

Overview of SNP array technology is shown on the Figure 2.1

2.2.1. Affymetrix SNP6.0 arrays. The Affymetrix Genome-Wide Human SNP Array 6.0 features 1.8 million genetic markers, more than 906,600 of which are for the assessment of single nucleotide polymorphisms (SNPs) and more than 946,000 probes for the detection of copy number variation, 202,000 of which located in known CNV regions and 744,000 which are spread evenly across the entire genome.

Each SNP is interrogated by six or eight perfect match probes - three or four replicates of the same probe for each of the two alleles (McCarroll et al., 2008). Therefore, intensity data for each SNP consists of two sets of repeated measurements. The copy number probes are designed to interrogate regions of the genome that do not harbour

SNPs, but which may be aberrant with regard to copy number. Each such copy number site is interrogated by only one probe.

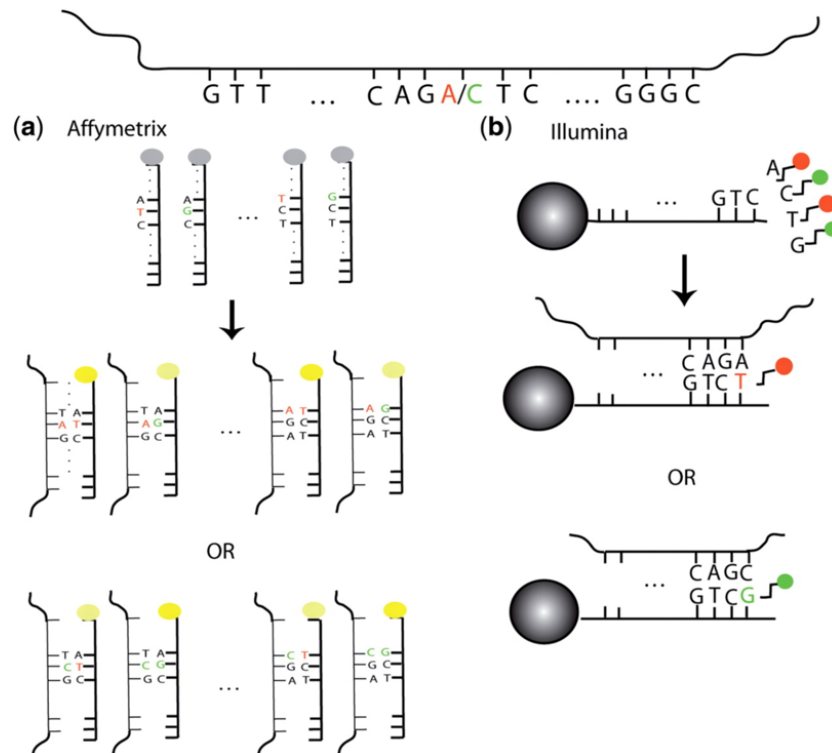


FIGURE 2.1. Overview of SNP array technology. At the top is the fragment of DNA harbouring an A/C SNP to be interrogated by the probes shown. (a) In the Affymetrix assay, there are 25-mer probes for both alleles, and the location of the SNP locus varies from probe to probe. The DNA binds to both probes regardless of the allele it carries, but it does so more efficiently when it is complementary to all 25 bases (bright yellow) rather than mismatching the SNP site (dimmer yellow). This impeded binding manifests itself in a dimmer signal. (b) Attached to each Illumina bead is a 50-mer sequence complementary to the sequence adjacent to the SNP site. The single-base extension (T or G) that is complementary to the allele carried by the DNA (A or C, respectively) then binds and results in the appropriately-coloured signal (red or green, respectively). For both platforms, the computational algorithms convert the raw signals into inferences regarding the presence or absence of each of the two alleles (LaFramboise 2009).

2.2.2. Methylation arrays. The Illumina 450k array is a high density BeadChip, that can assay over 450,000 CpG sites and analyse twelve samples in parallel. It covers 99% of RefSeq genes with multiple probes per gene, 96% of CpG islands from the UCSC database and CpG island shores (Bibikova et al., 2011).

The array technology uses beads with long, target-specific probes designed to target individual CpG sites, assembled on a slide. Each array contains two types of methylation assay design: The Infinium I method previously used in HumanMethylation27 arrays, and the newly developed Infinium II; DNA methylation is measured by quantitative “genotyping” of bisulfite converted genomic DNA.

Infinium I design consists of two query (Figure 2.2A) probes per CpG locus: “methylated” and “unmethylated”. The 3’ terminus of a probe is designed to match either protected cytosine or thymine base that results after bisulfite conversion (methylated and unmethylated design respectively). In this assay, each probe has a span of 50 bases and it was assumed that the methylation for adjacent CpG sites would be highly correlated and would change in the same way as the CpG being assessed. As there are separate probes for methylated and unmethylated signals, both signals are measured with the same colour.

The Infinium II assay requires one probe per locus for CpG sites, located in regions of low CpG density (Figure 2.2B). The 3’ terminus of the probe complements the base directly upstream of the query site and single base extension adds a G or A base complementary to either “methylated” C or “unmethylated” T. Therefore, the methylation status of the query site is assessed independently of the methylation status of adjacent CpG sites. As only one probe is used for both methylated and unmethylated signal, methylated signal intensity is measured in the green colour and unmethylated in the red colour channels.

The methylation level is measured using two methods: β -value method and M-value method. The Beta-value is the ratio of the methylated probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities): $\beta = \text{intensity of the Methylated allele (M)} / (\text{intensity of the Unmethylated allele (U)} + \text{intensity of the Methylated allele (M)})$, which represents percentage provides a continuous measure of levels of DNA methylation in samples, ranging from 0 (or 0%) indicating that all copies of the CpG site in the sample were completely unmethylated (no methylated molecules were measured) to 1 (or 100%) in completely methylated sites (Du et al., 2010) (Du et al., 2010). The second method is the log₂ ratio of the intensities of methylated probe versus unmethylated probe; using this approach, parametric statistical methods can be applied to methylation array data (Irizarry et al., 2008):

$$M = \log_2 \frac{\beta}{1 - \beta}$$

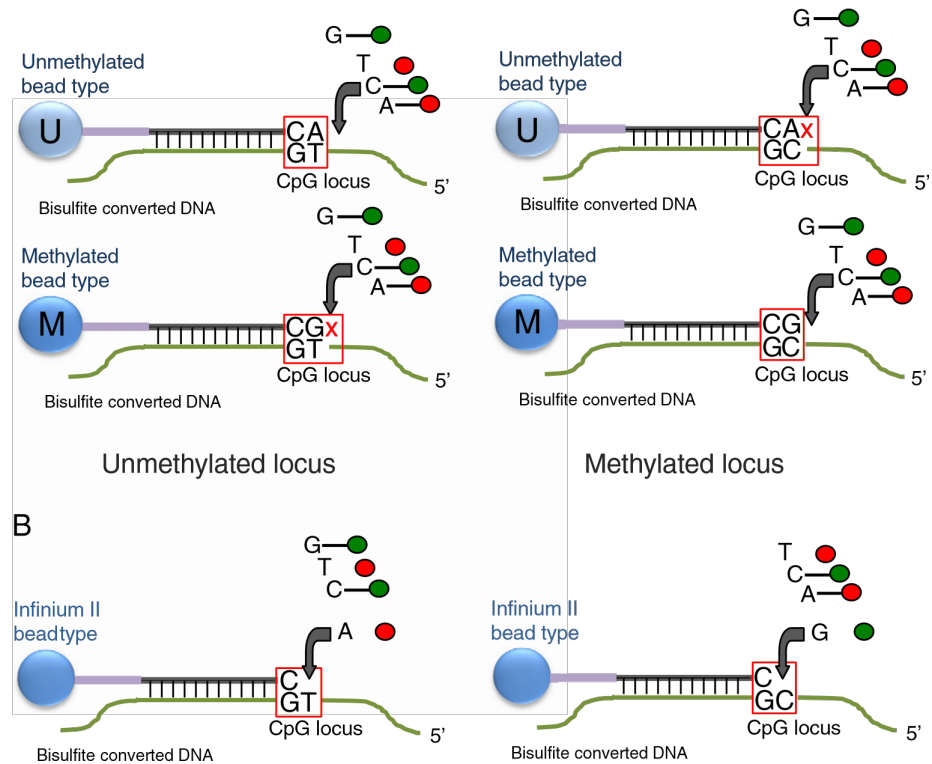


FIGURE 2.2. Infinium Methylation Assay scheme.

A. Infinium I assay. Two bead types correspond to each CpG locus: one bead type — to methylated (C), another bead type — to unmethylated (T) state of the CpG site. Probe design assumes same methylation status for adjacent CpG sites. Both bead types for the same CpG locus will incorporate the same type of labelled nucleotide, determined by the base preceding the interrogated “C” in the CpG locus, and therefore will be detected in the same colour channel.

B. Infinium II assay. One bead type corresponds to each CpG locus. Probe can contain up to 3 underlying CpG sites, with degenerate R base corresponding to C in the CpG position. Methylation state is detected by single-base extension. Each locus will be detected in two colours. In the current version of the Infinium II methylation assay design, labelled “A” is always incorporated at unmethylated query site (“T”), and “G” is incorporated at methylated query site (“C”). Figure adapted from Bibikova et al. (2011)

2.3. Methylation microarrays data pre-processing and quality control

The standard file format generated during Illumina BeadChip processing is IDAT (stands for "Intensity DATA file"). IDAT files generated during the scanning of methylation or genotyping arrays are binary files (one for each the green and red channels). Each file contains four fields: the bead-type ID on the array, the mean and standard deviation of intensities of each bead-type and the number of beads of each type. Additionally the files also contain metadata information such as the type of BeadChip, specific software version used and the date the array was scanned.

2.3.1. Initial quality control: identification of unusual samples and technical artefacts. The 450K arrays contain several internal control probes that are used to assess quality of different sample preparation steps during initial quality assessment of the arrays QC reports preparation. Types of these probes and their purpose are summarised in Table 2.2

Control probe	Purpose
Bisulfite conversion	Test efficiency of bisulfite conversion by query of C/T polymorphism
Hybridisation	Test the overall performance of Infinium assay using synthetic targets (not DNA) at three concentrations
Specificity controls	Test for non-specific detection of methylation signal over unmethylated background. This controls are designed against non-polymorphic T sites (G/T mismatch)
Extension	Test efficiency of extension of A, T, C and G nucleotides from a hairpin probe (sample independent). The perfect match hairpin controls should result in high-signal and mismatch in low signal.
Staining	Test efficiency and sensitivity of staining step (independent of hybridisation/extension steps)
Target removal	Test efficiency of stripping step after extension
Non-polymorphic	Query a non polymorphic base A,T, C and G to test overall performance of the assay from amplification to detection

Control probe	Purpose
Negative	Randomly permuted bisulfite-converted sequences containing no CpGs. They should not hybridise to DNA. The mean intensity of these probes determines the system background

TABLE 2.2. Illumina HumanMethylation450k control probes

Illumina HumanMethylation450k arrays also contain 65 control probes, that assay highly-polymorphic single nucleotide polymorphisms (SNPs) rather than DNA methylation and are included on the array to allow sample quality control to check for relatedness between individuals and enable the detection of potential sample mislabelling or unintended duplicates. The signal from these probes is expected to cluster into three distinct groups and could be used to provide an indication of the degree of technical variance between samples, as unintentional duplicate samples from the same individual would cluster together.

In order to identify any technical artefacts on the arrays and unusual samples, initial quality control procedure involves plotting distributions of the samples e.g. β -value densities, box-plots or density bean plots. Also it is useful to plot distribution of Type I and Type II probes separately to observe differences in chemistry and enrichment for different element.

Next steps after initial quality control in pre-processing raw microarray data for the analysis are filtering and normalisation, which are described below.

2.3.1.1. *Filtering.* Each assessed locus has an associated detection p-value: the probability that the target signal was distinguishable from background noise. One approach to overcome this issue is drop individual beta values if their detection p-value exceeds a threshold, typically 0.01 or 0.05.

The filtering steps related to technical issues are: dropping probes with known SNPs residing in their probe sequence, and also dropping probes that have the potential to anneal to multiple genomic locations.

Filtering steps related to analytical choices include: dropping probes on X and Y chromosomes, those with lowest variation, and with extreme methylation levels (e.g. median = 0% or 100%).

2.3.1.2. *Normalisation and batch correction.* Normalisation and batch correction techniques aim to identify and remove any remaining non-biological variation by modifying and standardising the measurements.

Causes of technical variation and batch effect (systematic differences across groups of samples, such as samples processed on different days) include differences in sample handling and preparation (e.g. extraction methods), DNA processing (e.g. amplification, labelling, hybridisation), scanning of arrays/chips (e.g. background noise), location of sample on chips and technical biases, such as dye and probe bias.

Details of pre-processing, filtering and normalisation process used in this work are given in the section 2.3.2.

In addition to initial standard data pre-processing and normalisation, normalisation of cancer samples against a set of normal controls was performed during copy number analysis (see sections 2.6 and 2.7 for details), in order to neutralise the effects of potential residual artefacts as described in (Sturm et al., 2012) and (Hovestadt and Zapatka, 2015). This step is required for correcting for probe and sample bias e.g. caused by GC-content, type I/II probe differences or technical variability (Hovestadt and Zapatka, 2015). Therefore, no specific correction for "genomic wave" artefact (a probe effect that correlates with surrounding genomic GC content), e.g. loess correction, which estimates and removes the wave effects (Feber et al., 2014), was performed prior to CNA analysis.

2.3.2. Minfi . Minfi (Aryee et al., 2014) is a suite of computational tools that incorporate statistical techniques for analysis of DNA methylation, which includes methods for preprocessing, quality assesment and detection of differentially methylated regions.

Minfi analysis starts from reading .idat files that contain green and red channel intensities. This data is organised into an object of RGChannelSet class. This initial object contains the raw green and red intensities as well as the intensities of the internal control probes, phenotype data and a manifest object that contains probe design information of the array. Once these data are preprocessed, it can be stored in one of the four additional classes: MethylSet, GenomicMethylSet, RatioSet, GenomicRatioSet. The prefix Genomic in the class name indicates association with genomic location, that is, the class includes genomic annotation.

Objects of class `MethylSet` contain the methylated and unmethylated signal. This class can be constructed using raw preprocessing function `preprocessRaw`, which uses the array design to combine different probes and colour channels to construct the methylated and unmethylated signal and does not perform any normalisation.

Objects of class `RatioSet` store the data as methylation ratios (beta values) or log ratios of beta values (M-values), as well as an optional copy number matrix $CN = \log_2(\text{Methylatedsignal} + \text{Unmethylatedsignal})$.

A `ratioSet` can be created from a `MethylSet` with function `ratioConvert`, which converts methylated and unmethylated signals to beta-values or M-values. `GenomicMethylSets` and `GenomicRatioSets` can be created from by mapping them to the genome with the function `MapToGenome`. A `MethylSet` can be either converted to `RatioSet` first and then mapped to genome or mapped first and then converted to `RatioSet` (see Figure 2.3). These conversions are irreversible.

Objects of class `MethylSet` can also be constructed and normalised at the same time by using the `preprocessIllumina` function. This method implements the functionality of the preprocessing tool available in `GenomeStudio` - Illumina's standard preprocessing software. It performs background subtraction and control normalisation. These functionalities are optional and turning off these options is equivalent to raw preprocessing discussed above.

Other normalisation methods implemented in `minfi` are: Subset-Quantile Within Array Normalisation (SWAN), Functional normalisation (FunNorm), Quantile and normal-exponential out-of-band (Noob) normalisation.

The function `preprocessQuantile` performs stratified quantile normalisation preprocessing, where the normalisation procedure is separately applied to methylated and unmethylated intensities. The distribution of both type I and II signals are forced to be the same, first quantile normalising the type II probes across samples and then interpolating a reference distribution to which the type I probes are normalised (Fortin and Hansen, 2014).

The function `preprocessSWAN` performs Subset-Quantile Within Array Normalisation (SWAN) (Maksimovic et al., 2012), a within array normalisation correction for the technical differences between Type I and Type II probe designs. The algorithm applies within-array quantile normalisation separately for different types of probes

divided by CpG content. The input is either a non-normalised `MethySet` or `RGChannelSet`, in which case the function will first call `preprocessRaw` and apply the SWAN normalisation. The output in both cases is a normalised `MethySet`.

Functional normalisation, performed by the function `preprocessFunnorm()`, is a between-array normalisation method for the Illumina Infinium HumanMethylation450 platform. It removes unwanted variation by regressing out variability explained by the control probes present on the array (Fortin and Hansen, 2014).

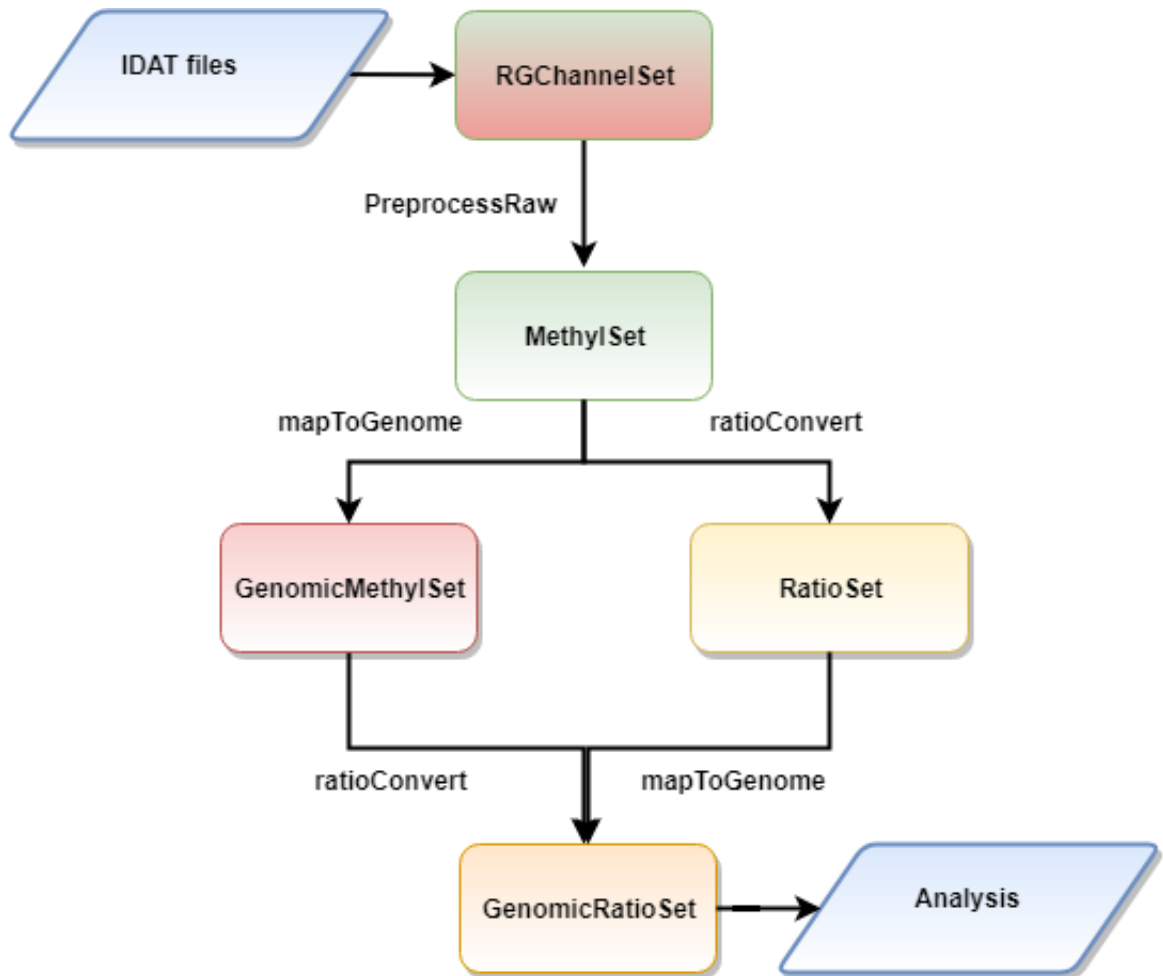


FIGURE 2.3. Minfi preprocessing flow chart

Minfi provides quality control and filtering functionality to perform the corresponding steps described in section 2.3, by providing simple quality control plot functions, a control probes plot, functions to filter probes, including probes with proximal SNPs, sex chromosome probes and cross-reactive probes.

2.4. Sample quality control: Derivative Log Ratio Spread

Sample quality control is absolutely critical during copy number analysis. If not addressed early, poor quality DNA in a sample can lead to poorly defined copy number

segments, which may result in false-positive and non-replicable findings. Therefore, samples which have poor signal-to-noise properties may pose difficulties for accurate CN aberration detection and would need to be excluded from analysis.

One of the procedures to identify poor quality log ratio data (e.g. excessive noise) is the derivative log ratio spread (DLRS). DLRS is a measurement of point-to-point noisiness in log ratio data and is calculated as the absolute value of the \log_2 ratio variance from each probe to the next adjacent probe, averaged over the entire genome.

The distribution of the DLRS can vary dramatically, depending on factors such as DNA source, e.g. tissue type or DNA archiving method (FFPE samples are noisier than fresh frozen, for example) and from one array platform to another, e.g. Illumina HumanMethylation arrays are typically noisier than Affymetrix SNP6.0 arrays. In order to determine which samples to exclude, a standard approach is the calculation of the outlier threshold, where the inter-quartile range (IQR) of the DLRS distribution is calculated and the outlier threshold defined as 1.5 IQRs from the third quartile (GoldenHelix, 2014).

2.5. Circular binary segmentation

Circular binary segmentation (CBS) algorithm is a method that translates noisy intensity measurements into contiguous regions of equal copy number. This method is based on a model that gains and losses of DNA copy number are discrete. DNA copy number aberrations occur in contiguous regions of a chromosome that often makes up an entire chromosome arm or whole chromosome. The approach uses thresholds based on variability of data on the array.

The data from array-based copy number experiments are the intensity measurements of DNA across the genome, called a test sample, and these are often compared against reference sample intensities. It is assumed that reference samples have a normal, diploid genome. Probes with normalised test intensities significantly greater than the reference intensities are indicative of copy number gains in the test sample in the same positions and lower intensities in the test samples are indicative of losses. Therefore, methods for analysing copy numbers are aimed at identifying locations of copy number gains and losses. The data used in the method are log-ratio values ordered by the location of the probe on the chromosome. The locus where a change of copy number

occurs is called a change-point. There may be multiple change-points on the chromosome. The circular binary segmentation algorithm is based on a binary change-point detection method which works in a sequential way: first the check-point is detected from the whole series of data, e.g. across a chromosome, and then the input is split around this change-point and the operation is repeated on the two resulting parts of data. This procedure is based on a test to detect a single change and a potential problem with this method is that it cannot detect a smaller changed segment hidden in a larger segment (Venkatraman and Olshen, 2007). In the CBS algorithm, a modification of the binary segmentation method is used to address this problem, where each segment is spliced to form a circle. Once each chromosome is partitioned, the copy number of the segments is estimated using additional information, such as the ploidy of the chromosome and the location of copy number aberration is reported (Olshen et al., 2004).

The first application of CBS algorithm to a brain tumour methylation array data was described in Sturm et al., 2012. This custom approach used the sum of the methylated and unmethylated signal at each probe. The log-ratio of the test samples intensities to the median value of a set of control samples intensities were calculated and sample noisiness was determined as the median absolute deviation of adjacent probes. Probes were combined by joining 20 adjacent probes, creating genomic windows that were iteratively merged until windows of greater than 100kb and smaller than 5Mb were created. Any genomic windows of more than 5Mb were excluded from analysis. For each remaining genomic window, the median probe value was calculated and shifted to minimise the median absolute deviation from all windows to zero for every sample. After this normalisation, the genome was then segmented using the CBS algorithm.

2.6. Copy Number analysis

Copy number aberrations (CNAs) are genomic alterations caused by structural genomic rearrangements such as duplications, deletions, translocations and insertions that result in abnormal number of one or more genes. Copy number analysis refers to the analysis of data produced by a test for DNA copy number variation in patient samples. Copy number variation can be detected with various types of tests such as fluorescent in situ hybridization, comparative genomic hybridization (Kallioniemi et al., 1992) and with high-resolution array-based tests based on array comparative

genomic hybridization (or aCGH), SNP array (McCarroll et al., 2008; Pinkel et al., 1998) and methylation array technologies (Bibikova et al., 2011). The methods used to assess copy number developed in this thesis were primarily focused on Illumina Human Methylation 450k arrays, which were used for both method development and discovery. Affymetrix SNP6.0 arrays (McCarroll et al., 2008) were used to provide "gold-standard" method for comparison.

A custom approach to CNA detection that used the sum of both methylated and unmethylated signals, initially described in Sturm et al. (2012), was the first of a kind approach that became a fundamental part of the methods used in this project, as well as multiple other methods developed by various research groups.

In this approach, probes were combined by joining 20 adjacent probes, and resulting genomic windows smaller than 100kb were merged in an iterative way with adjacent windows of smaller size. Windows of more than 5Mb were excluded from analysis, consistent with the approach employed by Sturm et al. This process partitioned the genome into 8,654 windows. For each window, the median probe value was calculated and shifted to minimise the median absolute deviation from all windows to zero for every sample. Segmentation was performed by applying the circular binary algorithm (Olshen et al., 2004, section 2.5). The median value of windows contained in each segment was calculated, and classified as homozygous or hemizygous deletion, neutral, gain or high-level amplification by using empirically-derived thresholds.

2.7. Conumee

Conumee is a package for copy number analysis using Illumina 450k and EPIC DNA methylation arrays. The conumee package is easily integrated with other Bioconductor packages, such as minfi (see 2.3.2), which is recommended for preprocessing and loading of data. Segmentation is performed by an implementation of circular segmentation algorithm (see 2.5) in the DNACopy package. The conumee package also creates publication-grade copy number plots of the whole genome, selected chromosomes or pre-defined regions of interest and a set of text-based output files for downstream processing and visualisation in other tools, such as IGV browser.

The conumee processes flow chart is shown on Figure 2.4. Before CNV analysis can be performed, data is loaded as an Mset object and an annotation object is created.

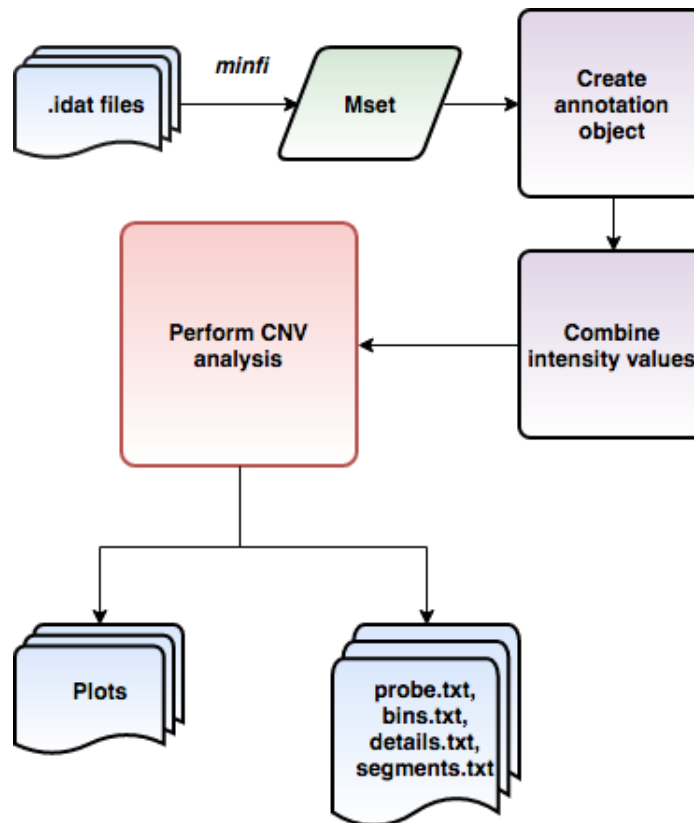


FIGURE 2.4. conumee flow chart

This object contains information such as minimum number of probes per bin, the minimum bin size, defined regions to be excluded from analysis and regions to be examined in detail, such as known oncogenes or tumour suppressor genes, where focal amplification or deletion may be expected. This object is only generated once, irrespective of the number of samples required for analysis. In the next step, for each array/tumour sample, "methylated" and "unmethylated" signal intensities are combined.

The main CNV analysis step consists of four parts.

- Firstly, a single query sample is normalised to a set of control samples by multiple linear regression and the \log_2 -ratio of probe intensities of the query sample versus combination of control samples are calculated.
- Secondly, probes are combined within genomic bins predefined in the annotation object. Intensity values are shifted to minimise median absolute deviation from all bins to zero in order to determine the copy number neutral state.
- Thirdly, the detailed regions predefined in annotation object are analysed. This step is optional, but required if the regions should be output in plots and text files.

- Finally, the genome is segmented into regions of the same copy-number state functions from Bioconductor DNACopy package.

The output of the package consists of plots of the complete genomes, one or multiple chromosomes or individual detailed regions, as predefined in the annotation object as well as four types of text output files, containing information about probes, bins, details specified in annotation files, and segments.

2.8. High-throughput sequencing: RNA-seq

RNA-seq is an approach to transcriptome profiling based on next-generation sequencing (NGS) which provides measurements of levels of transcripts and their isoforms (Wang et al., 2009b). RNA-seq allows to look at alternative gene spliced transcripts, post-transcriptional modifications, gene fusion, mutations/SNPs as well as changes in gene expression over time and differences in gene expression in different groups or treatments (Maher et al., 2009). RNA-Seq can characterise different populations of RNA, including total RNA, mRNA transcripts, small RNA, such as miRNA, tRNA and ribosomal profiling (Ingolia et al., 2012). RNA-Seq is also used to determine exon/intron boundaries and verify or amend previously annotated 5' and 3' gene boundaries (Lee et al., 2014).

2.9. Survival analysis in medulloblastoma

Survival analysis is a set of statistical approaches used to analyse data that measures the time to an event of interest. Survival analysis is used to analyse data collected prospectively in time such as cohort study or data collected retrospectively and can answer questions such as the proportion of the population surviving past a certain time without experiencing an event, at what rate the event occurs and what characteristics can influence the probability of survival.

In cancer, there are different types of events such as disease progression, relapse or death of the patient.

Time is defined as the time from the beginning of the observation, e.g. diagnosis or commencing of the treatment to (i) an event, or (ii) end of the study, or (iii) loss of contact or withdrawal from the study.

Survival analysis focuses on the duration of the time until the occurrence of an event of interest. In cases where the end of a study occurred, the patient was healthy

when last seen, or if a patient is lost to follow up during the study before the event happened, their event time is not observed. These incomplete observations can't be ignored but have to be handled differently. Handling of these situations in survival analysis is called *censoring*.

In cancer studies, two outcomes are typically measured: Overall survival (OS) and Event-free or Progression-free survival (EFS or PFS) which for medulloblastoma is defined as the time to death or relapse respectively. In medulloblastoma studies, PFS is favoured over OS due to long-term survival after relapse being rare and the time between relapse and death being variable, due to the factors unrelated to the disease severity, since treatments following disease relapse may focus on palliation or life-extension. Consequently, PFS is a more accurate measure.

Two probabilities are used to describe survival data: *survival probability* and *hazard probability*.

Survival probability, or the survival function $S(t)$, is the probability that a patient survives from the start time (e.g. diagnosis or beginning of the treatment) to a specific time point t in the future.

The hazard probability, or hazard function $h(t)$, is the probability that a patient under observation at a particular time point t has an event at this time.

Most cancer survival analyses utilise the following methods:

- *Kaplan-Meier plots* for survival curves visualisation
- *Log-Rank test* for comparison of survival curves for two or more groups
- *Cox Proportional Hazard models* to describe the effect of various factors on survival

Kaplan-Meier plots and log-rank test are examples of univariable analysis, which describe the survival taking into account only one factor and ignoring the impact of any others. This type of analysis, described in 2.9.1, is particularly useful when a variable is categorical, e.g. one medulloblastoma subgroup vs another, but is not applicable to quantitative variables, such as age or gene expression.

An alternative method is multivariable analysis. A multivariable analysis method, capable of assessing the effect of several risk factors on survival time, both categorical and quantitative, simultaneously, used in this research is Cox Proportional Hazard

regression analysis and is described in 2.9.2. Cox proportional hazards models are also appropriate to analyse single variables in univariable analysis.

2.9.1. Univariable analysis: Kaplan-Meier plots and log-rank tests . The Kaplan-Meier method is a non-parametric estimate of survival probability from observed survival times (Kaplan and Meier, 1958). It is used to measure the proportion of patients for whom an event did or didn't occur for a certain period of time, for example from diagnosis or beginning of treatment. With medulloblastoma, it is important to subdivide the patient cohort according to specific clinico-pathological or molecular features, and study of the survival within molecular subgroups has enabled the investigation of de-escalation of treatment intensity of patients with good prognosis.

An example of a Kaplan-Meier plot is shown on 2.5. The survival probability is a step function which changes value (drops down) only when the event (relapse) occurs. Censored individuals are shown on the plot as vertical lines or pluses. The table below the graph is life-table or at-risk table and shows the population size for each risk group in 6 month intervals.

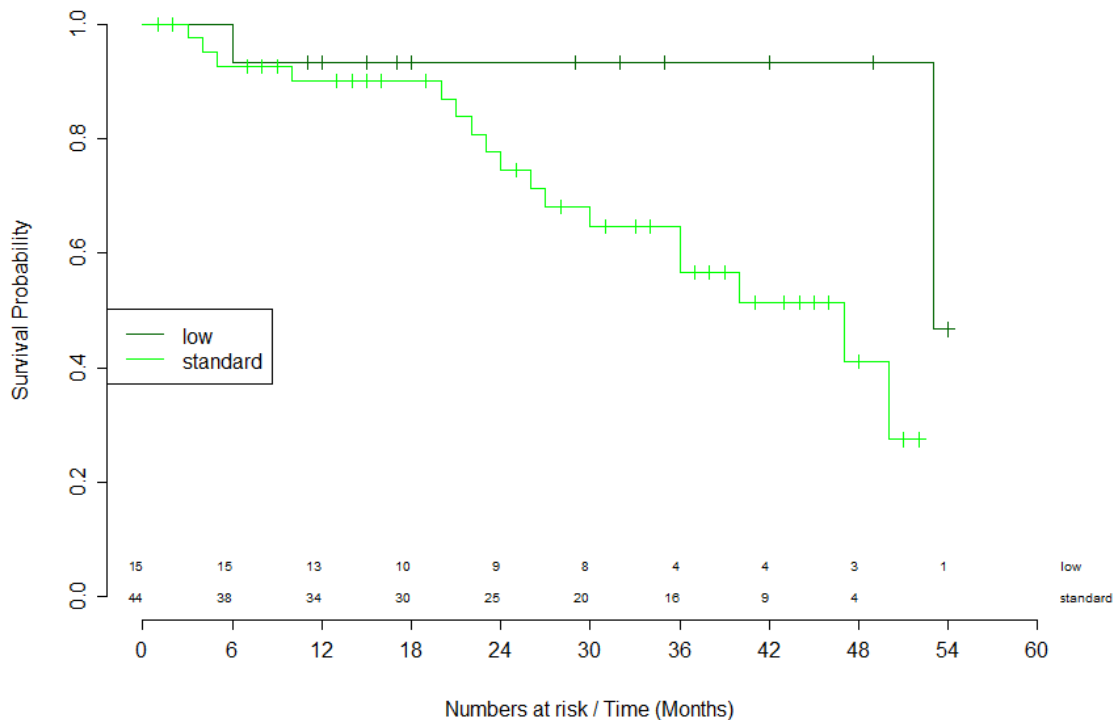


FIGURE 2.5. Kaplan-Meier plot of hypothetical cancer cohort assessing difference in survival between two groups of patients

The log-rank test (Bland and Altman, 2004; Mantel, 1966) is a non-parametric test that compares survival times of two or more groups and is applicable to censored data. It tests the null-hypothesis that survival for the groups in question is the same and determines if the observed number of events in groups is significantly different from the expected numbers. The log-rank statistic has an approximately chi-squared distribution with one degree of freedom and a chi-squared test of the null hypothesis can be used to calculate a p-value.

2.9.2. Multivariable analysis: Cox proportional hazards models . The Cox Proportional Hazards model (Cox, 1972) is the most commonly used multivariable approach for analysing survival data in clinical research. It is a regression model that is used in statistical cancer research in order to investigate association between patient survival (expressed by hazard function) and predictor variables. This method is suitable for both quantitative and categorical variables as well as for the univariable or multivariable assessment of several risk factors on survival time (Bradburn et al., 2003).

The Cox model is expressed by the hazard function $h(t)$ which can be interpreted as a risk, of an event, such as death or relapse, happening at a particular point of time t . The purpose of the model is to evaluate the effect of several factors on survival at the same time.

When interpreting a Cox model, the following factors for each covariate are reported: hazard ratio, confidence interval and p-value. Hazard ratio (HR) also known as exponential coefficient shows the effect of a covariate. $HR=1$ means the covariate has no effect on survival, a covariate with $HR<1$ indicates a reduction in the hazard and is often called a good prognostic factor in cancer studies, $HR>1$ indicates an increase in hazard (a bad prognostic factor). The confidence interval is calculated in order to determine whether or not this interval includes 1, indicating a non-significant covariate. An example output from a Cox model analysis is shown in the Table 2.3

2.9.3. Median follow-up time . The duration of the follow-up is a measure of the quality of a clinical trial and is measured as the median follow-up time. It represents how long, on average, patients have been followed. Simply finding a median of all the survival times, whether censored or not, has a disadvantage: a trial with many early deaths but long observation period would appear not to have a long median follow-up

Covariate	Hazard Ratio (HR)	95% CI	p-value
Age	1.01	0.99-1.03	0.23
Sex	0.58	0.41-0.79	0.00098
Histology	1.59	1.27-1.98	<0.0001

TABLE 2.3. An example Cox model from a hypothetical dataset that includes age, sex and histology as covariates. The p-value for age is over the 5 percent significance cut-off and the confidence interval for HR includes 1, indicating it does not contribute significantly to the model; conversely, sex and histology are significant factors ($p < 0.05$ for both). Being a certain gender may be associated with good prognosis and a particular histological variant is associated with poor survival.

time. A way to overcome this disadvantage is the so called "reverse" Kaplan-Meier method, which looks at median survival as the "potential" median survival. In this method the censoring indicators are reversed, that is, a "censored" observation becomes the "event" and *vice versa*. Then a Kaplan-Meier "survival" estimate is computed using this reversed censoring indicators and the median survival is found, which will be the median follow-up time.

2.10. Confusion matrices

One way to determine if a classification algorithm or method produces satisfactory results is by building a confusion matrix.

A confusion (error) matrix is a specific table layout that visualises the performance of an algorithm or method. The rows of the matrix represent the instances in the predicted class and columns represent the actual, true class or *vice versa* (Stehman, 1997; Powers, 2011), example of such a table can be seen in Table 2.4. A table of confusion is a 2x2 contingency table, that contains the number of incorrectly predicted instances: false negatives (FN) and false positives (FP), and the number of correctly predicted instances: true positives (TP) and true negatives (TN).

Based on these four descriptors, several further metrics can be calculated: sensitivity (true positive rate (TPR) or recall), specificity (true negative rate, TNR), positive predictive value (PPV, also known as precision), negative predictive value (NPV) and accuracy. Sensitivity and specificity show the ratio of positive and negative cases correctly identified by the methods. Positive and negative predictive values are conditional probabilities that a positive and negative cases are predicted as positive and negative, respectively. Accuracy is the proportion of correct predictions (Vihinen, 2012).

		Actual class		
		Positive	Negative	Metric
Predicted class	Positive	True positive TP	False positive FP	Positive predictive value (PPV) $TP/(TP+FP)$
	Negative	False negative FN	True negative TN	Negative predictive value (NPV) $TN/(FN+TN)$
	Metric	Sensitivity $TP/(TP+FN)$	Specificity $TN/(FP+TN)$	Accuracy $(TP+TN)/(TP+FP+FN+TN)$

TABLE 2.4. Table of confusion displaying outcome of predictions (in blue) and statistical metrics calculated based on it. Table adapted from Vihinen (2012)

2.11. Over-fitting

Survival prediction from methylation microarray experiment data and other high-dimensional methylomic or genomic data are associated with the problem of having a much larger number of potential prognostic covariates than patient samples in the study, potentially leading to over-fitting (Bovelstad et al. 2007). This is common problem in cancer research, especially in cancers with relatively low incidence, such as medulloblastoma, where cohort sizes are typically small. Over-fitting can be defined as ”the production of an analysis which corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably” (Bovelstad et al., 2007; Gareth James, Trevor Hastie, Robert Tibshirani, 2014).

2.12. Cross-validation

In the absence of independent validation cohorts, one approach to avoid over-fitting is to use a process called cross-validation. Cross-validation is a set of methods for measuring performance of a predictive model on a new dataset. In this project k-fold cross-validation (k-fold-CV) was used to reduce the possible effects of over-fitting. The k-fold-CV method evaluates performance of the model on different subsets of training data and then calculates the average prediction error rate. This approach randomly divides the set of observations into k sub-sets, or folds, of approximately equal size. The first fold is treated as a validation set, and the model is fit on the remaining k-1 folds. This procedure is repeated k times using a different group of observations as a validation set each time (Gareth James, Trevor Hastie, Robert Tibshirani, 2014).

2.13. Selection of testable, potentially prognostic methylation probes

In order to identify potentially testable methylation probes and to avoid unnecessary statistical testing, unsuitable probes with low variability were filtered out. The 20,000 most variably methylated CpG probes (measured by standard deviation) were selected as candidates for investigation of their prognostic potential.

2.14. Bimodality

An idealised methylation markers would be mainly either hypermethylated or hypomethylated, and follow the bimodal distribution with modes at 0 and 1 that correlate with survival (Fan and Chi, 2016). In order to select the most bimodal probes, a measure called bimodality index (Wang et al., 2009a) can be calculated for the most variably methylated probes and probes were ranked in order of bimodality.

2.15. GLMnet

It is not difficult to find predictors that perform excellently on the fitted data, but fail in external validation data not used in model training as result of over-fitting (Gareth James, Trevor Hastie, Robert Tibshirani, 2014). The standard linear model performs poorly in this situation. The effective way to handle this problem with high-dimensional data applied to survival analysis is through using cross-validation-based methods that utilise Cox proportional hazards models. Parameter estimates were obtained by a dimension reduction or parameter shrinkage estimation technique. This is known as penalised regression. That is, a linear regression model that is penalised for having too many variables by adding a constraint in the equation (Gareth James, Trevor Hastie, Robert Tibshirani, 2014). The rationale for imposing this penalty is to reduce (*i.e.* shrink) the coefficient values towards zero in order to remove coefficients that are less important by setting their contribution to zero or close to zero. The most commonly used penalised regression methods are ridge regression, lasso regression and elastic net regression.

2.15.0.1. *Ridge regression.* Ridge regression shrinks the regression coefficients, so that variables with minor contribution to the outcome have their coefficients close to zero. A penalty term called L2-norm, which is the sum of the squared coefficients, is used for shrinkage. Ridge regression performs better when the outcome is a multiple

predictor function with coefficients of similar size (Gareth James, Trevor Hastie, Robert Tibshirani, 2014).

2.15.0.2. *LASSO*. LASSO (Least Absolute Shrinkage and Selection Operator) regression uses a penalty term called L1-norm, which is the sum of the absolute coefficients, to force the coefficients with minor contribution to the model to be exactly zero. LASSO regression may perform better when the predictor function has a mixture of large and small coefficients.

2.15.0.3. *Elastic Net*. Elastic Net regression penalises regression model using both the L1-norm and L2-norm. As a result it effectively shrinks some coefficients (like in ridge regression) and sets some coefficients to zero (as in LASSO).

2.16. Concordance index

Concordance index, also known as C-statistic, or C-index, is a measure of the goodness of fit for binary outcomes in a logistic regression and is used to measure how well a biomarker predicts the time to an event (Brentnall and Cuzick, 2018).

In survival analysis, the C-index gives the probability that a randomly selected patient who experienced an event (e.g. a relapse or death) had a higher risk score than a patient who had not experienced the event. The C-index is equal to the area under the Receiver Operating Characteristic (ROC) curve and ranges from 0.5 to 1 (Figure 2.6). The value of 0.5 (red line on the Figure 2.6) means that the model is no better in predicting an outcome than a random chance; values over 0.7 indicate a good model, values over 0.8 indicate a strong model and a value of 1 (blue line on the Figure 2.6) means that the model perfectly predicts which group members will experience an event and which will not. C-index is paired with confidence interval and in general, any results are not significant if the confidence interval includes 0.5 even if it includes the C-index (Hosmer and Lemeshow, 2005).

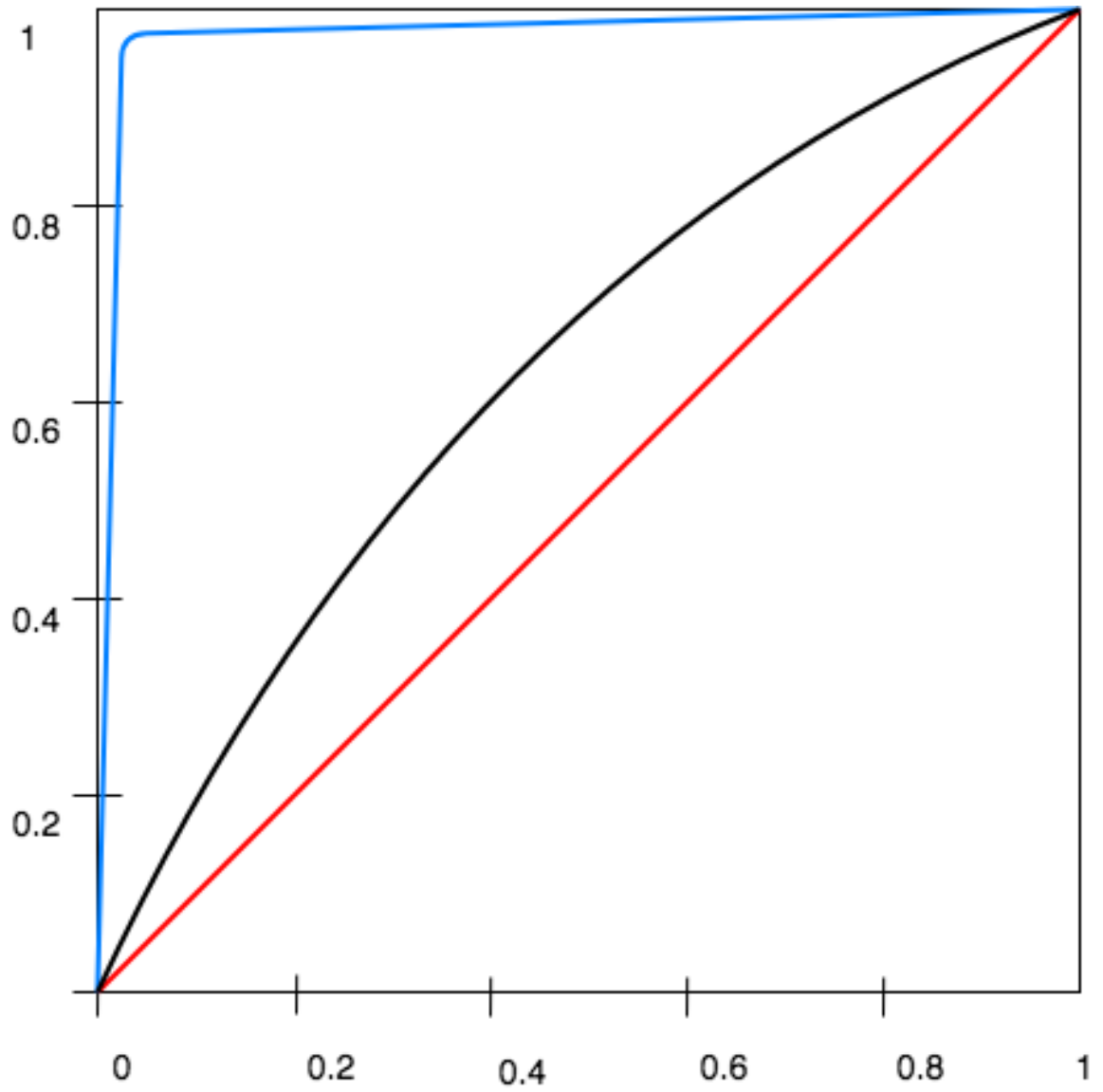


FIGURE 2.6. The concordance statistic is equal to the area under the ROC curve (shown in black for a hypothetical cancer survival cohort). Red line represents a model with C-index=0.5, blue line represents a model with C-index close to 1

2.17. Statistical Analysis

Statistical analyses were performed using the R statistical programming language (version 3.4.4) in RStudio integrated development environment (version 1.0.143). Statistical significance was assessed using t tests, Fisher's exact test, and the Chi-squared test. Survival associations were tested using log-rank tests (section 2.13.2) and Cox proportional hazards test (section 2.14). Pearson's product-moment correlation coefficient (r) was calculated to assess the direction and strength of the linear association between CpG site methylation and the level of gene expression. Concordance between gold standard CN calls from SNP arrays and methylation CN estimates was measured using Cohen's kappa statistic, a measure that assesses agreement between measures while allowing for agreements by chance (a kappa of 0-0.20 is considered as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1 as almost perfect agreement by Landis and Koch (1977). Fleiss considers kappa > 0.75 as excellent, 0.40-0.75 as fair to good, and < 0.40 as poor (Fleiss, 1981). 'P' values were used to assess significance, with the cut-off of 0.05. Further details of specific analyses are discussed in subsequent chapters where appropriate.

Development and implementation of the workflow and methods for copy number analysis using Illumina Human Methylation arrays

3.1. Introduction

Genomic changes, such as copy number alterations (CNAs) as well as epigenomic ones, e.g. DNA methylation patterns have been described and associated with the development and progression of many human cancers, including medulloblastoma (Northcott et al., 2009; Thompson and Fuller, 2006; Nord et al., 2012).

The first platforms for assessing genomic copy number changes were described in the 1990s and refinements over the following years increased genomic resolution (Pinkel et al., 1998; Jiang et al., 2012). At the same time, various methods were developing for the assessment of epigenetic alterations, specifically for locus-specific DNA methylation. Initially, these platforms were based on immunoprecipitation (MeDIP), or enzymatic digestion followed by hybridisation to a bacterial chromosome (Irizarry et al., 2008; Weber et al., 2005). Subsequently, microarrays using the same technology as SNP detection arrays were designed for bisulfite converted DNA in order to establish methylation status (Bibikova et al., 2009). Subsequently, the most common platforms for assessment of genome-wide methylation status have been the third and fourth generation Illumina Infinum Human Methylation arrays: HumanMethylation450k BeadChip array (Sandoval et al., 2011), covering 485,000 CpG sites, now discontinued, and replaced with the MethylationEPIC BeadChip array (Moran et al., 2016) which now assays 850,000 CpG loci.

DNA methylation profiling is now successfully used in classification and subgrouping of all brain tumour entities (see section 1.3.11, (Capper et al., 2018)) and is now routinely applied in clinical diagnostics, where the standard way to archive tissue material is formalin-fixed paraffin-embedded tissue (FFPE). FFPE is an excellent way to prepare tissue for histopathological assessment and to store samples for decades at

room temperature. However, formalin fixation causes fragmentation and cross-linking of the DNA (Gilbert et al., 2007) and, therefore, DNA extracted from FFPE samples is often of poor quality and degraded to various degrees which makes analysis of such DNA challenging.

Integration of genomic and epigenomic data has become commonplace, in order to gain a better understanding of the complex mechanisms underlying the development and progression of cancer. However, the use of separate platforms to assess both copy number and DNA methylation status presents a number of issues: This not only increases the cost of analysis, which is the most obvious problem, it also increases the amount of specimen used, which in the case of the brain tumours, is a limiting factor, due to the often scant amounts of material available for analysis (Feber et al., 2014; Gerlinger et al., 2012). Therefore, there is a need for methods that would allow an integrated genomic and epigenomic data analysis of the same DNA sample, saving precious sample materials and reducing costs.

Previous research showed that Illumina Human Methylation arrays are a suitable platform for integrated genomic and epigenomic analysis of FFPE solid tumour samples (Feber et al., 2014; Sturm et al., 2012; Kling et al., 2017). Genotyping SNP arrays, alongside whole-genome sequencing, are the currently accepted gold standard for assessment of aneuploidy. The Infinium methylation arrays measure methylation status by quantifying methylation-dependent SNPs induced by treatment of DNA with sodium bisulfite (see 2.2), and their high genomic coverage enables the repurposing of these arrays to assess CNA as well as DNA methylation profiles (Feber et al., 2014).

One of the main purposes of this project was to find a way to leverage genomic information from a large, historic medulloblastoma archive of the Newcastle University Brain Tumour Group, which consists of over 500 unique samples that have already been processed using Illumina Infinium HumanMethylation450k arrays. This cohort consists of both fresh frozen and FFPE samples, collected from the CCLG and other sources, mainly from Europe.

This required methods for genomic CN changes detection specifically applicable to Illumina Infinium HumanMethylation450k methylation arrays and the subset which were profiled from FFPE tissue.

This chapter focuses on developing a CNA analysis workflow that includes newly-developed methods as well as the incorporation of previously reported methods, tailored specifically for CNA analysis using FFPE samples processed on the Illumina Methylation microarray platform. This workflow is modular and extendable by design in order to be easily tailored and updated for use with emerging advances in DNA methylation analysis. Robust QC methods are described, and the workflow is applied to the Newcastle Medulloblastoma archive.

3.2. Aims

The aims of the work reported in this chapter are:

- Design and implement a bioinformatics workflow and develop and validate methods for assessing disease-specific copy number changes using Human Methylation arrays that enable integrated genomic and epigenomic data analysis from the same sample on a single platform.
- Develop and implement Quality Control methods that would enable robust copy number analysis using Human Methylation arrays for clinical material archived as FFPE tissue.
- Extend and adapt existing methods for copy number analysis using methylation arrays to enable them to identify focal, locus-specific copy number alterations.
- Develop a method for automatic assessment of large-scale DNA copy number changes, such as chromosomal arm level gain/loss.
- Develop methods for automatic reporting of both large-scale and narrow, focal abnormalities identified in both human and machine-readable copy number profiles.

3.3. Materials and methods

3.3.1. Cohorts. A test cohort of 135 medulloblastoma samples run on both Illumina HumanMethylation450 arrays and gold-standard genotyping Affymetrix SNP 6.0 arrays was used for the copy-number analysis. Methylation array-derived large-scale, chromosomal arm level CN calls were compared against gold-standard calls. Calls made from SNP6.0 genotyping arrays were available after QC and processing; this was performed, prior to this project, and was kindly provided by Dr S. Nakjang of the Bioinformatics Support Unit, Newcastle University. The Bioconductor package *DNAcopy* was used to assess DNA copy number. Cut-offs for calling genomic gain/loss were set based on the median of the \log_2 ratio for each array $+2\sigma$ or -2.5σ (where σ is the standard deviation of the \log_2 ratio for each array), as described in (Curtis et al.,2012).

The two most commonly focally amplified oncogenes in medulloblastoma, *MYC* and *MYCN*, were used to test the methods developed during this project for detection of focal locus-specific high-level amplifications. A cohort of 203 samples that passed

pipeline-specific quality control was used for the evaluation of the pipeline's suitability and performance for detection of focal amplification. The reference "true" *MYC* and *MYCN* copy number statuses were previously established by a consensus call using both fluorescence in situ hybridisation (FISH) molecular cytogenetic technique and multiplex ligation-dependent probe amplification (MLPA) and was performed by Dr J. Lindsey and Dr R. Hill.

3.3.2. Overview of copy number analysis workflow . The bioinformatics workflow described in Figure 3.1 was designed for automated CNA analysis of DNA derivatives from FFPE and fresh-frozen tissue samples that were processed using methylation arrays, and included the following functions:

- Preparation of input data for downstream analyses, which included pre-processing of raw intensities data input in .IDAT format (see 2.3), general and pipeline specific quality control.
- Performing CNA analyses of medulloblastoma samples using bioinformatics methods developed for this project as well as methods developed using existing bioinformatics tools.
- Extraction of relevant output of CNA analysis methods and assembly of CNA profiles.

3.3.3. Input data preparation. The first stage of the workflow consists of two main steps: initial universal preprocessing and quality control and pipe-line specific quality control.

STEP 1: The first step of initial input data preparation stage was developed as a set of methods implemented as R-scripts that perform microarray data pre-processing and quality control steps described in section 2.3, for mainstream CNA analysis. It is designed and implemented as an R wrapper-method for standard bioinformatics methods and tools implemented in Bioconductor package minfi and includes the following steps:

- reading red and green channels .IDAT files and creating RGset object to store the information loaded from these files;
- running the standard minfi initial QC report;

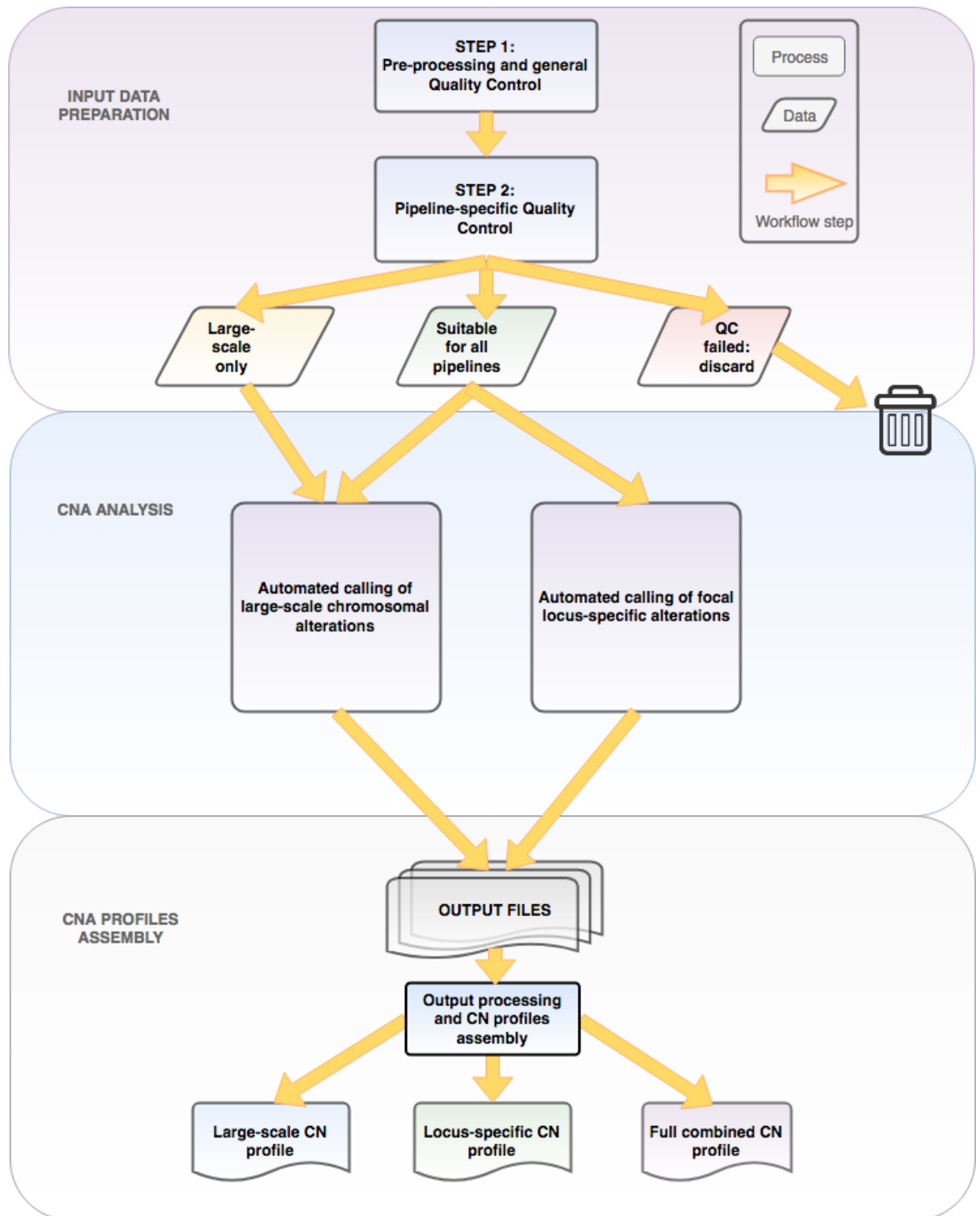


FIGURE 3.1. Summary of copy number analysis workflow. The overall workflow consists of three stages: (1) Input data preparation, (2) Copy number alteration analysis, and (3) Copy number alteration profile assembly. The first stage is divided into two steps: initial pre-processing and quality control (see section 2.3 and additional sample quality control (see section 2.4)). This additional quality control step was designed during this project in order to separate DNA samples processed using methylation arrays of varying degree of copy-number quality; some samples which were unsuitable for locus-specific CNA analysis, could still be analysed to identify large-scale chromosomal CNA. The second stage can be divided into two pipelines, large-scale and locus-specific CNA analysis. Each of these pipelines can be run independently. The third stage involves processing the output of the previous stage and CN profile assembly.

- preprocessing the raw data and performing standard initial normalisation using a choice of standard minfi methods: Illumina, SWAN, Quantile, Funnorm and Noob (see section 2.3.2 for details).
- removal of probes with potentially-confounding SNPs within their sequence, sex-chromosomes and cross-reactive probes.

By default, this step includes the removal of sex chromosomes. However, in some cases, it is desired to retain the sex chromosomes for copy number analysis. In this instance, an additional, method-specific normalisation step, which is performed during CN analysis stage and described in the next section, is required. For the analysis, a set of control samples is required. These control samples should be obtained from normal diploid tissue; if sex chromosome CNA is required, the control cohort needs to be filtered to contain samples from a single sex, to avoid confounding by mixed ratios of the X chromosomes in male and female samples.

The output of this step will be either a `GenomicMethylSet` or `GenomicRatioSet` object (see 2.3.2), depending on the normalisation method chosen.

Step 2 is designed as an additional, pipeline-specific sample quality control step that separates pre-processed sample data into DNA quality-dependent subsets of data. The method for pipeline-specific sample quality control method was developed as an R implementation of calculation of the outlier threshold using DLRS (see Section 2.4). This step is necessary when working with FFPE samples, whose DNA derivatives are often degraded and therefore unsuitable for identification of focal, locus-specific amplifications, but from which larger-scale whole chromosome arm level alterations can be detected. Quality of the DNA samples, e.g. level of degradation, is reflected in the noisiness of the resulting log ratio data, which is measured by DLRS. Tumour cell content, such as normal cell contamination or clonal heterogeneity does not correlate with DLRS, DNA extracted from FF and FFPE samples from the same tumour, however will have different DLRS values, higher for FFPE. During this step, the DLRS of each sample is calculated and compared to a predefined set of thresholds in order to determine suitability for a particular type of the CNA analysis (see 3.3.4 below). The samples are separated into subsets based on their suitability for a particular method of CNA analysis and marked accordingly.

3.3.4. Pipeline specific sample quality control . This method was developed as an R implementation of the method for calculation the outlier threshold using DLRS (see section 2.4) in order to identify samples suitable for focal, locus-specific CNA analysis and samples that are unsuitable for focal analysis but which are still amenable to large-scale, arm-level CNA.

The method calculates the DLRS of log ratio values of each sample. Samples are graded using the following empirically-determined cut-offs, based on recommendations from GoldenHelix (2014):

- Less than 0.2 - excellent quality DNA, likely derived from fresh-frozen or blood sample;
- 0.2-0.3 - good quality DNA, suitable for processing through both pipelines;
- 0.3 - 0.5 - fair quality DNA, unsuitable for locus-specific CNA analysis, but still acceptable for processing with chromosomal arm-level pipeline;
- Over 0.5 - significantly degraded DNA. Unsuitable for analysis using the methods developed in the project;

3.3.5. CN analysis methods. The second stage of the workflow performs the CNA analysis. The following CNA analysis methods were developed for this project:

- automated detection of large-scale, chromosomal arm-level copy number changes;
- automated detection of very narrow, focal locus-specific copy number changes from a set of multiple DNA samples simultaneously.

The initial implementation of the methods for automated calling of both focal and large-scale chromosomal alterations was based on an implementation of the circular binary segmentation (CBS) algorithm for CNA analysis using methylation arrays described in 2.5. The original method was published by Sturm et al. (2012) as a custom approach for copy number calling in glioblastoma (referred to as "original" further in the text) and later re-developed and implemented as a Bioconductor package called *conumee* by the members of the same research group. The calling of CNAs using *conumee* was implemented (referred to as "*conumee*"-based further in the text) as an option within the CN calling pipeline described in this chapter, which demonstrated the extensibility of the workflow.

The techniques developed during this project are described below.

3.3.5.1. *Methods for automatic assessment of CNA on chromosomal arm-level* .

One of the initial requirements of the project was to develop a method for automated detection of one of the most common chromosomal defects found in medulloblastoma patients: isochromosome 17q (where the p-arm is lost and q-arm is gained for chromosome 17, present in 40% of cases), as well as the whole chromosome status of other commonly aberrant chromosomes, e.g. chromosome 7 gain, chromosome 8 and 11 loss. The method was developed using chromosome 17 as a test chromosome, and then the method was further extended for use in automatic processing of all human chromosomes, including acrocentric chromosomes. Acrocentric chromosomes are autosomes that have a very short p-arm; in the human genome, these are chromosomes 13, 14, 15, 21 and 22 . The 450k methylation arrays contain no probes mapping to p-arms of the acrocentric chromosomes and therefore these regions remain unassessed.

In order to assess copy number status of a chromosome arm, the median probe log ratio (LRR) for each arm of the chromosome were calculated and then chromosomal arms were assessed individually as neutral, lost or gained using cut-off values that were initially estimated and then optimised by comparing methylation-derived CN calls against gold-standard Affymetrix SNP6.0 arrays calls that were available from the same samples from the same cohort.

Optimal cut-offs for calling chromosomal gain and loss from conumee-derived LRR values from methylation arrays were determined by maximising agreement with confirmed gain or loss from gold standard Affymetrix SNP6.0 arrays. These optimised LRR cut-offs are: loss (this included both hemizygous and homozygous deletions) called when median LRR is below -0.22, normal, unchanged copy number - between -0.22 and 0.12, between 0.12 and 0.6 – gain, more then 0.6 – high-level amplification (the last cut-off was only used to determine focal high-level amplifications, see section 3.3.5.2 and 3.2).

3.3.5.2. *Methods for assessing focal amplification* . The methods described in this section were developed to fulfil one of the aims of this project - the detection of focal, locus-specific CN changes.

The original method by Sturm et al. (2012) was modified to allow "zooming in" to the area ± 100 kbp of a locus of interest and was wrapped as a standalone R-script. However, this method relies heavily on visual curation of the output plots to identify

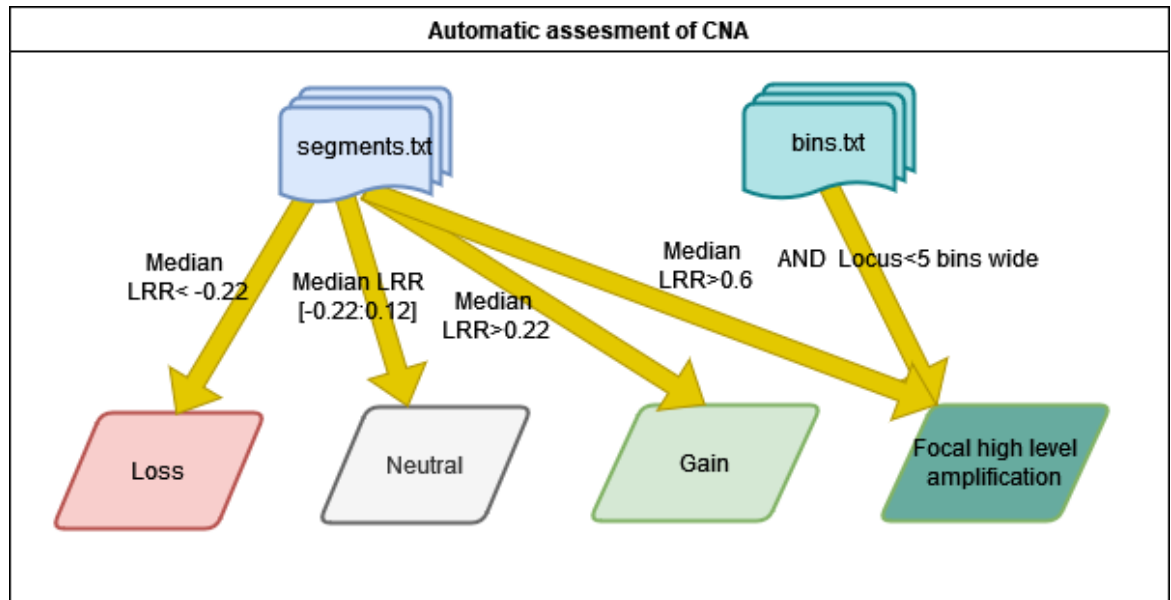


FIGURE 3.2. Copy number changes assessment: loss (both hemizygous and homozygous deletions) called when median LRR is below -0.22, normal, unchanged copy number - between -0.22 and 0.12, between 0.12 and 0.6 – gain, more then 0.6 – high-level amplification, used to determine focal high-level amplifications in section 3.3.5.2)

amplification and was therefore not amenable to automation and, consequently, is not a good choice for using with a large number of samples, since it is labour intensive and relies on subjective user opinion.

In order to automate the process for detection of focal CN changes, a new method, based on the Bioconductor package *conumee* was developed. One of the outputs from the Conumee copy number changes detection method is a text file containing the mean LRR values for every single 50kb genomic bin. By examining the genomic coordinates of the gene of interest, it is possible to extract mean log-ratio values for only the bins adjacent to and spanning the gene and compare them to the values of adjacent regions, highlighting regions with 5 bins amplified both upstream and downstream of the locus of interest as "attention required" as this might indicate large-scale CNA, thereby detecting focal changes in DNA copy number in a semi-supervised manner. The cut-off used for the high-level amplification detection was 0.6 (see section 3.3.5.1 and Fig 3.2).

3.3.6. CNA profiles assembly methods. In the third stage (module) of the workflow, the output files from the CNA analysis stage are processed, copy number

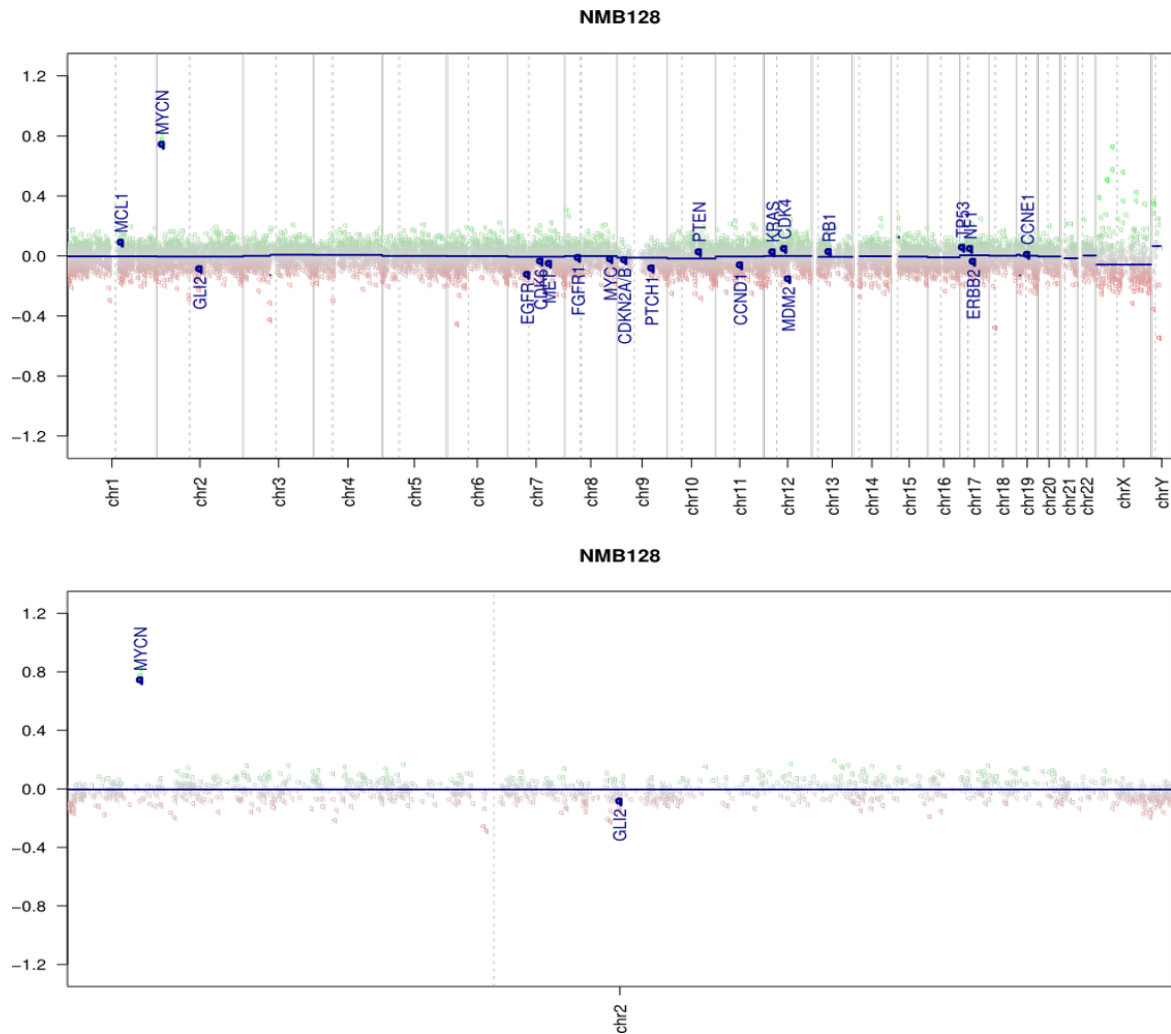


FIGURE 3.3. Conumee graphic output: Whole genome view (top) of a medulloblastoma sample and zoomed to chromosome 2 view of the same medulloblastoma sample. The *MYCN* gene is amplified but amplification is not picked up by the standard method (horizontal blue line, indicating genomic segmentation is still close to zero but *MYCN* is positioned well above the horizontal line since it is amplified).

information for each processed sample is extracted and copy number profiles are assembled. Four types of profiles are generated as output of this stage, which are grouped as follows:

- Large-scale CN profiles (Profiles type I and II)
- Locus-specific CN profiles (type III)
- Full combined CN profiles (type IV)

The full combined profile is a combination of both large-scale and locus-specific profiles for the same dataset.

3.3.6.1. *Large-scale CN profiles*. CN profiles can be generated as arm-level (profile I) and whole chromosome calls (profile II).

P-arm/Q-arm	Q-arm Loss	Q-arm Gain	Q-arm Neutral
P-arm Loss	Whole chromosome loss	Isochromosome-q*	P-arm loss
P-arm Gain	Isochromosome-p**	Whole chromosome gain	P-arm gain
P-arm Neutral	Q-arm loss	Q-arm gain	Balanced
Cut-offs***:			
Focal amplification: >0.6	Loss: <-0.22	Gain: >0.12	Neutral: [-0.22:0.12]

TABLE 3.1. Rules for whole chromosome status assessment from statuses of individual chromosomal arm. *-this is currently restricted to chromosome 17 only as isochromosome-q re-arrangement is only confirmed for this chromosome in medulloblastoma; **- provisional rule, no isochromosome-p rearrangements have been reported in medulloblastoma; ***-loss, gain and neutral state cut-offs are applicable to both arm and whole chromosome

- *Profiles type I: Whole genome arm-level status profile.* This profile type contains the following information: list of sample IDs, p- and q-arm status of either all chromosomes or only autosomes for each sample. P-arm and q-arm statuses are either arm-level loss, gain, neutral (normal) or QC fail if the DLRS status of a sample exceeds 0.5.
- *Profiles type II: Individual chromosome status profile.* Whole chromosome status is inferred from chromosomal arm statuses and determined according to the rules presented in table 3.1. This profile is either an optional direct output from the large-scale CNA pipeline or can be generated later using whole genome arm-level status profile as an input.

3.3.6.2. *Locus-specific CN profiles* . Locus-specific CN profiles (Profiles type III) are generated by the focal, locus-specific pipeline. This profile type contains the individual genes of interest for each sample in the sample list. The potential outputs for an individual gene/locus are 0 when there are no CN changes detected, 1 - CN change detected, "attention required" - this flags up the situation when the algorithm is unable to verify the changes and manual curation of the sample using a graphic output is required to confirm potential CN change and "QC fail". The latter indicates that the sample did not pass DLRS sample quality control step and its DLRS status exceeds 0.3, thereby making any focal estimation of CN unreliable. This profile assembly method takes the output of stage II of the workflow as its input and the profiles are generated as a direct output of the locus-specific pipeline.

3.3.6.3. *Profile type IV: full combined CN profiles* . This profile assembly method is not pipeline-specific and combines outputs from large-scale and locus-specific pipelines.

Profiles generated using this method contain joint information from profiles of type I and III for the same set of samples: list of sample IDs, p-arm status and q-arm status of either all chromosomes or only autosomes, and CN statuses of individual gene loci for each sample in the sample list.

3.3.7. CNA analysis pipelines developed in this project . The copy number analysis workflow employed a set of methods that were developed specifically for the project, as well as existing bioinformatics tools and methods. These methods were organised into the following CNA analysis pipelines:

- Large-scale (chromosomal arm-level) CNA pipeline
- Focal, locus-specific CNA pipeline

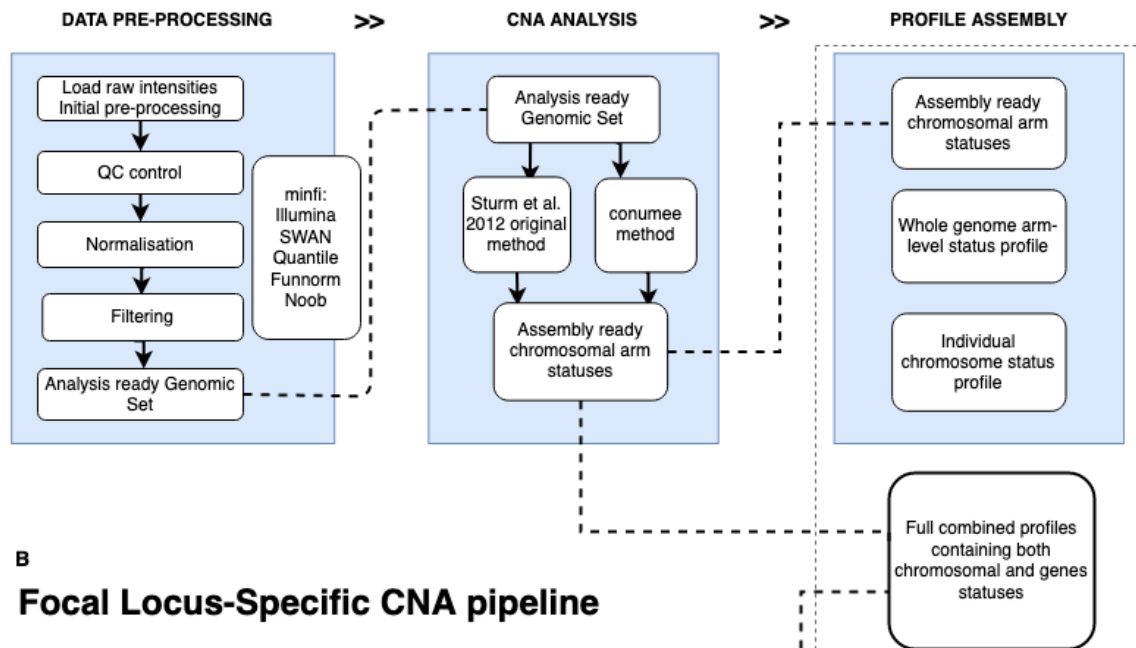
These pipelines bind together different bioinformatics tools and allow the analysis to be performed in appropriate order with no or minimal external involvement. In this project, all bioinformatics pipelines required a set of raw intensity files. Pipeline-specific analysis results can be organised into pipeline-specific profiles as well as into combined cross-pipeline profiles that consist of a combination of results from both pipelines. The pipelines follow the same general workflow scheme.

3.3.7.1. *Pipeline for automated calling of large scale chromosomal alterations.* One of the goals of this project was to develop a robust method for automated detection of large-scale chromosomal structural defects frequently seen in medulloblastoma, such as chromosomal arm gains and losses and the automatic inference of whole chromosome status that can be classified as balanced (normal, unchanged chromosomal CN status), whole chromosome loss, gain and isochromosomes (when one of the arms lost and the other gained).

This pipeline is presented on Figure 3.4, A.

- *Data pre-processing.* The pipeline loads raw intensity data using standard *minfi* loading functions and performs two-tier quality control: the general quality control offered by external *minfi* package quality control options (section 2.3.2), as well as pipeline-specific sample quality control using DLRS (section 2.4), developed for this project. Initial implementation of general quality control allowed Illumina and SWAN normalisation, however, along with the later version of *minfi* becoming available, Quantile, Funnorm and Noob normalisation options have been added. The output of this stage of the pipeline

A

Large-Scale (arm-level) CNA pipeline

B

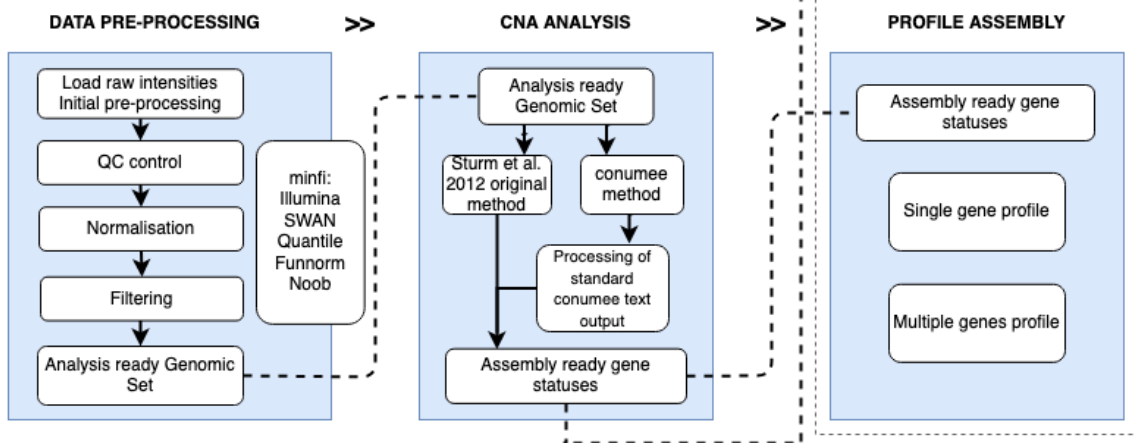
Focal Locus-Specific CNA pipeline

FIGURE 3.4. Copy number analysis pipelines: Large scale (chromosomal arm-level) pipeline (A) and focal locus-specific CNA (B)

can be passed directly onto the next stage for CNA analysis or stored as an RData class file for future use.

- *CNA analysis.* This stage of the pipeline uses the output from the data pre-processing stage or previously pre-processed data. The CN status of chromosomal arms can be established by either methods described in section 3.3.5.1 (in-house developed method based on description found in Sturm et al. (2012) and based on external package *conumee*). The output of this stage is a data frame of mean log-ratio values of chromosomal arms which are either passed

directly to the next stage for profile assembly or, alternatively, can be stored as a .csv file for future use, as well as for creation of plots of whole genome and individual chromosomes of interest and, in the case of the *conumee*-based method, additional information contained in text format, as described in section 2.7

- *Profile assembly.* This stage uses the data frame output from the CNA analysis stage as an input and defines chromosomal arms statuses using cut-offs introduced in section 3.3.5.1 and assembles profiles of type I (arm-level) and II (whole chromosome-level). Methods for this stage are specifically developed for this project.

3.3.7.2. *Pipeline for automated calling of focal, locus-specific alterations.* Another goal of this project was to develop a method capable of detecting focal amplifications of commonly amplified oncogenes in medulloblastoma, such as *MYC* and *MYCN*. These focal changes are difficult to detect using methylation arrays, especially from FFPE archived samples due to the combination of factors such as the changes being relatively narrow (*i.e.* less than 10 Mb) and are often missed during segmentation process (see Figure 3.3 for a representative example), as well as the inherent noisiness of FFPE samples processed on methylation arrays.

This pipeline consists of two sub-routines each based on the methods described in section 3.3.5.1 (Figure 3.4 B).

- *Data pre-processing.* This stage of the pipeline is identical to data pre-processing stage of the arm-level pipeline.
- *CNA analysis.* This stage uses the output from the previous stage to establish the CN status of focal, locus-specific changes and employs the methods described in 3.3.5.2. The original method is not automated and produces a graphical output suitable for manual curation. The alternative, *conumee*-based method uses Conumee's text output to determine CN status of genes of interest in a semi-automated way, and only requires manual curation when a gene's status is flagged as "attention required". The output from this method is passed directly to the profile assembly stage.

- *Profile assembly.* In this stage the type III profile (narrow locus-specific amplifications) is assembled and the gene statuses marked "attention required" can be manually curated if needed.

3.3.8. Performance and suitability evaluation metrics. The metrics used to evaluate suitability and performance of the methods developed for this project relied on statistical classification methods such as confusion matrices (also known as error matrix), tables of confusion (sometimes also called confusion matrices) and four values calculated based on the above: sensitivity, specificity, positive and negative predictive values (see 2.10).

To assess the performance of the workflow in term of calling both large-scale and locus-specific CNA, calls were compared to gold-standard calls derived from Affymetrix SNP6 arrays and these statistical metrics were calculated as shown in Table 2.4.

3.4. Results

This section describes suitability and performance of the workflow and associated methods for CNA analysis of DNA samples processed on Illumina Methylation arrays archived as FFPE tissue and the summary of the results generated from each developed pipeline.

3.4.1. Pipeline specific sample quality control. The representative examples of medulloblastoma samples shown in Figure 3.5 were graded according to the cut-offs described in section 3.3.4:

- Less than 0.2: excellent quality DNA, likely derived from fresh-frozen or blood sample (Figure 3.5, A) and 0.2-0.3 - good quality DNA (Figure 3.5, B). These samples are suitable for processing through both large-scale and locus-specific pipelines;
- 0.3 - 0.5 - fair quality DNA, unsuitable for locus-specific CNA analysis but still acceptable for processing with chromosomal arm-level pipeline (Figure 3.5, C);
- Over 0.5 - significantly degraded DNA. Unsuitable for analysis using the methods developed in the project (Figure 3.5, D).

For the focal changes pipeline the use of external and in-house methods is similar to the large-scale pipeline.

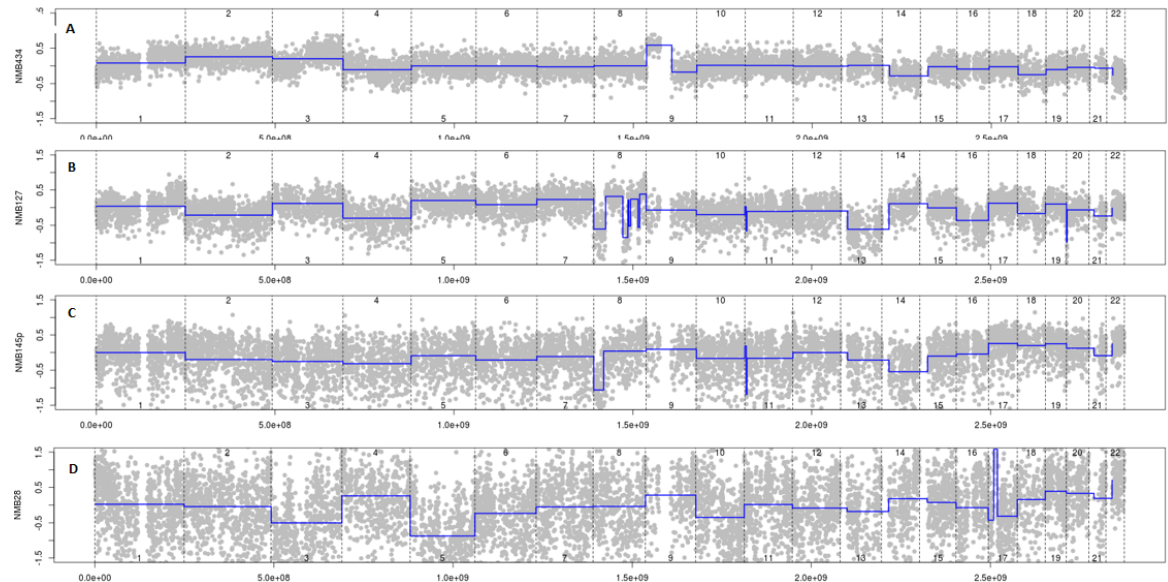


FIGURE 3.5. Pipeline specific quality control: A - excellent quality DNA, DLRS status 0.18 (Fresh Frozen sample); B - very good quality DNA, DLRS status 0.22 (FFPE), suitable for both arm-level and locus-specific pipelines; C- fair quality DNA, DLRS status 0.36. Unsuitable for locus-specific pipeline, but still acceptable for large-scale arm level pipeline; D - largely degraded DNA, unsuitable for analysis using pipelines developed in this project. The blue lines show genomic segmentation

3.4.2. Evaluation of suitability and performance of methodology for automated calling of large scale chromosomal alterations. The most common chromosomal defect in medulloblastoma, isochromosome-17q (where the p-arm is lost and q-arm is gained, present in 40% of cases), was used to compare methylation array CN calls to SNP6.0 calls.

Agreement between methylation arrays derived CN calls and gold-standard CN calls derived from SNP6.0 arrays for chromosomal arm-level events was measured using Cohen's kappa statistic (see Section 2.17), which measures concordance of data values in the main diagonal of a cross-tabulation of chromosomal calls from 450k and SNP6.0 arrays, and then adjusts these values for the agreement expected to happen by chance. Results of a grid search for cut-off values between ± 0.1 - 0.25 (- for arm loss, + for arm gain) were represented as a heatmap with kappa values varying on the scale from dark blue (poor) to dark red (excellent) in Figure 3.6.

Concordance between the gold standard and methylation CN estimates was good. The cut-offs shown in Table 3.2; maximised kappa (kappa=0.892) with the confusion matrix shown in Table 3.3, and therefore these settings were considered optimal.

The results represented in the confusion matrix (Table 3.3) mean that, out of 83 aneuploidy calls made from methylation arrays, a total of 77 calls were concordant with

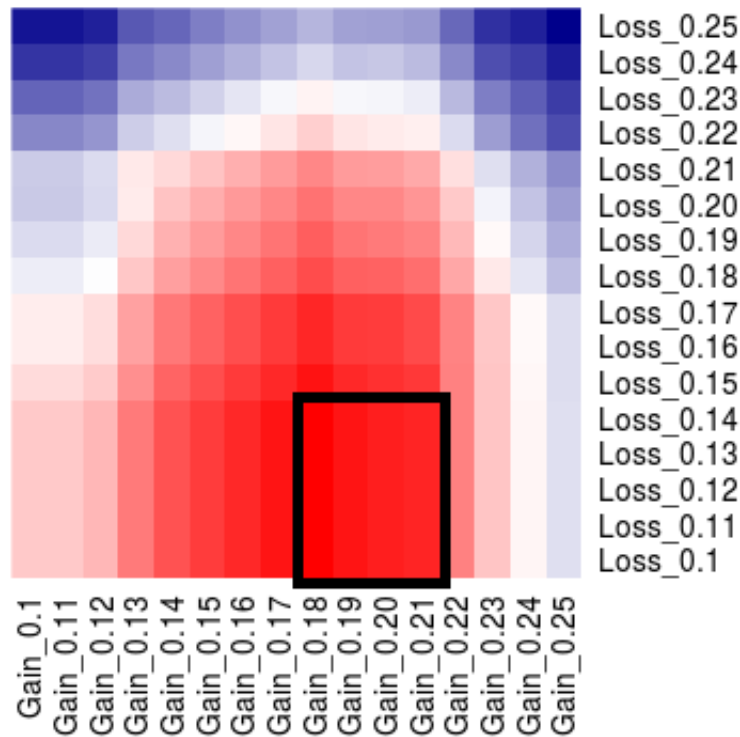


FIGURE 3.6. Results of grid search for cut-offs between ± 0.1 - 0.25 (- for arm loss, + for arm gain) represented as a heatmap with kappa values varying on the scale from dark blue (poor) to dark red (excellent). The box over the red area shows the area of optimal cut-offs.

Status	Arm Loss	Arm gain	Neutral
Median Log Ratio	Less than -0.14	More than 0.18	[-0.14-0.18]

TABLE 3.2. Optimal cutoffs for median LRR for calling arm gain / loss from 450k array data

SNP6.0 array							
		balanced	i17q	p-loss	q-gain	WCG	Total
HumanMethylation450k array	balanced	15	-	-	3	3	21
	i17q	-	40	-	-	-	40
	p-loss	-	-	4	-	-	4
	q-gain	-	-	-	17	-	17
	WCG	-	-	-	-	1	1
	Total	15	40	4	20	4	83

TABLE 3.3. Confusion matrix showing performance of CNA analysis workflow for Illumina Methylation arrays in comparison with gold standard SNP 6.0 arrays using optimal cut-offs defined in Table3.2). WCG - whole chromosome

SNP6-derived calls (the number of concordant calls stratified by CNA type are shown in bright yellow). Using the confusion matrix above (Table 3.3), its corresponding performance metrics, calculated as described in Table 2.4 are presented in Table 3.4.

Representative examples of the comparison of copy number estimation between Illumina Human Methylation 450k and SNP6.0 (images of SNP6.0 samples are courtesy of Dr S. Nakjang, Bioinformatics Support Unit, Newcastle University) arrays are

	TP	FP	TN	FN	PPV	NPV	Sensitivity	Specificity
Balanced	15	6	62	0	71.42%	100%	100%	91.17%
i17q	40	0	43	0	100%	100%	100%	100%
p-loss	4	0	79	0	100%	100%	100%	100%
q-gain	17	0	63	3	100%	95.45%	85%	100%
WCG	1	0	79	3	100%	96.34%	25%	100%

TABLE 3.4. Performance of the project's pipeline for automated calling of large scale chromosomal abnormalities. TN = True Positive, FN = False Negative, TP = True Positive, FP = False Positive, PPV = Positive Predictive Value, NPV = Negative Predictive Value

demonstrated in Figures 3.7 to 3.11 for assessing chromosome 17 CN status using the adaptation of the original method described by Sturm et al. (2012). The vertical axes on all the figures represent log-ratios of sample and the horizontal axes show genomic position.

Balanced chromosome 17 – both mean and median LRR are close to zero (Figure 3.7). This is a flat diploid chromosome without any copy number alterations.

Representative plots showing tumours with i17q, 17p loss, 17q gain, WCG (whole chromosome gain) are shown in figures 3.8-3.11, respectively.

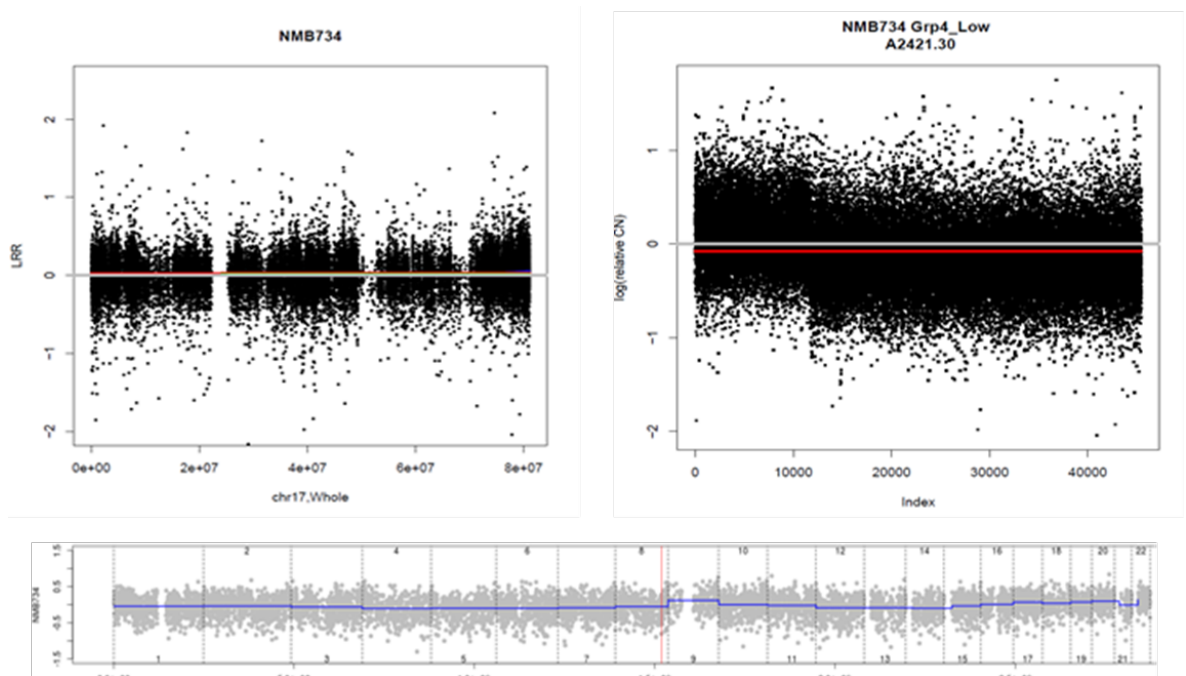


FIGURE 3.7. Balanced chromosome 17; Illumina Human Methylation 450k (top left), Affymetrix SNP6.0 (top right), whole genome view (bottom, Illumina Human Methylation 450k). In the whole-genome view, individual numbers of chromosomes are also indicated on horizontal axes with blue lines showing genomic segmentation. On the individual chromosome view, red lines represent median log-ratios, green - mean.

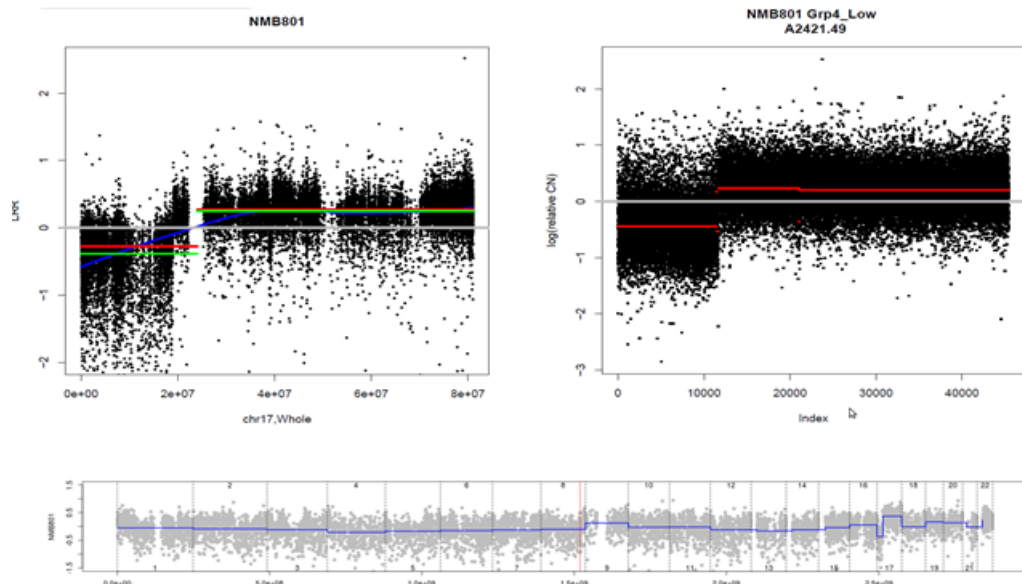


FIGURE 3.8. Isochromosome 17q; Illumina Human Methylation 450k (top left), Affymetrix SNP6.0 (top right), whole genome view (bottom, Illumina Human Methylation 450k). In the whole-genome view, individual numbers of chromosomes are also indicated on horizontal axes with blue lines showing genomic segmentation. On the individual chromosome view, red lines represent median log-ratios, green - mean.

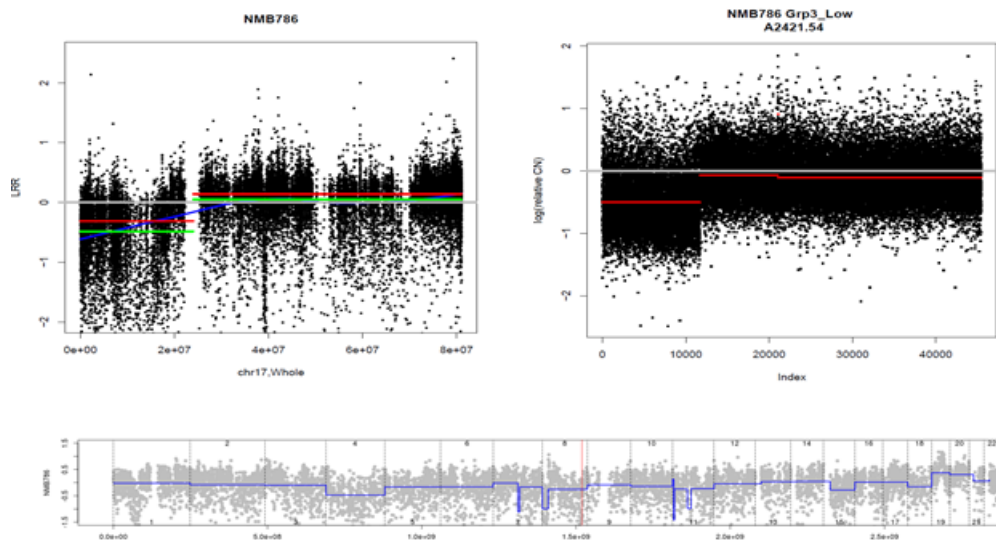


FIGURE 3.9. Chromosome 17 p-arm loss; Illumina Human Methylation 450k (top left), Affymetrix SNP6.0 (top right), whole genome view (bottom, Illumina Human Methylation 450k). In the whole-genome view, individual numbers of chromosomes are also indicated on horizontal axes with blue lines showing genomic segmentation. On the individual chromosome view, red lines represent median log-ratios, green - mean

Figure 3.12 shows chromosome 17 calls made from the same samples as shown on figures 3.7-3.11 using the conumee method.

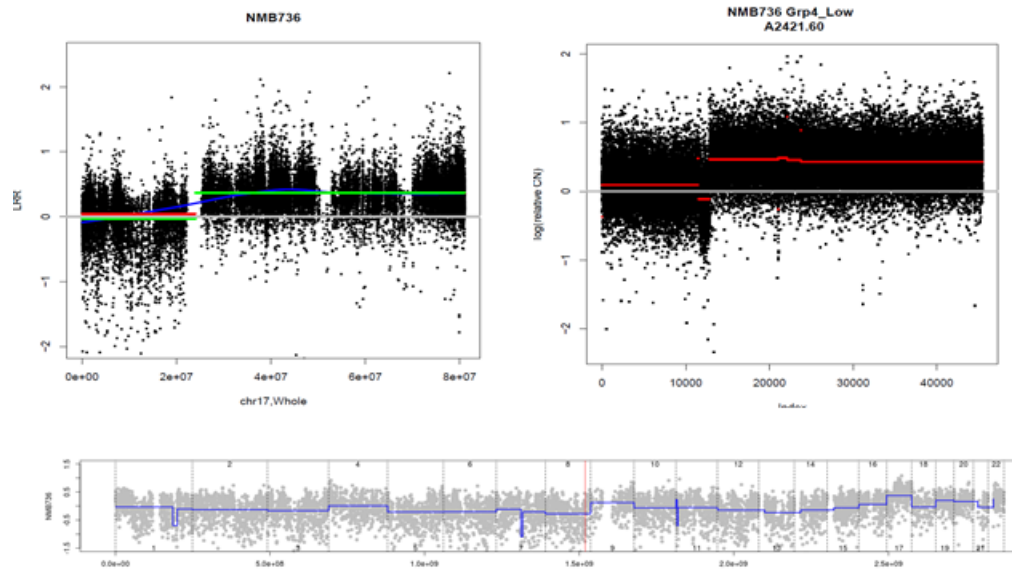


FIGURE 3.10. Chromosome 17 q-arm gain; Illumina Human Methylation 450k (top left), Affymetrix SNP6.0 (top right), whole genome view (bottom, Illumina Human Methylation 450k). In the genome-wide view, individual numbers of chromosomes are also indicated on horizontal axes with blue lines showing genomic segmentation. On the individual chromosome view, red lines represent median log-ratios, green - mean

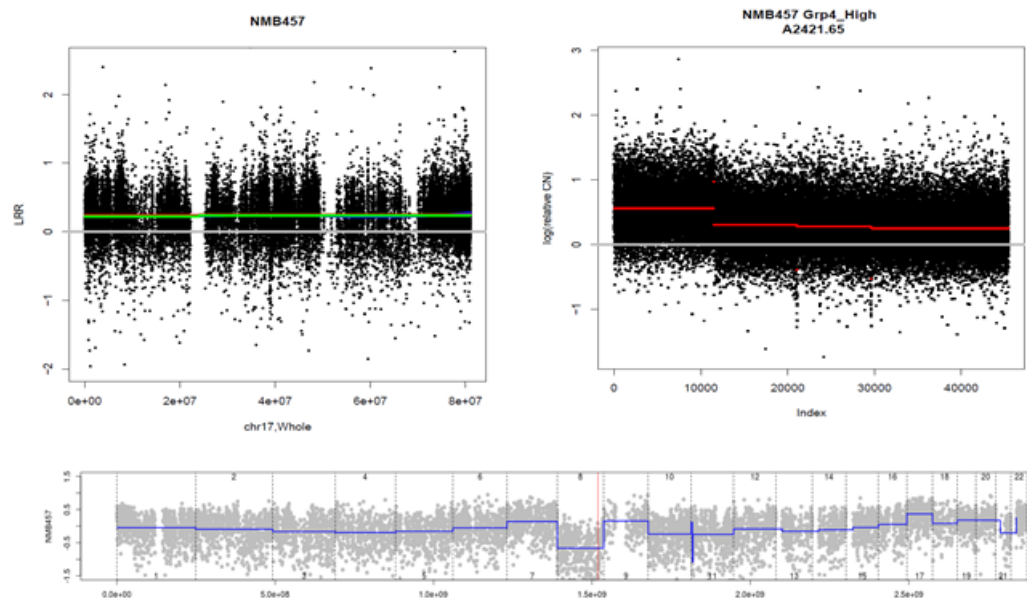


FIGURE 3.11. Whole chromosome 17 gain; Illumina Human Methylation 450k (top left), Affymetrix SNP6.0 (top right), whole genome view (bottom, Illumina Human Methylation 450k). In the genome-wide view, individual numbers of chromosomes are also indicated on horizontal axes with blue lines showing genomic segmentation. On the individual chromosome view, red lines represent median log-ratios, green - mean

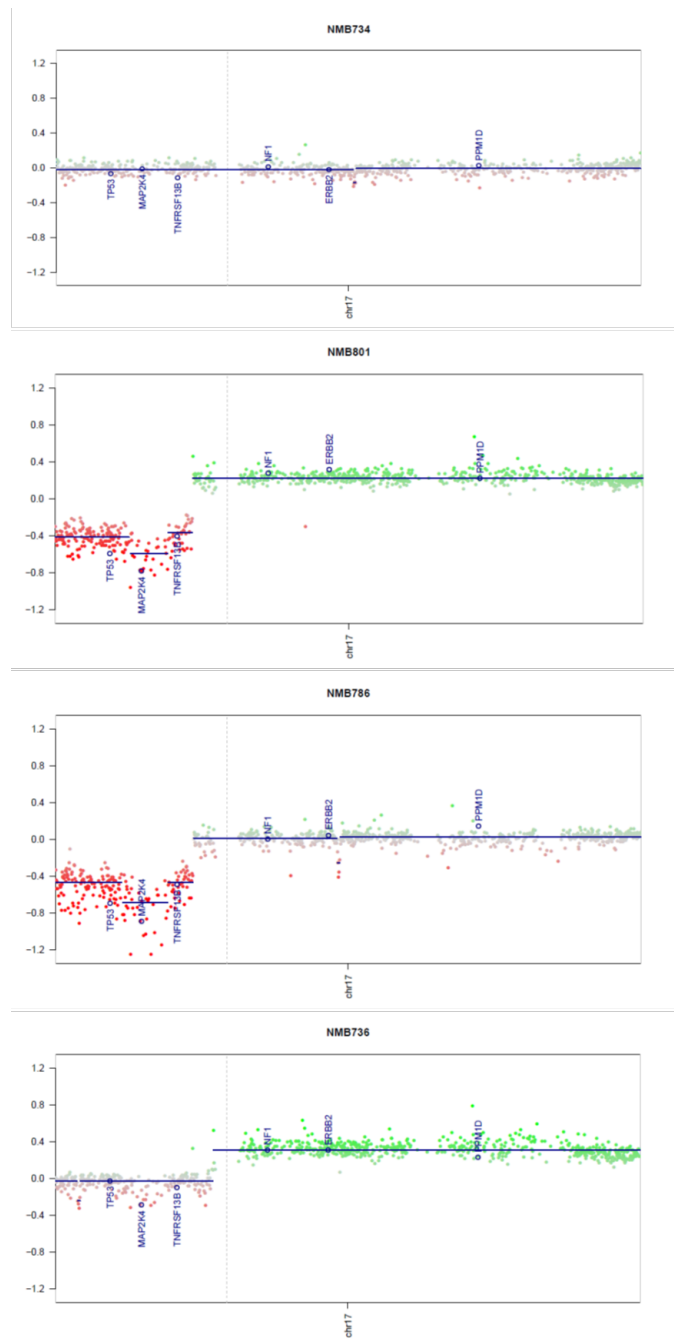


FIGURE 3.12. Calls made from Illumina Human Methylation 450k using method based on the Bioconductor package conumee. From top to bottom plots: balanced chromosome 17, isochromosome 17q, 17p loss, and 17q gain.

3.5. Evaluation of semi-automated calling of focal, locus-specific alterations

The confusion tables, describing the performance of automatic locus-specific CNA analysis pipeline for calling focal high-level *MYC* or *MYCN* oncogenes amplifications are presented in Table 3.5.

The corresponding performance metrics, calculated as described in Table 2.4 are presented in the Table 3.6

Human Methylation 450k array, automated locus-specific pipeline	Reference		
	<i>MYC</i> amplification	Positive	Negative
	Predicted positive	5	2
	Predicted negative	2	194
	<i>MYCN</i> amplification	Positive	Negative
	Predicted positive	14	2
Predicted negative	5	182	

TABLE 3.5. Confusion table describing performance of the CNA locus-specific pipeline. TN = True positive, FN = False negative, TP = True positive, FP = False positive.

	TP	FP	TN	FN	PPV	NPV	Sensitivity	Specificity
<i>MYC</i>	5	2	194	2	71.42%	98.97%	71.43%	98.98%
<i>MYCN</i>	14	2	182	5	87.5%	97.32%	73.68%	98.91%

TABLE 3.6. Performance of the pipeline for semi-automated calling of locus specific focal amplifications. TN = True Positive, FN = False Negative, TP = True Positive, FP = False Positive, PPV = Positive Predictive Value, NPV = Negative Predictive Value

The positive predictive rate of 71.42% for *MYC* and 87.5% for *MYCN* means that out of 100 samples predicted to be *MYC* or *MYCN* amplified about 29 and 12 were not respectively *MYC* or *MYCN* amplified. The negative predictive rate of 98.97% for *MYC* and 97.32% for *MYCN* means that out of 100 predicted *MYC* or *MYCN* negative samples about 1 sample was actually positive for *MYC* and about 2 per 100 were actually positive for *MYCN* amplifications, respectively.

The sensitivity of 71.43% and 73.68% for *MYC* or *MYCN* amplification respectively means that out of 100 *MYC* or *MYCN* amplified samples 29 and 27 cases of amplification respectively were not detected.

The specificity of 98.98% for *MYC* and 98.91% for *MYCN* means that out of 100 of samples called not amplified for *MYC* or *MYCN* respectively, only one of each was actually amplified.

Representative examples of the comparison between Illumina Human Methylation 450k calls made using the locus-specific pipeline are shown on Figure 3.13.

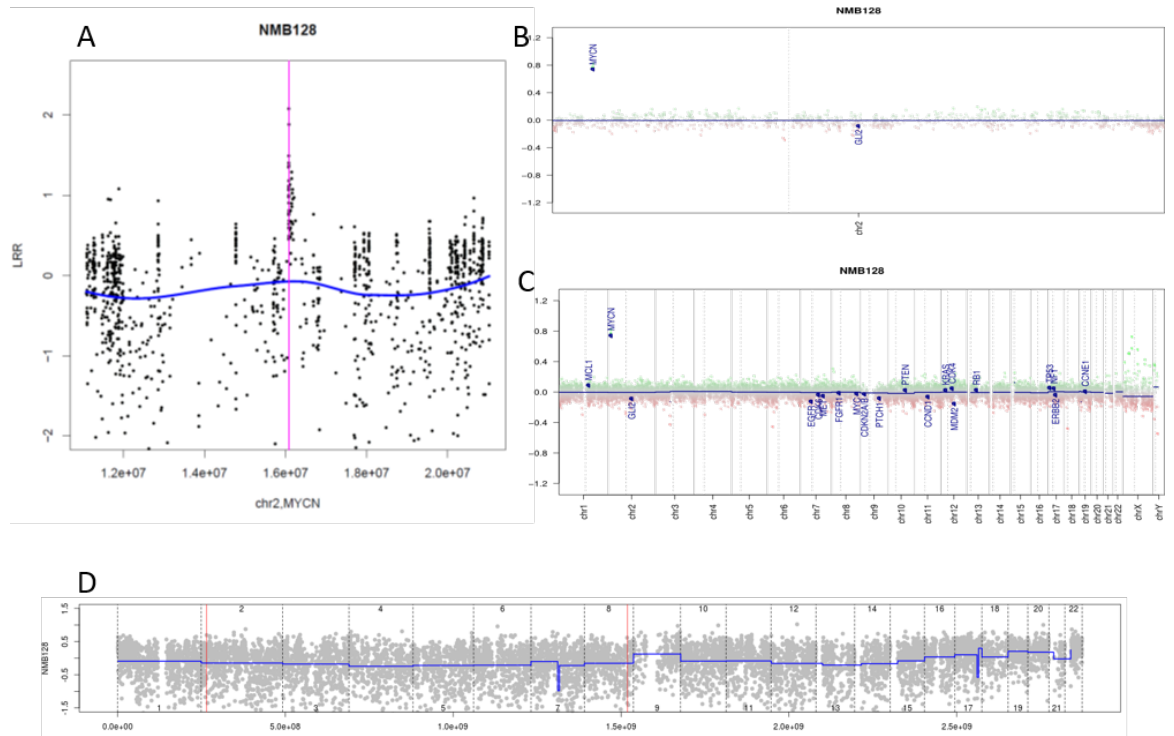


FIGURE 3.13. Focal high-level amplification of *MYCN* oncogene. Sturm et al. (2012) based method: zoomed-in plot of chromosome 2 at *MYCN* locus (A) and whole-genome view (D); conumee-based method: zoomed-in plot of chromosome 2 at *MYCN* locus (B) and whole-genome view (C).

3.6. Discussion

The Newcastle Medulloblastoma archive currently contains over 500 unique patient DNA samples processed using Illumina HumanMethylation450k arrays and much of the DNA was archived as FFPE tissue. Most of the samples that comprise this archive are collected from various sources over a large number of years and for many of them, not enough tissue is available to perform any additional analysis.

This chapter investigated the utility of the methylation arrays as a single platform tool for integrated genomic and epigenomic analysis of DNA samples. A bioinformatics workflow that uses methods, both developed specifically for this project as well as existing methods, tailored for working with FFPE samples processed on methylation arrays, was designed and implemented.

The ability to generate genomic copy number data from DNA extracted from FFPE tissue using methylation arrays offers potential to perform confirmatory or investigatory work on historic archival cohorts for which FF tissue is not generally available and scarce availability of archived material does not allow perform additional analysis using multiple platform and methods.

3.6.1. Development of the bioinformatics workflow and methods: rationale and implementation. With the recent advances of cancer research at a molecular level, CNAs have been studied as important components in the initiation, development, and progression of cancer. Many researchers have focused on the detection of chromosomal regions having amplifications and deletions using arrays and sequencing data. At the outset of this project state of the art methods for CNA were such methods as GISTIC2 (Beroukhi et al., 2007; Mermel et al., 2011) and its rival WIFA (Hur and Lee, 2011).

GISTIC2 (Genomic Identification of Significant Targets in Cancer) is a statistical approach for identifying regions of aberration that are more likely to drive cancer pathogenesis. The method identifies those regions of the genome that are aberrant more often than would be expected by chance, with greater weight given to high-amplitude events (high-level copy-number gains or homozygous deletions) that are less likely to represent random aberrations. This method includes identification of focal as well as larger and arm-level CN aberrations (Mermel et al., 2011).

WIFA (wavelet-based identification of focal genomic aberrations) uses the wavelet analysis, and is a multi-resolution approach, that integrates multiple cancer samples so that it enables the detection of the consistent aberrations across multiple samples (Hur and Lee, 2011).

However, these methods were developed and tested using SNP arrays, including application to medulloblastoma (Northcott et al., 2012b) and were not readily applicable to methylation arrays.

Methylation array specific approaches were limited to those described in Feber et al. (2014) and Morris and Beck (2014), which was based on method described Feber et al. (2014), and Sturm et al. (2012).

The Chip Analysis Methylation Pipeline (ChAMP) package is a pipeline that integrated currently available 450k analysis methods and also offers its own novel functionality. ChAMP takes the raw IDAT files as input, using the data import, quality control and normalisation options offered by minfi. CNAs were identified using circular binary segmentation in the Bioconductor package DNACopy (Morris and Beck, 2014; Feber et al., 2014). This is not optimised for FFPE samples.

The method described by Sturm et al. (2012) was specifically focused on the assessment of CN from DNA derived from FFPE tissues, but their publication did not provide source code for its implementation. This custom algorithm was developed for glioblastoma research and reverse-engineered for application to medulloblastoma by Dr E. Schwalbe of joint Northumbria/Newcastle University Brain Tumour research group ahead of this project (unpublished).

Therefore, this method was selected and implemented during this project, as no usable implementation was available until it was repackaged by Hovestadt and Zapatka (2015) into the Bioconductor package "conumee".

This original method was capable of detecting both large arm-level chromosomal changes and narrow high-level amplifications but it exclusively relied on visual assessment of graphic output for every sample, which was impractical when analysing large cohorts. No publications were available at the time to confirm the utility of Illumina HumanMethylation450k microarrays for use in medulloblastoma. Therefore, in order to be able to use the Newcastle Medulloblastoma archive for this project, it was necessary to develop methods that would be able to reliably call copy number

changes from methylation arrays but that would not require a researcher to visually assess potentially hundreds of plots. This was achieved by developing methods that are specifically designed for methylation arrays, optimised for FFPE archived DNA and combine functionality of gold standard methods, such as detection of narrow focal high-level amplification and arm-level CNAs.

The methods described in this project were developed as a first attempt for the automated detection of genomic CNA using methylation arrays for medulloblastoma research. This allowed rapid analysis of large cohorts. The methods included the integration of sample pre-processing, quality control, profile assembly and genomic copy number estimation methods embedded in a single modular, 3-stage workflow. Each module was developed to function in a stand-alone manner

The modules can be executed successively, forming a pipeline. The first stage conducts data pre-processing and quality control. Then pre-processed and quality controlled data can be saved or passed directly to the second stage, where copy numbers are calculated. The results again are either saved for later use or passed on to the third stage where they are collated to create copy-number profiles containing final CNA calls. This stage is semi-automated, as some results (e.g. from the locus-specific pipeline) need to be manually curated before they can be integrated into a profile. The workflow currently contains two main pipelines for large-scale (chromosomal arm level) and focal locus-specific (narrow high-level amplifications) copy number calling (Figure 3.4). Each pipeline has two alternative implementations: the first that relies on the original algorithm available at the beginning of the study and second that employs Bioconductor package *conumee* that emerged during this project, and is based, as well as the first one, on the original algorithm by Sturm et al. (2012), see also section 2.6. Both implementations of the large-scale pipeline allow fully automatic CN profiling and are able to identify both broad gains and losses, such as arm-level and whole chromosome aberrations. The *conumee* based implementation of the locus-specific pipeline for focal CN changes is semi-automatic (requires visual assessment of the graphic output in a small number of cases highlighted as "attention required" by the script). The implementation based on the original algorithm relies on visual assessment.

3.6.1.1. *Bioinformatics pipeline for large-scale copy number detection confidently identifies arm-level CNAs.* Both implementations of the large-scale CNA analysis method

were able to provide fully automated aneuploidy detection and both showed good results. Out of 83 samples, there were 6 incorrect calls using the original method's implementation; using the *conumee* implementation, 4 incorrect calls were made. Therefore, the performance of the method based on *conumee* was only marginally better than of the original method and both methods are suitable for automatic calling of whole chromosome statuses.

There are limitation of the large-scale pipeline regardless of the algorithm used such as the inability to distinguish homozygous and hemizygous deletions.

3.6.1.2. *The focal, locus-specific method is able to detect high-level amplifications.*

As this project required detection of focal high-level amplifications in a large number of samples, development of another method was needed that would be sensitive enough to pick up most amplifications while keeping the false discovery rate relatively low and would not require a researcher to visually assessing large numbers of plots.

The algorithms described in sections 2.6 and 2.7 were used again as the core of this CNA analysis module. Both algorithms allow detection of various copy number changes of different sizes as some parameters, such as number of probes in a genomic window in the original method described by Sturm et al. (2012) and the minimum number of probes per bin and minimum number of bins per segment in Bioconductor package *conumee* are adjustable to suit a particular task. However, settings of both algorithms were optimised to work with FFPE samples, which rendered them to be too insensitive to detect narrow CNA such as some frequently amplified medulloblastoma oncogenes like *MYC*, *MYCN* or *GLI2*. This allows the algorithm to identify smaller CN changes when the DNA is of relatively good quality, e.g. from frozen tumour tissue. However, in the case of noisy FFPE samples, where the DNA is partially degraded, reducing the minimum amount of bins per segment or amount of probes per bin will increase the amount of false positive calls, e.g. noise might be called as amplification.

A simple solution to detect the narrow high-level amplifications was to examine the regions of genes of interest and the regions directly up- and down-stream, which was implemented in this chapter.

The limitations of this method include a semi-automatic analysis as it still involves some degree of visual curating of the regions highlighted as "attention required" and

inability to detect a single copy gain (e.g. *SNCAIP* gene commonly gained in medulloblastoma) or loss due to the lack of sensitivity.

3.6.2. Illumina Human Methylation array is a robust and accurate platform for DNA copy number estimation. This chapter successfully employed methylome profiling technology for genome profiling task, identifying both large-scale and focal copy number changes and therefore confirmed findings of Feber et al. (2014) that Infinium arrays detect copy number alterations and show that high-density methylation arrays provide a robust and economic platform for detecting copy number and methylation changes in a single experiment.

The sensitivity and specificity of the large-scale methods compared to SNP-array derived calls was good: 85%-100% for most types of chromosomal aberrations, apart from only 25% sensitivity when assessing whole chromosome gain for the implementation of the original algorithm. However, the total number of this type of CNA in the discovery cohort was 4, which is too small for reliable statistical inference and therefore, does not reflect actual sensitivity of the method. The locus-specific pipeline showed specificity nearing 99%, however sensitivity was significantly less than in the large-scale pipeline - just over 70%.

There is no specific provision in the pipelines to account for germline CNV showing as false somatic events. However, the possibility of any focal CNA giving false positive result is minimised by normalisation against a panel of normal cerebella and specific alterations such as *MYC*, *MYCN*, *GLI1/2* high level amplifications in MB, for which this pipelines were designed, are not normal germline CNVs. The arm level events that are detected using these methods are large-scale chromosomal changes that would have an obvious phenotype, e.g. Down or Turner syndromes, and therefore do not pose any issue for false-positive discoveries.

The possibility to estimate most DNA copy number aberration status from methylomic data generated by methylation arrays from FFPE preserved tissue offers potential for further investigations using archival cohorts, such as the Newcastle Medulloblastoma archive, for which fresh frozen tissue is not always readily available.

3.6.3. Future work. The bioinformatics methods and the workflow developed and described in this chapter have the potential to be further developed into series of stand-alone applications or software package.

As the microarray technology advances, newer methods for effective analysis of the growing amount of data that is continuously generated are always in demand. Examples of such methods, that were developed recently are CopyNumber450kCancer, that provides a novel functionality of baseline correction in segmentation data obtained from CN calling tools such as ChAMP described above (Marzouka et al., 2016), and RUBIC (Recurrent Unidirectional Break Identification by Clustering), an approach that detects recurrent copy number breaks, rather than recurrently amplified or deleted regions (van Dyk et al., 2016).

Some frequent types of copy number changes, that play a significant role in medulloblastoma, such as ploidy and copy neutral loss-of-heterozygosity (Jones et al., 2012) are impossible to detect using methylation arrays due the design of this platform. This could be addressed by designing additional pipeline that would use whole-genome methylation sequencing (e.g. enzymatic methyl-seq (EM-seq) or bisulphite sequencing) for both genomic and epigenomic analysis.

The workflow is modular and each part of it is potentially scalable. New methods can be added to each module in the same way as the *conumee*-based method was integrated, or deprecated methods can be replaced with current best practice.

EPIC methylation arrays have a greater amount of probes and provide a more dense genome coverage in comparison to 450k arrays. Therefore, they should be more sensitive to the narrow non-high-level CN changes like single copy gain or loss. Integration of the EPIC array analysis will be straight forward. The current version of the Bioconductor package *minfi* has been updated to work with EPIC arrays and the following functions have been added to allow working with all types of Illumina methylation arrays - Table 3.7. The function `read.metharray` (reads idat files into R) has replaced the deprecated function `read.450`, and has been already used in this project. The new functions `combineArrays` (allows to combine data from different platforms) and `convertArray` (casts an array platform into another) would be useful for mixed platform experiments.

The addition of a method that would be able to detect chromothripsis would also be beneficial, however the limited resolution of these arrays may preclude its detection using this platform.

Function	Description	Platforms
read.metharray	Read idat files into R	27k, 450k, EPIC
convertArray	Cast an array platform into another	27k, 450k, EPIC
combineArrays	Combine data from different platforms	27k, 450k, EPIC

TABLE 3.7. Updated minfi functions supporting multiple versions of Illumina HumanMethylation arrays (Fortin et al., 2017).

3.6.4. Summary. In this chapter a methylome profiling technology was employed for the identification of DNA copy number status from both FFPE and fresh-frozen tissue, thus confirming that Illumina Human Methylation arrays are a suitable platform for integrated genomic and epigenomic analysis of FFPE solid tumour samples and can be successfully used as an alternative to genotyping SNP arrays or sequencing for copy number analysis, alongside their intended use for methylomic analysis.

The bioinformatics workflow described in this chapter successfully addressed a need for a multilevel, modular approach that can evolve to accommodate changes in array technology and best practice bioinformatics techniques and provides a platform for prognostication and biomarker identification in the disease, discussed in chapters 4 and 5.

Application of the developed methodology to validation of existing prognostication schemes and determination of the utility of PNET4 stratification scheme for prognostication of patients with unavailable clinical risk factors

4.1. Introduction

The fundamental goal of most statistical analyses is to identify factors important to the outcome, *i.e.* to determine whether the outcome is affected by measurable covariates (Ensor, 2014). For cancer research, covariates can be specific molecular biomarkers, present alongside classical clinical markers such as histopathology or extent of resection; biomarker identification can enable targeted therapy and personalised cancer treatment. Biomarkers, or biological markers, are characteristics that are objectively measured and evaluated as indicators of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic agent (Hunter et al., 2012). Because biomarkers play an important role as prognostic and predictive factors of clinical outcome, it is important to define and understand the common statistical implications of biomarker validation. Biomarker validation is the process of assessing the biomarker and its characteristics, and determining the range of conditions under which the biomarker will give reproducible and accurate data (Wagner, 2002; Lee et al., 2006).

Cancer biomarkers are used to characterise tumours and to explain the heterogeneity between different types of tumours. The heterogeneity is reflected by the wide range of sub-classifications (diagnostic markers) and risk-stratifications (prognostic markers) in many cancer types, as well as by the number of biomarkers able to predict the response of patients to personalised therapies (predictive markers) (Patel et al., 2014).

It is now recognised that medulloblastoma is a group of heterogeneous tumours with variable demographics, transcriptomes, genetics, and clinical outcomes (Northcott et al., 2012a; Kool et al., 2012; Shih et al., 2014).

The WNT and SHH subgroups are characterised by activating aberrations that affect critical regulators of the corresponding signalling pathways. Groups 3 and 4 are characterised by recurrent chromosomal alteration and a low incidence of mutations. These subgroups are more related to each other than to WNT and SHH activated tumours and are listed as non-WNT/non-SHH MB in the revised WHO 2016 classification (Louis et al., 2016).

Discovery and validation of clinically meaningful medulloblastoma features in previous clinical trial cohorts has enabled improvements in the clinical management of the disease. This includes systemic therapeutic approaches that take into account molecular subgroups. There are now ongoing biomarker driven clinical trials that investigate the possibility of reducing intensity of treatment for patients with WNT-activated MB, who consistently show favourable survival rates and evaluate prospects of using SHH pathway inhibitors for treatment SHH-activated disease at diagnosis or refractory or recurrent SHH-activated tumours.

The non-WNT/non-SHH tumours represent a complex challenge, since these tumours account for over two thirds of all MBs, are largely molecularly and clinically heterogeneous, and there is currently a lack of specific targeted treatment strategies for this category of patients.

Current risk stratification is based primarily on clinical variables, with high-risk patients identified as having leptomeningeal metastases at presentation and/or an incomplete resection (Gajjar et al., 2006; Northcott et al., 2012a; Lannering et al., 2012).

Even though in general these features are considered high-risk, in some groups of patients, the presence of these markers does not always result in poor outcomes and in others, absence of high-risk markers does not guarantee favourable outcomes. Often, these risk indicators are not available at the time of the diagnosis for various reasons, which may contribute to decisions on the treatment regimen, and, subsequently, on the overall outcome. Numerous research groups have attempted to identify biomarkers to either support or supplant clinical risk stratification (Goschzik et al., 2018; Northcott et al., 2012a; Remke and Hielscher, 2011a).

In this chapter, two copy-number based, biomarker risk stratification schemes, described in Shih et al. (2014) (referred to as the Shih scheme further in the text) and Goschzik et al. (2018) (referred to as the PNET4 scheme in the text from here) were

validated using the copy number analysis workflow for methylation arrays described in chapter 3. Additionally, the impact of missing/unavailable clinical risk-factor data in newly diagnosed patients on treatment regimen and outcomes and the potential of the PNET4 scheme to overcome this issue is evaluated. Finally, the PNET4 scheme, which was derived from a clinical trial of standard-risk medulloblastomas, was tested on high-risk non-SHH/non-WNT medulloblastoma.

4.1.1. Shih risk stratification scheme . The Shih scheme is based on subgroup-specific biomarkers (*GLI2*, *MYC* amplifications, chromosome 11 loss, chromosome 14 loss, 17p and 17q statuses) that identify very low-risk and very high-risk patients within SHH, Group 3 and Group 4 medulloblastoma subgroups. This scheme was originally developed from CNAs identified using genotyping arrays (Affymetrix SNP6.0) and published in Shih et al. (2014). A risk model based on these subgroup-specific biomarkers is presented on Figure 4.1.

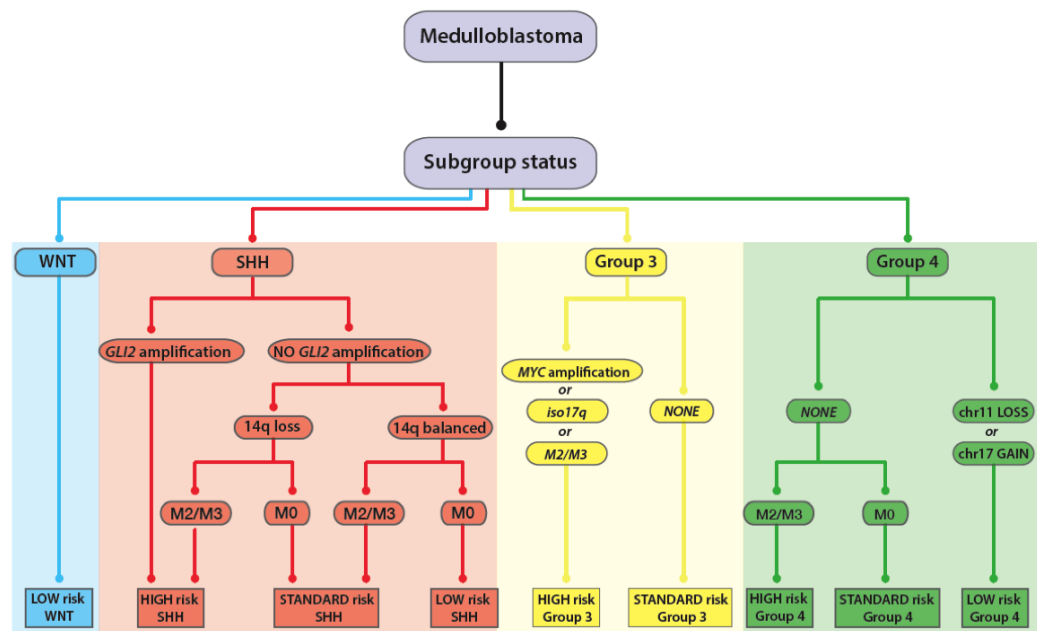


FIGURE 4.1. Shih prognostication scheme: subgroup-specific risk model stratifies medulloblastoma by molecular subgroup, clinical and cytogenetic biomarkers(Shih et al., 2014).

4.1.2. PNET4 risk stratification scheme. The PNET4 scheme (derived from HIT-SIOP-PNET4 clinical trial by Goschzik et al. (2018)) involves the presence of the PNET4 signature - at least two of chromosome 7 gain, chromosome 8 loss and chromosome 11 loss, and which is associated with favourable risk (near 100 percent survival)

within non-WNT/non-SHH standard risk medulloblastomas, further stratifying patients of this risk group into favourable-risk and high-risk subgroups. A risk model defined by these non-WNT/non-SHH specific cytogenetic biomarkers within standard risk disease is presented in Figure 4.2.

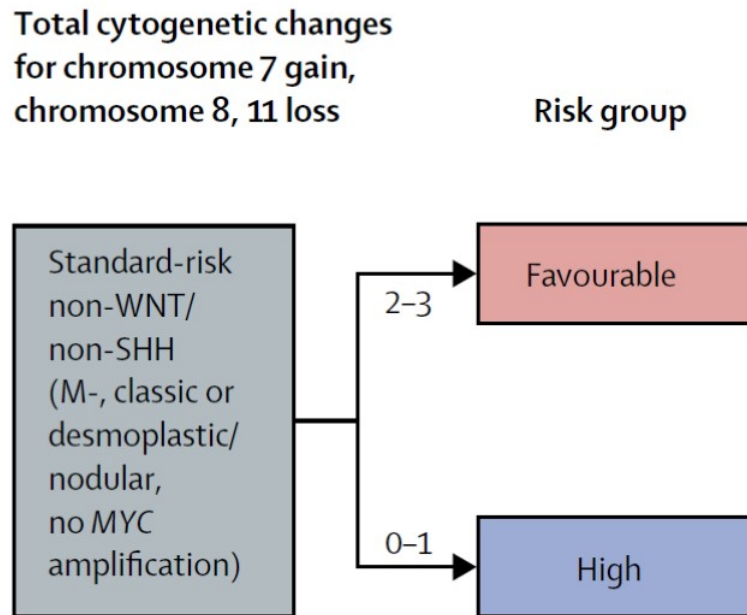


FIGURE 4.2. PNET4 prognostication scheme: non-WNT/non-SHH-specific risk model stratifies SR medulloblastoma by cytogenetic biomarkers chromosome 7 gain, chromosomes 8, 11 loss, into two risk groups: favourable - 2 or 3 biomarkers present, and high - 0 or no biomarkers present (Goschzik et al., 2018).

4.2. Aims

This part of the study aimed to investigate:

- Whether prognostication schemes described in Shih et al. (2014) and Goschzik et al. (2018) can be validated using methodology, developed and described in Chapter 3, and DNA methylation array approach to CNA analysis
- Application of PNET4 scheme described in Goschzik et al. (2018) to high-risk medulloblastoma
- Whether PNET4 scheme has a potential for rapid risk assessment of medulloblastoma patients when clinical risk factors are unavailable at the time of diagnosis and treatment regimen decision

4.3. Materials and methods

4.3.1. Shih scheme biomarkers validation. In order to validate the suitability of the CNA analysis methods developed during this project, the results of a recent cytogenetic prognostication paper (Shih et al., 2014) were validated on an independent disease cohort using the above-mentioned methodology.

4.3.1.1. *Discovery cohort.* The prognostic biomarkers were previously discovered and described in Shih et al. (2014) using a discovery set of 673 medulloblastoma samples with clinical follow up and risk stratified by the scheme represented on Figure 4.1. As WNT-pathway activated medulloblastoma subgroup has consistently shown a favourable risk in previous studies (Shih et al., 2014), this work focused only on validation of biomarkers associated with SHH, Group 3 and Group 4.

A summary of the clinical data for this cohort available from Shih et al. (2014) is presented in Table 4.1. Subgroup information was not available for 162 samples, which, along with 58 WNT subgroup samples, were excluded from analysis.

4.3.1.2. *Validation cohort.* A total of 338 patients (96 SHH, 99 Group 3 and 143 Group 4, initially, though some samples had to be removed later due to inconclusive metastatic status) were analysed using Illumina 450k methylation arrays and genomic copy number changes and focal amplifications were detected using the techniques detailed in Chapter 3. The clinical data of this cohort (referred to as Newcastle Medulloblastoma, or NMB cohort further in text), was demographically comparable to Shih et al's discovery cohort, excepting the absence of adult (aged over 16) patients in the NMB cohort, and is shown in Table 4.1.

Only summary demographic information for the cohort from Shih et al. (2014) was available, therefore no statistical testing for differences between cohorts was performed.

4.3.1.3. *Copy-number profiling.* Copy-number profiling of the NMB cohort was performed using the copy-number analysis workflow described in section 3.3.2 and outlined on Figure 3.1.

Large-scale, *i.e.* arm-level/whole chromosome copy-number profiling for the following copy-number aberrations: chromosome 14q loss, isochromosome 17q, whole chromosome 11 loss, whole chromosome 17 gain, was assessed on the NMB cohort using the large-Scale (arm-level) CNA pipeline as described in section 3.3.7. Whole genome

arm-level status profile was assembled for the validation cohort and from this, individual chromosomal copy number status for chromosomes 11, 14 and 17 was inferred, as described in section 3.3.6.1.

Locus-specific copy-number profiling for *GLI2* and *MYC* oncogene amplifications was performed using the focal, locus-Specific pipeline described in section 3.3.7. Two single gene profiles for *GLI2* and *MYC* genes were assembled.

Finally, a full combined profile containing statuses for chromosomes 11, 14, 17 and genes *GLI2* and *MYC* was assembled, as described in section 3.3.7.

Subgroup	Discovery cohort (Shih et al., 2014)				Validation cohort (NMB)				p-value		
	SHH	Group 3		Group 4		SHH	Group 3		Group 4		
N (%)	158 (30.9)	98 (19.2)		202 (39.5)		96 (28.4)	99 (29.3)		143 (42.3)		
Age											
Median	6.5	5.3		8.5		2.83	4.32		7.95		
Range	0-47	1-17		1-36		0-15	0-16		0-16		
Group	Infant Child Adult	Infant	Child	Adult	Infant	Child	Adult	Infant	Child	Adult	
N	49 72 35	11	85	2	9	180	13	0	6	137	
%	31.4 46.2 22.4	11.2	86.7	2.0	4.5	89.1	6.4	0	4.2	95.8	
Gender	Male Female	Male	Female		Male	Female		Male	Female		
N	85 69	61	37		139	61		73	46		
Ratio (male : female)	1.23 : 1	1.65 : 1		2.27 : 1		1.4 : 1		2.8 : 1		2.1 : 1	
Metastatic status	M0 M1 M2/M3	M0	M1	M2/M3	M0	M1	M2/M3	M0	M1	M2/M3	
N	95 7 18	41	10	24	104	11	59	23	49	11	
%	79.2 5.8 15.0	54.7	13.3	32.0	59.8	6.3	33.9	23.9	49.5	11.1	
Follow-up (years)†											
Median	4.25	5.18		6.25		6.28	5.83		4.6		
Survival (5-year) ‡											
%	63.8	55.1		69.0		54	70.6		82.1		
CI	55.5-73.4	45.1-67.4		61.8-76.9		38.6-75.5		56-89.1		72.5-93	

TABLE 4.1. Patient characteristics of the discovery (Affymetrix SNP6.0, Shih et al. (2014), with WNT subgroup excluded) and validation (Illumina HumanMethylation450k Arrays, Newcastle Medulloblastoma) cohorts. Infant, age <3; Child, 3 ≤ age < 16; Adult, age ≥ 16 years old. M0, no metastasis; M1, presence in cerebrospinal fluid; M2/M3, macroscopic metastasis. CI, 95% confidence interval. NMB, Newcastle Medulloblastoma cohort.

4.3.2. PNET4 scheme biomarkers validation and application to high-risk disease cohort. The original, standard-risk disease stratification scheme developed by Goschzik et al. (2018) using the CNA analysis methods on standard risk cohort, was validated on a standard-risk cohort, and subsequently applied to the high-risk medulloblastoma cohort.

4.3.2.1. *Discovery cohort.* The prognostic biomarkers were previously discovered for Group 3 and Group 4 standard risk disease and described in Goschzik et al. (2018) using a discovery set of 91 medulloblastoma samples from HIT-SIOP PNET4 clinical trial. In this cohort, patients with certain potential high-risk factors, such as *MYCN* amplification and subtotally resected tumour were retained as there was no evidence that these factors affected survival in this particular cohort. Clinical data for the original discovery cohort of 91 standard risk medulloblastoma samples is represented in the second column of Table 4.2.

4.3.2.2. *Validation cohort .* A total of 75 standard risk medulloblastoma samples (of which 16 are Group 3 and 59 are Group 4 samples) were analysed using Illumina 450k methylation arrays and aneuploidy was assessed using the CNA profiling methods developed described in chapter 3. This cohort was demographically matched to the original discovery cohort, excepting the absence of adult patients (aged over 16), and patients with *MYCN* amplifications and sub-totally resected tumours, and is represented in the third column of the Table 4.2.

4.3.2.3. *High risk cohort .* A 100 member non-SHH/non-WNT high-risk cohort was used to test the suitability of the PNET4 prognostication scheme for assessing high-risk disease patients. The entry criteria for this cohort was presence of at least one known high risk factor, namely metastatic disease, large-cell/anaplasia, *MYC/MYCN* amplification and subtotal resection. The clinical data for this cohort is represented in the fourth column of the Table 4.2.

	Non-WNT/non-SHH in HIT-SIOP PNET 4 (n=91)	Non-WNT/non-SHH in SR NMB cohort (n=75)	Non-WNT/non-SHH in HR NMB cohort (n=100)	p-value HIT-SIOP PNET 4 <i>vs</i> SR NMB cohort	p-value HIT-SIOP PNET 4 <i>vs</i> HR NMB cohort
Sex					
Male	61 (67%)	55 (73%)	71 (71%)		
Female	30 (33%)	20 (27%)	29 (29%)	0.4	0.6
Male:female ratio	2:1	2.75:1	2.44:1		
Age at diagnosis (years)*	8.0 (4-20) [6.0-10.0]	7.46 (3.25-15.92) [5.28-9.81]	7.04 (3.47-16.43) [5.32-10.47]		
Histology					
Classic	81 (89%)	68 (91%)	75 (75%)		
Desmoplastic/nodular	10 (11%)	7 (9%)	1 (1%)	0.8	<0.001
Large-cell/anaplastic	0	0	24 (24%)		
Resection					
Gross total resection	81 (92%)	75 (100%)	69 (69%)		
Subtotal resection	7 (8%)	0	31 (31%)	0.02	<0.001
Follow-up (years)	6.7 (5.8-8.2)	3.97 (0.25-28.99)	3.06 (0.18-16.01)		
Molecular subgroup					
Group3	15 (16%)	16 (21%)	33 (33%)		
Group4	76 (84%)	59 (79%)	67 (67%)	0.43	0.01
MYC amplification					
Amplified	0	0	8 (8%)		
Not amplified	91 (100%)	75 (100%)	92 (92%)	1	0.003
MYCN amplification					
Amplified	10 (11%)	0	10 (10%)		
Not amplified	81 (89%)	75 (100%)	90 (90%)	0.002	1
Metastasis					
M0	91 (100%)	75 (100%)	33 (33%)		
M+	0	0	67 (67%)	1	<0.001
Chromosome 17 (total assessed)	91	75	100		
iso17q	56 (62%)	35 (46%)	42 (42%)		
Others	35 (38%)	40 (54%)	58 (58%)	0.06	0.01

TABLE 4.2. Clinical and molecular characteristics of non-WNT/non-SHH HIT-SIOP PNET4 (Goschzik et al., 2018) and NMB cohorts. SR - standard risk, HR - high risk, M0 - no evidence of metastasis, M+ - metastatic disease, iso17q - isochromosome 17q, * Data are median (range)[IQR]; p-values derived using chi-squared statistical test.

4.3.2.4. *Procedures.* In the discovery cohort, due to the material being mostly of low quality and quantity, authors were unable to use conventional approaches (such as DNA Methylation arrays) for subgroup assessment and a mass spectrometry minimal methylation classifier (MS-MIMIC) assay (Schwalbe et al., 2017a) was developed and implemented for molecular subgroup assessment. Amplifications of *MYC* and *MYCN* oncogenes were assessed by interphase FISH and genomic copy number changes were identified using a molecular inversion probe array (Goschzik et al., 2018).

The validation and high risk disease cohort samples were analysed on the Illumina 450k DNA methylation microarray and chromosomal and focal copy number changes were estimated using the CNA analysis workflow developed for this project and described in section 3.3.2 and outlined on the Figure 3.1.

Progression-free survival was defined as the time from diagnosis to the first event - progression or death, or date of the last follow-up. If the patient's follow-up time exceeded 10 years, they were right-censored at 10 years as relapses after 10 years are almost never observed in medulloblastoma and it is assumed that the patients are cured of their disease.

4.3.2.5. *Copy-number profiling.* In both validation and high risk cohorts, large-scale copy-number profiling for the following copy-number aberrations: isochromosome 17q, whole chromosome 8 and 11 loss, whole chromosome 7 gain, was performed using the large-scale (arm-level) CNA pipeline as described in section 3.3.7. Whole genome arm-level status profile was assembled for the validation cohort and then individual chromosome status profile was assembled and included chromosomes 7, 8, 11 and 17 (section 3.3.6.1).

Locus-specific copy-number profiling for oncogenes *MYC* and *MYCN* amplifications was performed using the focal, locus-specific pipeline described in section 3.3.7. Two single gene profiles (section 3.3.6.1) for the *MYC* and *MYCN* oncogenes were assembled.

Finally, a full combined profile containing the CNA status for chromosomes 7, 8, 11, 17 and genes *MYC* and *MYCN* was assembled as described in section 3.3.7.

4.3.3. Assessment of missing information patterns within Newcastle Medulloblastoma (NMB) cohort. In order to evaluate the extent of the missing data problem in clinical and research settings, a total of 267 samples comprising SHH, Group

3 and Group 4 medulloblastomas were assessed for completeness of relevant clinical information such as subgroup, pathology, extent of resection, *MYC/MYCN* amplification, metastatic status and *TP53* mutant/wild-type status. This dataset included so called 'confident' standard risk samples, *i.e.* negative for all above mentioned risk factors and samples assessed as standard risk on available features, *i.e.* where some risk factor information was missing.

4.3.4. Statistical analyses. Statistical analyses were performed using the R statistical language, version R version 3.5.2, and R-Studio program, version 1.1.456. Patient survival was assessed using Kaplan-Meier method and survival associations were tested using log-rank tests (section 2.9.1). Significance of association was assessed using Fisher's exact test. 'P' values were used to assess significance and $p < 0.05$ were considered significant. Cox models were used to test the prognostic values of clinical markers (resection status (subtotal *vs* gross-totally resected disease), metastasis (metastatic *vs* non-metastatic disease), pathology (LCA *vs* non-LCA), *MYC(N)* amplification (amplified *vs* non-amplified), recurrent whole chromosomal or arm-level aberrations (presence *vs* absence)).

4.4. Results

The bioinformatics methods developed in chapter 3 were used to validate two previously published prognostication schemes - Shih scheme (Shih et al., 2014) and PNET4 (Goschzik et al., 2018). Thus the utility of the methods and methylation arrays for CNA analysis was investigated. Subsequently, performance of the PNET4 scheme was assessed within high-risk disease.

4.4.1. Assessment of missing information pattern within NMB cohort.

The patterns of missing information within the initial Standard-Risk cohort are represented in Figure 4.3. Two plots are shown, for the missing values of each variable, and the combinations of missing/non missing values. The bar plot on the left hand side represents the amount of missing values in each clinico-pathological variable: *TP53* mutation, subgroup, resection, *MYC/MYCN* oncogenes amplification, pathology and metastasis information. The aggregation plot on the right hand side shows all combinations of missing and non-missing values in the observations and proportions of combination of the missing values of variables per sample.

For example, 123/267 (46%) of samples that do not have any high-risk features, also have no missing information; 40/267 (15%) of samples are missing all the important information for risk stratification, such as pathology, resection, metastatic status and the rest have some of the information unavailable (Figure 4.3).

4.4.2. Shih biomarkers validation.

4.4.2.1. *Cohort characteristics* . Demographic characteristics (only available as a table from Shih et al. (2014)) of the discovery cohort, CNA data derived from Affymetrix SNP6.0 array (Shih et al., 2014), were compared against demographic characteristics of the independent validation (NMB) cohort, which is part of Newcastle Medulloblastoma Archive, assessed for CNA using Illumina HumanMethylation450k array, in order to be able to account for any discrepancies in survival characteristics of relevant subsets of these cohorts (see Table 4.1). The Shih cohort consists of mostly North American and Canadian samples, while NMB cohort contains predominantly European samples, therefore, an overlap is highly unlikely. No differences in metastatic status ($p=0.24$ for SHH, $p=0.57$ for Group 3 and $p=0.2$ for Group 4) and gender compositions ($p=0.69$

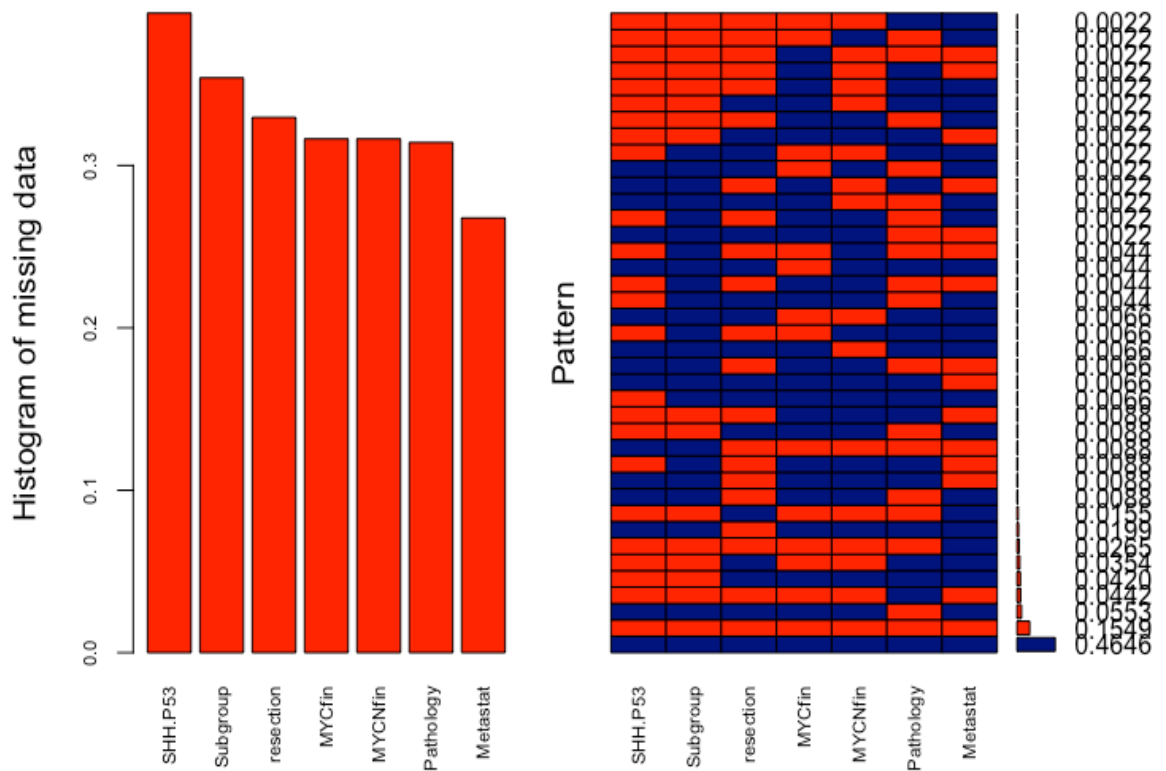


FIGURE 4.3. Missing data patterns within the initial Standard Risk cohort. 1) Histograms show missing information in terms of the proportion of specific clinico-pathological defects: SHH group with *TP53* mutation (labelled as SHH.P53), subgroup, resection, *MYC(N)* oncogene amplification (labelled as MYC(N)fin), pathology and metastasis information, labelled Metastat; 2) On the right hand plot, blue blocks represent complete information, red – missing. The numbers on right hand side represent fraction of samples with particular pattern of complete/missing data.

for SHH, $p=0.1$ for Group 3 and $p=0.8$ for Group 4) of the cohorts were found, however age group compositions of all the molecular subgroups within the cohorts differed significantly ($p<0.001$ for SHH, $p<0.001$ for Group 3 and $p=0.002$ for Group 4).

4.4.2.2. *Stratification of patients with SHH tumours and their survival characteristics.* =A total of 96 patients with SHH-activated tumours from Newcastle Medulloblastoma cohort processed on methylation arrays were stratified into 3 subsets according to the Shih scheme decision tree (Figure 4.4 (A), Shih et al. (2014)) and Kaplan-Meier analysis undertaken (see Figure 4.4 (C)).

The patient subsets carrying high-risk markers such as *GLI2* amplification and chromosome 14q loss were identified, totalling 3 patients (two *GLI2*-amplified and one metastatic with chromosome 14q loss). All the patients in this group died within 5 years of their diagnosis.

The second group was formed of 58 patients carrying standard-risk markers, such as metastatic tumours without chromosome 14q loss and non *GLI2*-amplified, and non-metastatic tumours with 14q loss. For this group, 5 year OS was 57% (Figure 4.4 (C)). 5 year OS in the original study was approximately 45% for corresponding standard-risk subset (Figure 4.4 (B); the exact rate was not available from the original paper (Shih et al., 2014)).

The third group comprised 35 patients that did not carry any high or standard risk features, *i.e.* not *GLI2*-amplified, non-metastatic tumours without 14q loss, with 68% 5 year OS rate (Figure 4.4 (C)). This subset corresponded to low-risk subgroup with the highest survival rate approximately 80% in the original study (Figure 4.4 (B), exact rate was not available in the original paper (Shih et al., 2014)).

The application of Shih stratification scheme to the NMB cohort resulted in three distinct subgroups with individual survival differences that resembled the results in the original paper, however the differences in the survival rates were not significant (p-value=0.06), as opposed to the original study, where p-value=0.001 (4.4 (B)-(C), Shih et al. (2014)). The likely explanation of these discrepancies is differences in demographic and clinical features of the Shih and NMB cohorts (see 4.1, 4.4 (A)-(D)).

4.4.2.3. *Identification of high-risk patients with Group 3 medulloblastoma was confirmed by metastatic status, iso17q and MYC amplification.* Application of the Shih group 3 stratification scheme (decision tree on Figure 4.5 (A)) identified two subgroups with significant survival differences (p=0.004, Figure 4.5 (C)). (decision tree on Figure 4.5 (A)) and Kaplan-Meier analysis undertaken, resulting in two subgroups with distinct survival differences (Figure 4.5 (C)).

The subset of 58 patients that was distinguished by presence of *MYC*-amplification, or isochromosome 17q, or displaying the signs of metastatic disease had significantly worse 5 year OS of 42% when compared with the subset of the patients that did not have any of these features and whose survival was 80%, p=0.004 (Figure 4.5 (C)).

The results were consistent with the results of the original study, where survival rates were about 40% for high-risk subset and 84% for the standard-risk, p<0.001 (Figure 4.5 (B), exact rates were not available in the original paper (Shih et al., 2014)).

4.4.2.4. *Low-risk group of patients with metastatic Group 4 medulloblastoma was identified.* The patient subsets carrying chromosome 11 loss and chromosome 17 gain

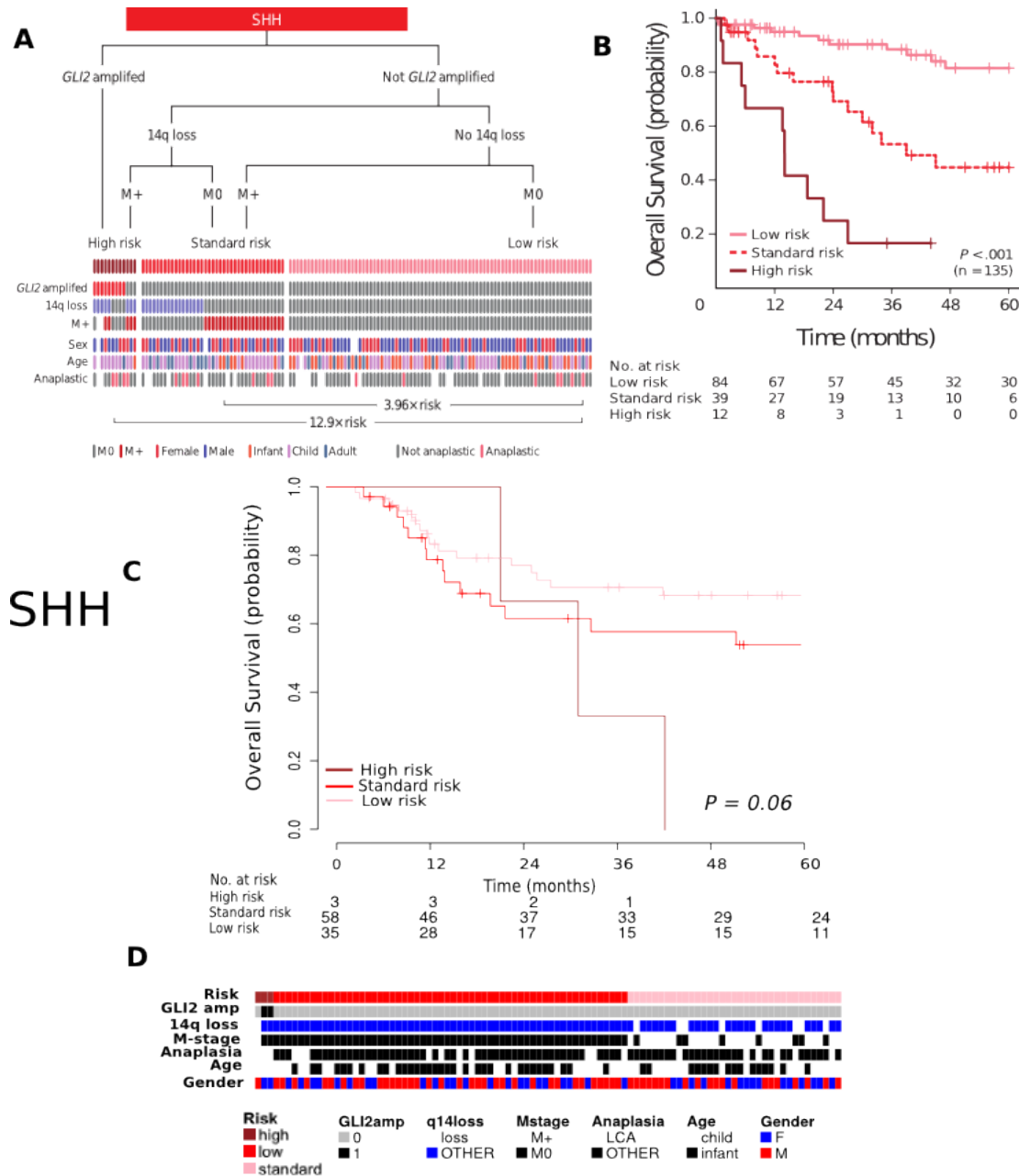


FIGURE 4.4. (A)-(B) Original clinical prognostication scheme developed by Shih et al. using genotyping SNP arrays of patients with SHH medulloblastoma. (A) Risk stratification of SHH medulloblastomas by molecular and clinical prognostic markers. Decision tree, with events plot depicting status of molecular and clinical markers across risk groups below. (B) Overall survival curves for SHH risk groups. (Shih et al., 2014) (C)-(D) Validation of Shih et al. stratification scheme using methylation arrays: (C) overall survival curves for SHH risk groups and events plot of molecular and clinical markers across SHH group (D) for Newcastle medulloblastoma (NMB) cohort

were identified. The cohort then was further stratified according to the Shih scheme (decision tree 4.6 (A)) and Kaplan-Meier analysis undertaken (see Figure 4.5), resulting in three subgroups with distinct survival differences. The result was consistent with the results of the original study, with the patients carrying ch11 loss or ch17 gain showing significantly better 5 year OS survival of 98% (n=18) than metastatic subset

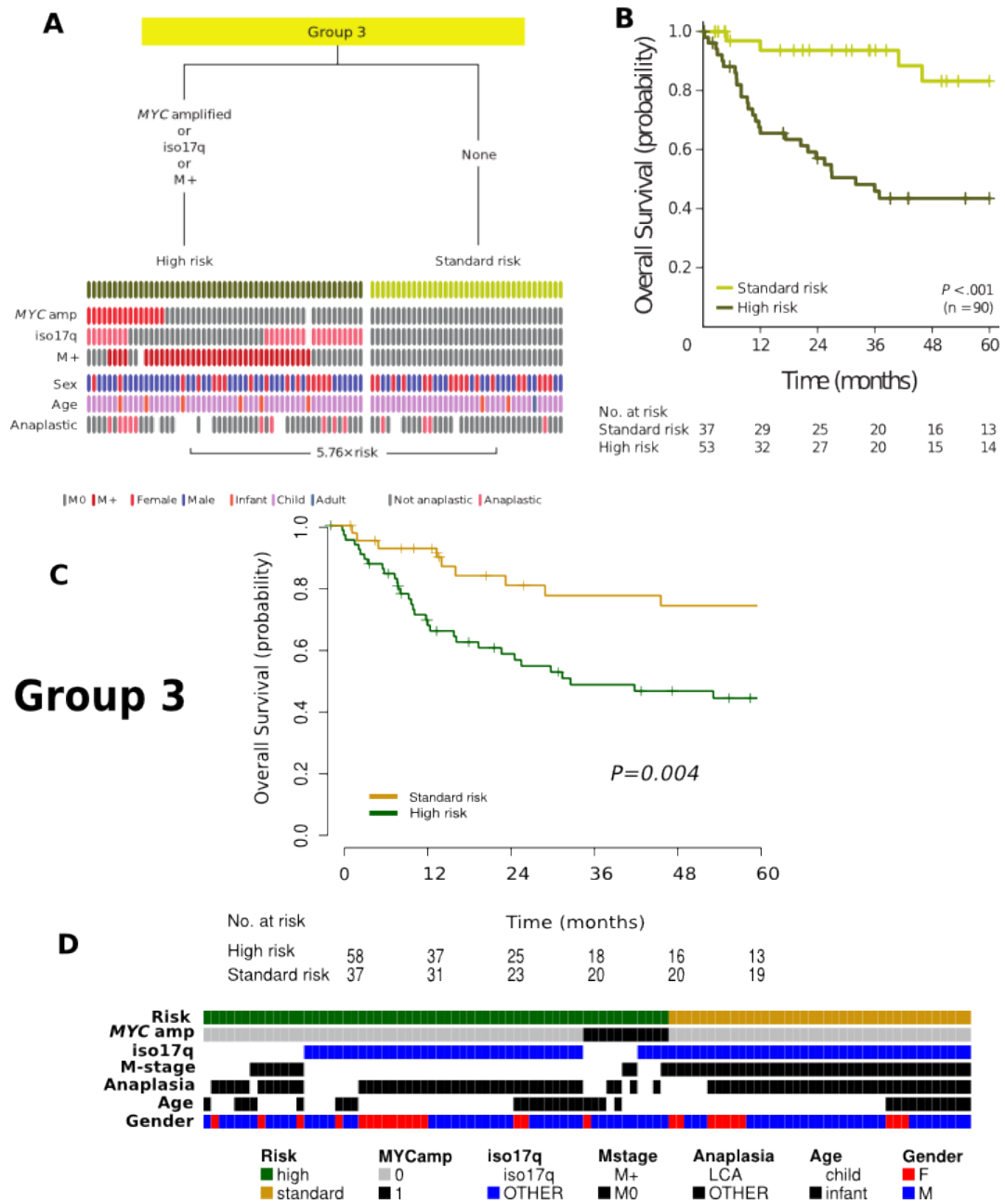


FIGURE 4.5. (A)-(B) Original clinical prognostication scheme developed by Shih et al. using genotyping SNP arrays of patients with Group 3 medulloblastoma. (A) Risk stratification of Group 3 medulloblastomas by molecular and clinical prognostic markers. Decision tree, with events plot depicting status of molecular and clinical markers across risk groups below. (B) Overall survival curves for Group 3 risk groups. (Shih et al., 2014); (C)-(D) Validation of Shih et al. stratification scheme using methylation arrays: (C) overall survival curves for Group 3 risk groups and events plot of molecular and clinical markers across Group 3 group (D) for Newcastle medulloblastoma (NMB) cohort

without either ch11 loss or ch17 gain (77%, n=48), and better than non-metastatic subset without chromosomal copy number changes (50%, n=73), and thus, assigned to lower risk subset (Figure 4.6 (C), $p < 0.001$). In the original study 5 year OS rates were approximately 98% for low-risk, 70% for standard-risk and 22% for high-risk patient

subsets with $p=0.0012$ (Figure 4.6 (B), exact rates were not available in the original paper (Shih et al., 2014)).

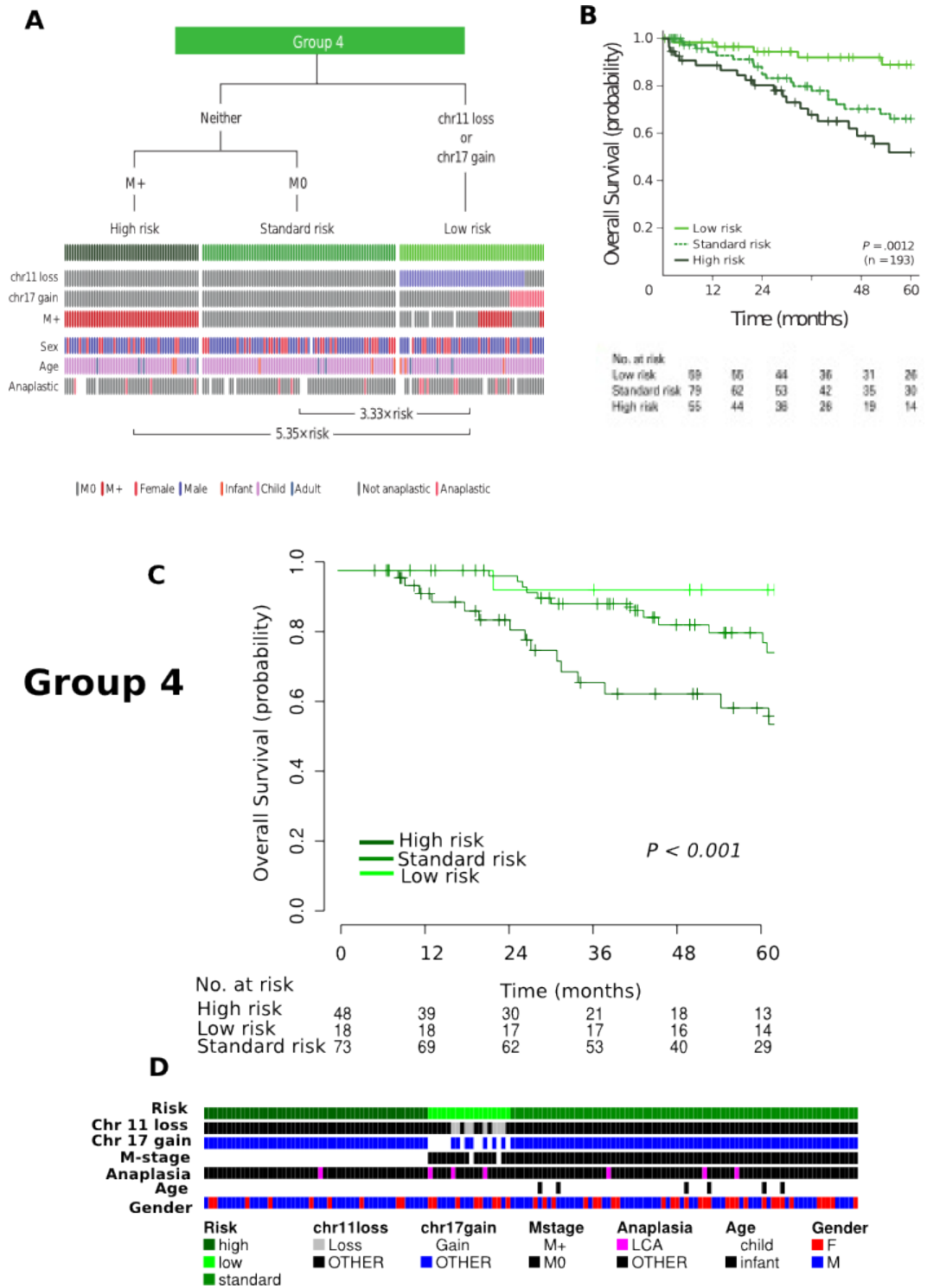


FIGURE 4.6. (A)-(B) Original clinical prognostication scheme developed by Shih et al. using genotyping SNP arrays of patients with Group 4 medulloblastoma. (A) Risk stratification of Group 4 medulloblastomas by molecular and clinical prognostic markers. Decision tree, with events plot depicting status of molecular and clinical markers across risk groups below. (B) Overall survival curves for Group 4 risk groups. (Shih et al., 2014); (C)-(D) Validation of Shih et al. stratification scheme using methylation arrays: (C) overall survival curves for Group 4 risk groups and events plot of molecular and clinical markers across Group 4 group (D) for Newcastle medulloblastoma (NMB) cohort

4.4.3. Validation of biomarkers discovered in PNET4 cohort in Newcastle Medulloblastoma cohort.

4.4.3.1. *Cohort characteristics.* Demographic characteristics of the non-WNT/non-SHH in HIT-SIOP PNET4 (n=91, all standard risk by the design of the trial) cohort, (Goschzik et al., 2018), were compared against demographic characteristics of non-WNT/non-SHH Standard Risk (SR) (n=75) NMB cohort and non-WNT/non-SHH High Risk (HR) NMB cohort (n=100) in order to be able to explain differences in survival characteristics of the relevant cohorts (see Table 4.2).

The tumour samples with PNET identifiers were from the PNET4 clinical trial while the NMB cohort comprised non-trials samples obtained from the Children' Cancer and Leukaemia Group (CCLG) in the UK, plus additional samples obtained from European collaborators.

When comparing non-WNT/non-SHH in HIT-SIOP PNET4 and non-WNT/non-SHH SR NMB cohort no significant differences were identified in sex composition (p=0.4), histology (p=0.8), molecular subgroup composition (p=0.43), *MYC*-amplification (p=1) and metastatic status (p=1). The following characteristics were different: resection (p=0.002 due to no subtotal resected samples in SR NMB cohort), *MYCN*-amplification (p=0.002 due to no *MYCN*-amplified samples in SR NMB cohort) and chromosome 17 status (p=0.003).

When comparing non-WNT/non-SHH in HIT-SIOP PNET4 and non-WNT/non-SHH HR NMB cohort no significant differences were identified in sex composition (p=0.6), *MYCN*-amplification (p=1) and metastatic status (p=1). The following characteristics were different: resection (p<0.001), histology (p<0.001), molecular subgroup composition (p=0.01), *MYC*-amplification (p=0.003) and chromosome 17 status (p=0.007).

4.4.3.2. *Validation of previously reported high-risk clinico-pathological markers .* The relationship between previously reported prognostic factors associated with high-risk disease, such as subtotal resection, LCA pathology, *MYC(N)* amplification, metastasis and presence of isochromosome 17q, and survival was investigated in high-risk NMB cohort consisting of 100 cases positive for one or more prognostic factor being tested. Log-rank tests and univariable Cox proportional hazard models of the prognostic risk factors identified significant relationship between survival and M+ disease

stage (Hazard Ratio (HR) 2.37, 95% CI 1.17-4.7, $p=0.01$; Figure 4.7 B). No significant relationships were identified between subtotal resection (HR 0.73, 95% CI 0.38-1.39, $p=0.3$), LCA pathology (HR 1.5, 95% CI 0.79-2.9, $p=0.2$), *MYC(N)* amplification (HR 1.4, 95% CI 0.51-4.01, $p=0.5$ for *MYC*, HR 0.68, 95%CI 0.25-1.9, $p=0.5$ for *MYCN*) and isochromosome 17q (HR 1.2, 95% CI 0.67-2.18, $p=0.05$) and worse event-free survival (Figure 4.7 A, C-F).

The prognostic potential of existing high-risk markers, defined in disease-wide cohorts, was tested in a high-risk disease cohort, defined by positivity for one or more validated high-risk markers, such as metastatic status, subtotal resection, LCA pathology . While the presence of metastatic disease was prognostic and conferred additional risk, other high-risk markers did not further stratify high-risk disease and were not significantly associated with progression-free survival. Thompson et al. (2016) noted that the prognostic importance of subtotal resection is unclear. Similarly, (Goschzik et al., 2018) have reported that *MYCN*-amplified Group3/4 medulloblastoma did not demonstrate worse survival in the PNET4 clinical trial (Goschzik et al., 2018). The survival of *MYC*-amplified tumours was not significantly different, however this could be due in part to the overall worse survival of this cohort and the small number of *MYC* amplified tumours observed ($n=8$). Finally, LCA disease was not associated with a worse survival; LCA disease has been validated as a poor-risk marker in multiple trials cohorts. Its non-significance in this cohort may be due its co-occurrence with other high-risk disease features. This demonstrates the need for novel biomarkers for high-risk disease, for improved patient stratification.

4.4.3.3. *Validation of PNET4 prognostication scheme: standard-risk disease* . The subset of 75 standard-risk non-WNT/non-SHH medulloblastomas (Table 4.2), carrying whole chromosome 8 and 11 loss and chromosome 7 gain was identified. The patients were then stratified into two subsets. The first subset comprised 16 patients (21%) with at least two of identified copy number aberrations and second subset of 59 patients who presented with one or no PNET4 scheme biomarkers. Five year event-free survival for the first group was 93.75% (95% CI 82.6-100) and for the second group 50.64% (95% CI 38.7-66.34), showing significant association between favourable event-free outcome and presence of at least two of whole chromosomal aberrations defined in the PNET4 prognostication scheme, and poorer survival for those with one or no biomarkers present

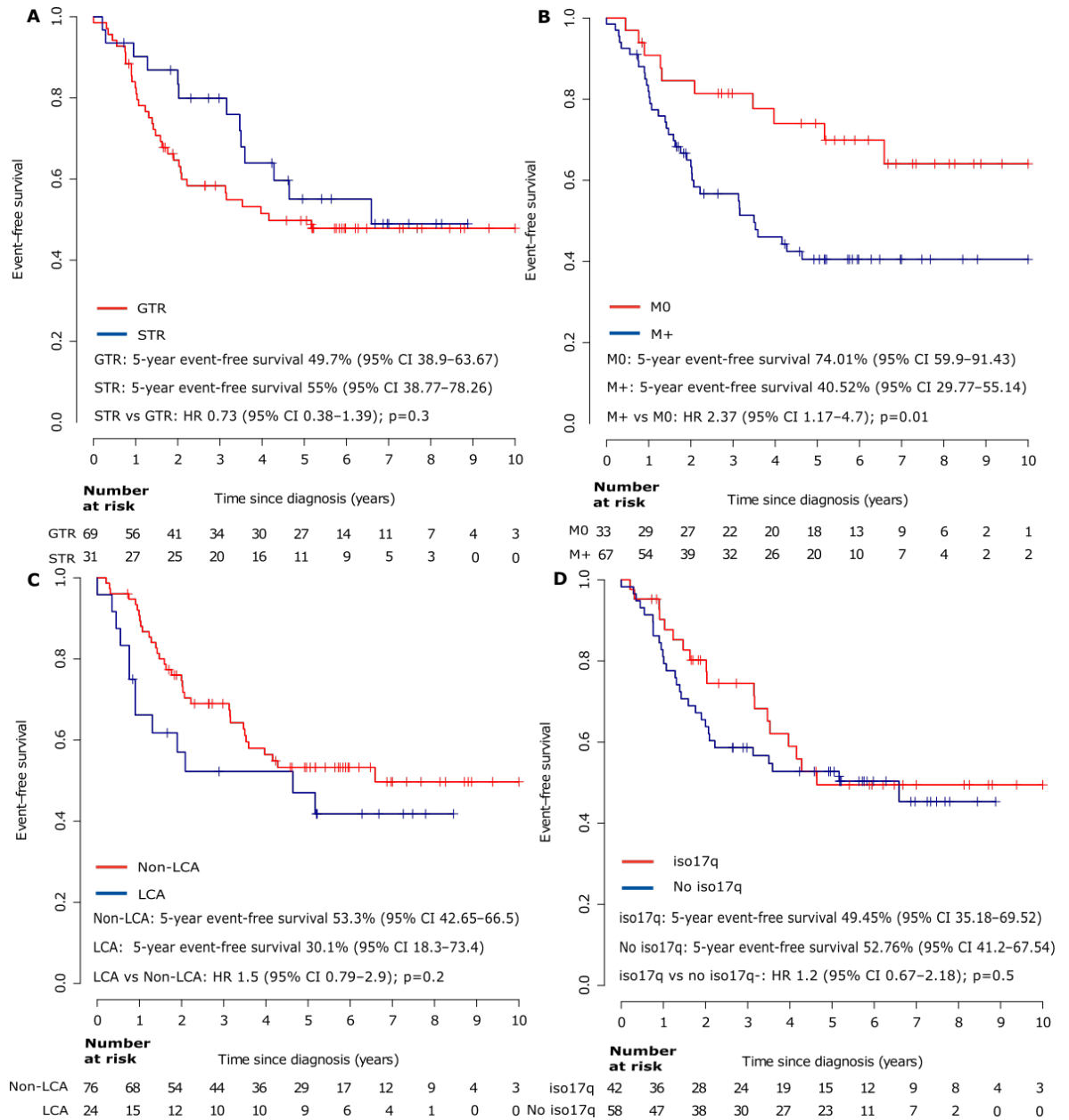


FIGURE 4.7. Continues on next page

(HR=0.1056 (95% CI 0.014-0.78); $p = 0.007$ Figure 4.9). This was consistent with the results of the original study, where 58 samples out 161 (38%) of non-WNT/non-SHH medulloblastomas were classified as favourable risk with 5-year event-free survival was 98.3% (95% CI 94.9-100) and 103 (64%) as high-risk with survival rate 64% (95% CI 52.9-73.2) (Figure 4.8, (Goschzik et al., 2018)).

4.4.3.4. *Application of PNET4 prognostication scheme to high-risk disease*. The PNET4 prognostication scheme was applied to the high-risk medulloblastomas using the methods described in the previous chapter and the favourable-risk subset carrying whole chromosome 8 and 11 loss and chromosome 7 gain was identified. Within the high-risk cohort, Table 4.2, 11 (11%) patients were classified as favourable risk, which

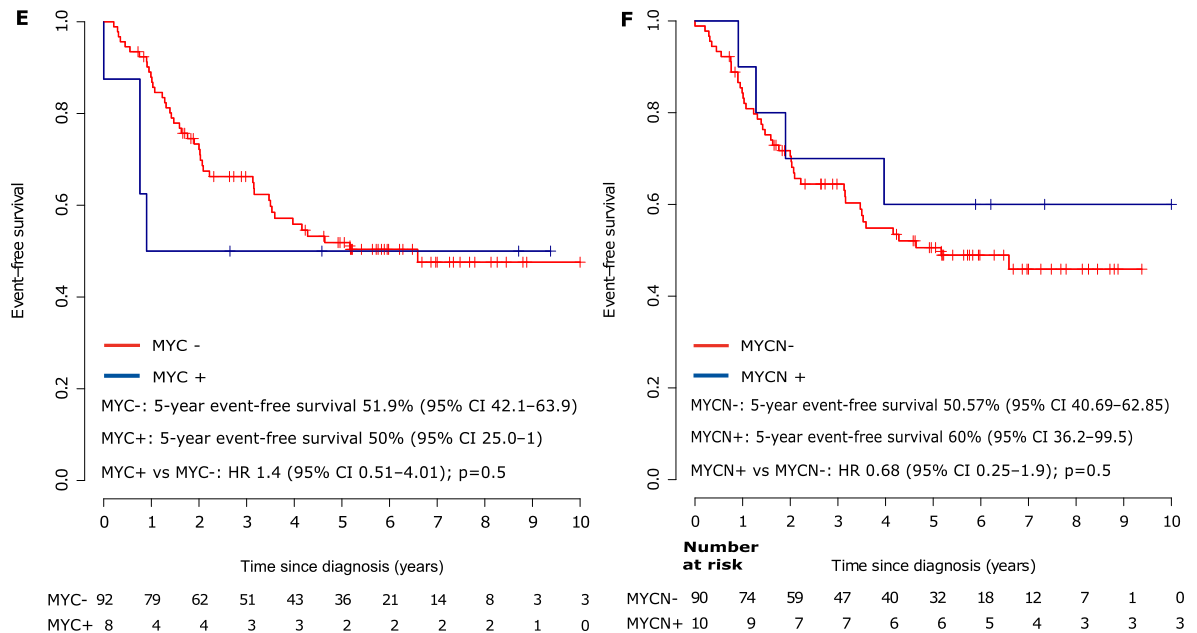


FIGURE 4.7. Event-free survival in the high-risk cohort by clinical and disease-associated molecular features. Patients grouped as: (A) patients with gross total resection *vs* subtotal resection, (B) non-metastatic (M0) *vs* metastatic disease (M+) at presentation, (C) large-cell or anaplasia (LCA) *vs* any other histopathology, (D)-presenting with i17q or not, (E) patients with *MYC*-amplified *vs* non-amplified tumours, and (F) *MYCN*-amplified *vs* non-amplified tumours. HR=hazard ratio.

showed lower incidence of biomarkers signature compared to the standard risk cohort (21%, section 4.4.3.3) and 89 (89%) as high-risk. The presence of at least two whole chromosomal aberrations as defined in the PNET4 prognostication scheme showed significant association with favourable event-free outcome, despite 6 (54%) of them being metastatic: 5 year event free survival was 87.5% (95% CI 67.3-100) *vs* 47% (95% CI 37.7-59.7) in the high-risk group (HR 0.135, 95% CI 0.018-0.98; p=0.02).

The PNET4-scheme derived biomarkers signature-bearing (chromosome 7 gain, and chromosomes 8 and 11 loss) favourable-risk group was observed within methylation subgroups described in recent publications, was found to be significantly associated with subtype VI of Group 3 and Group 4 (Sharma et al., 2019) with p < 0.01, Table 4.3) and Group 4 low risk subtype medulloblastomas (Schwalbe et al., 2017b) with p < 0.01 (Grp4-LR, Table 4.4). The high-risk group was significantly associated with the subtype VIII (Sharma et al., 2019), with p < 0.01 (Table 4.3) and high-risk Group 4 medulloblastomas (Schwalbe et al., 2017b) with p < 0.01 (Grp4-HR, Table 4.4).

4.4.4. Comparison of performance of PNET4 risk stratification scheme: standard-risk and high-risk cohorts. Application of biomarker-driven PNET4 stratification scheme for standard-risk medulloblastoma to high-risk cohort revealed that the

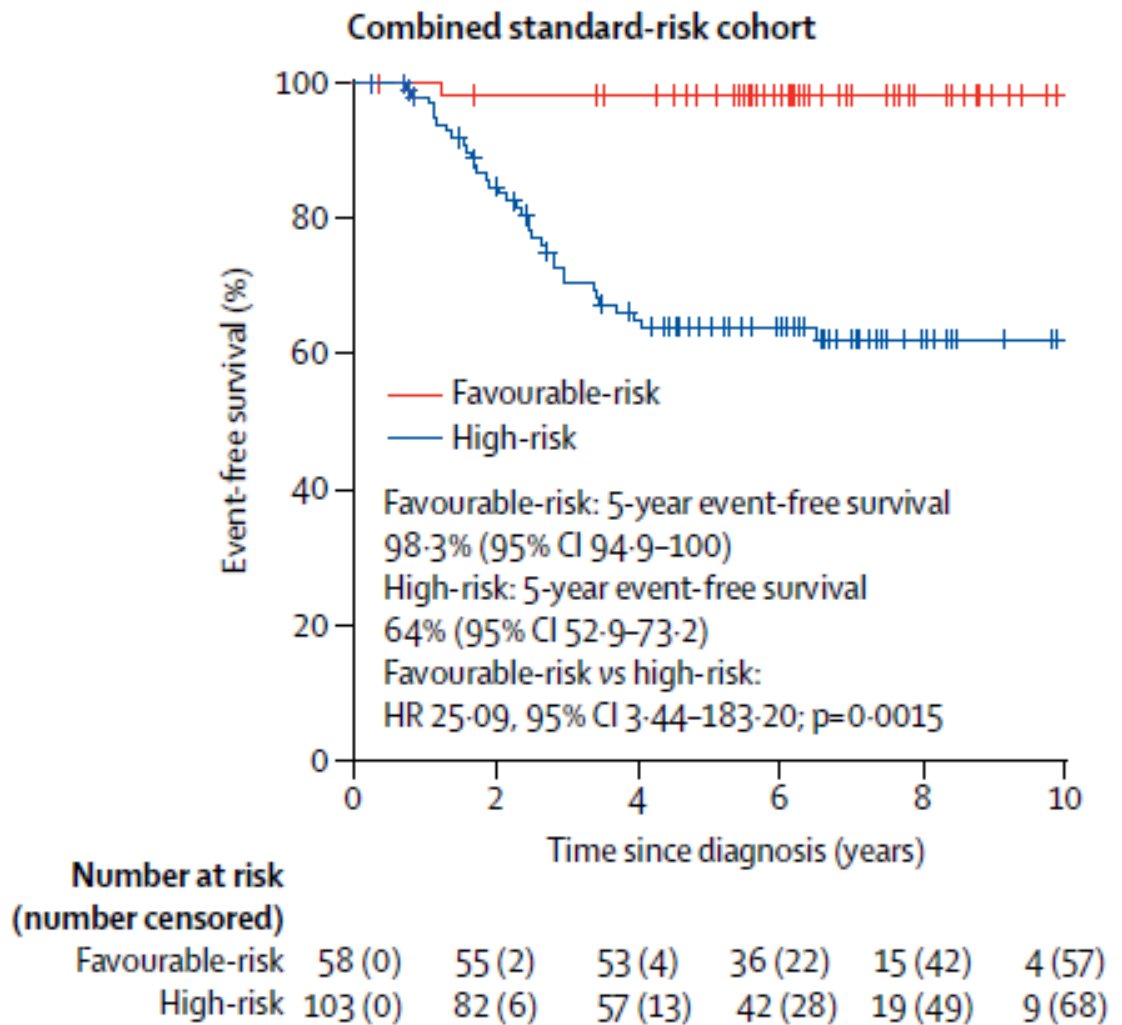


FIGURE 4.8. Original event-free survival curves for non-WNT/non-SHH in combined (HIT-SIOP PNET4 and SR NMB, Table 4.2) standard-risk cohorts by the scheme shown in Figure 4.2. Figure taken from Goschzik et al. (2018).

Subtype	High-risk	Favourable-risk
I	2	0
II	10	1
III	15	1
IV	2	2
V	11	0
VI	5	4
VII	15	1
VIII	22	0
MBNOS*	7**	2***
$p < 0.01$		

TABLE 4.3. Cross-tabulation of PNET4 scheme-defined risk groups and Group 3/4 disease subtypes described by Northcott et al. (2017) in HR NMB cohort. *MBNOS-Medulloblastoma not otherwise specified. **- one of the samples was classified as MBNOS/low confidence subtype V, ***- one of the samples was classified as MBNOS/low confidence subtype VI.

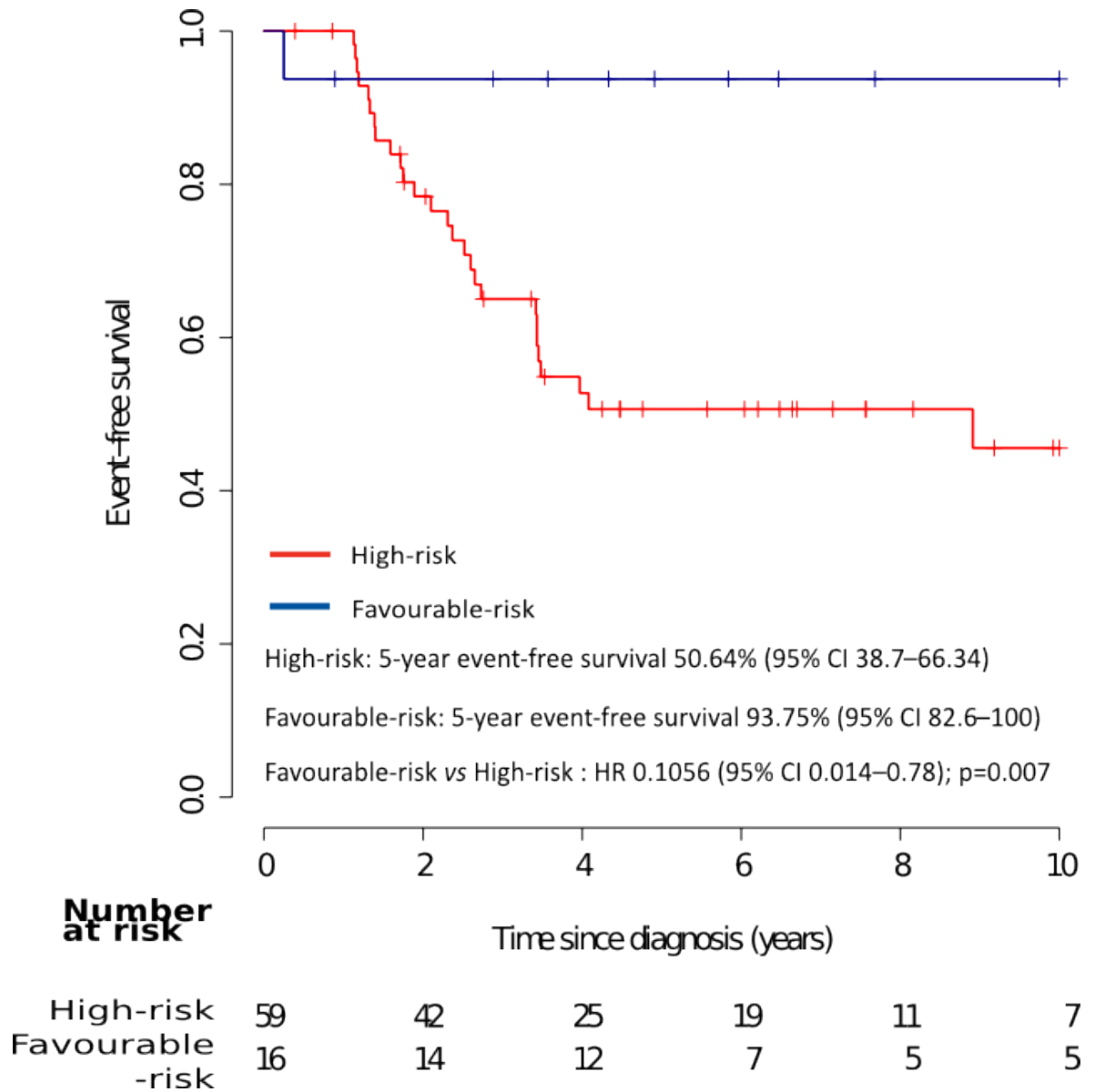
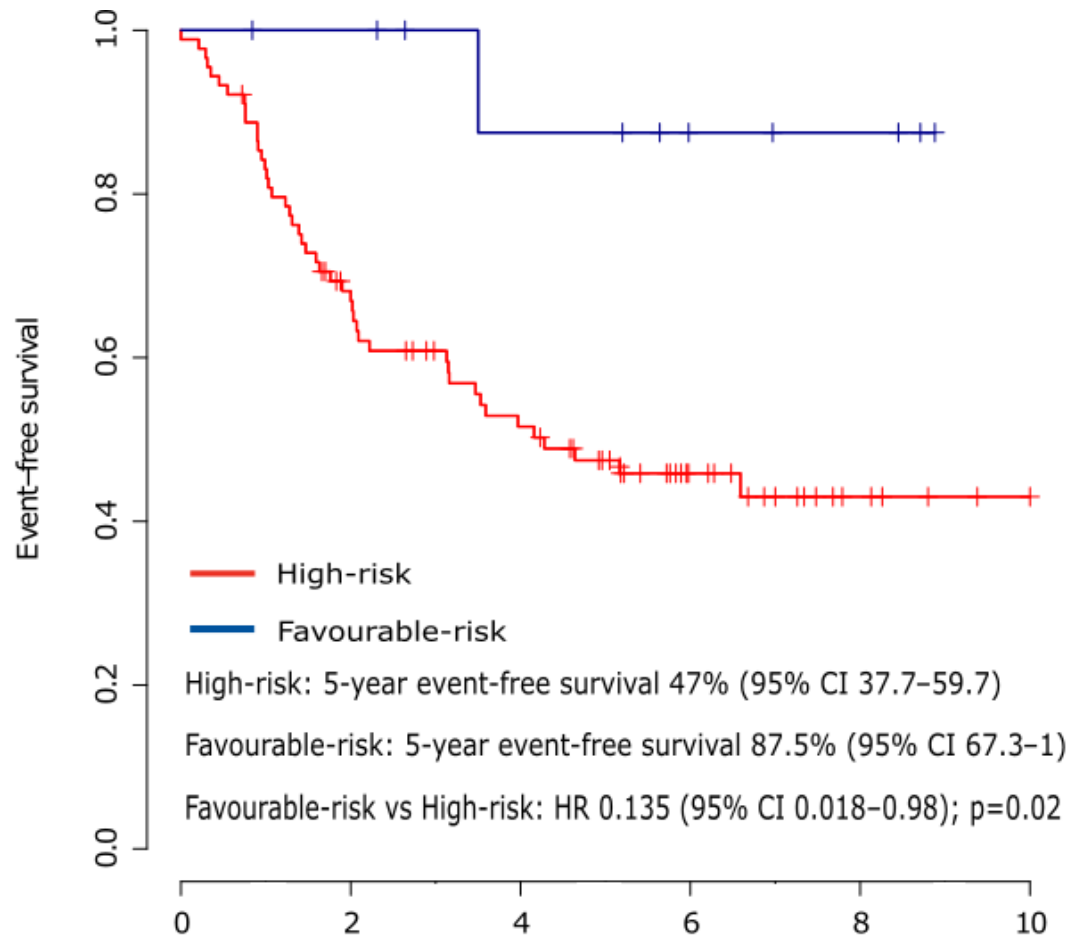


FIGURE 4.9. PNET4 prognostication scheme validation: standard-risk disease. Risk stratification of non-WNT/non-SHH medulloblastomas by the scheme shown in Figure 4.2 applied to standard-risk validation cohort (see 4.3.2.2).

Subtype	High-risk	Favourable-risk
Group 3-HR	26	1
Group 3-LR	5	3
Group 4-HR	31	0
Group 4-LR	27	7
p<0.01		

TABLE 4.4. Cross-tabulation of PNET4 scheme-defined risk groups and Group 3/4 disease subtypes described by Schwalbe et al. (2017b) in HR NMB cohort. HR - high-risk, LR - low-risk.

scheme originally developed for non-infant standard-risk medulloblastoma performed surprisingly well within the high-risk cohort.



Number at risk	Time since diagnosis (years)										
	0	1	2	3	4	5	6	7	8	9	10
High-risk	89	73	56	46	39	31	19	13	7	4	3
Favourable-risk	11	10	10	8	7	7	4	3	3	0	0

FIGURE 4.10. PNET4 prognostication scheme application to high-risk cohort of non-WNT/non-SHH medulloblastoma.

The high-risk cohort comprised of samples that were positive for at least one known high-risk marker, but was otherwise demographically matched to the standard-risk cohort. The proportion of the signature-bearing samples within the high-risk cohort was only 11% which was almost twice less than within the standard-risk cohort (21.3%, see Figures 4.8 and 4.9).

5 years event-free survival in favourable-risk subset of high-risk cohort was less than in standard-risk (87.5% (95% CI 67.3-100) *vs* 93.75% (95% CI 82.6-100) correspondingly Figures 4.8 and 4.9 and Table 4.5).

PNET4 prognostication scheme: 5 years survival comparison		
	Standard Risk disease (SR NMB cohort)	High Risk disease (HR NMB cohort)
Favourable-risk group	93.75% (95% CI 82.6-100)	87.5% (95% CI 67.3-100)
High-risk group	50.64% (95% CI 38.7-66.34)	47% (95% CI 37.7-59.7)
Hazard Ratio (Favourable-risk <i>vs</i> High-risk)	0.105 (95% CI 0.014-0.78; p=0.007)	0.135 (95% CI 0.018-0.98; p=0.02)

TABLE 4.5. PNET4 prognostication scheme comparison table for 5 years survival in non-WNT/non-SHH medulloblastoma: standard risk *vs* high risk disease

4.5. Discussion

An evaluation of suitability and performance of the methodology for CNA analysis described in Chapter 3, demonstrated that this methodology is able to detect both large-scale CN aberrations and focal amplifications. This chapter described the application of the methodology to:

- validation of a prognostication scheme published by Shih et al. (2014) and developed using gold-standard high-density genotyping arrays, thus demonstrating the suitability of the developed methods, as well as confirming utility of methylation arrays, for copy number analysis;
- validation of the previously described PNET4 standard-risk disease stratification scheme and assessment of its utility for stratification of high-risk disease patients and patients whose risk status is unknown.

4.5.1. Copy number analysis of cohorts processed using methylation arrays validates previously reported survival markers and classification schemes.

The stratification of medulloblastoma accordingly to the Shih scheme using methylation arrays in a total of 338 patients from Newcastle non-WNT medulloblastoma cohorts (96 SHH, 95 Group 3 and 139 Group 4) identified survival trends concordant with the ones observed in the original study for all three subgroups.

Group 3 tumours were stratified into high-risk and standard-risk variants, whereas in Group 4 was divided into three subsets: high-risk, standard-risk and low-risk, all with distinct differences in survival ($p = 0.004$ for Group 3 and $p < 0.001$ for Group 4).

The SHH subgroup was stratified into three subsets, first with a dismal 5-year survival rate of 0% corresponding to high-risk, second with 58% corresponding to standard-risk and the last group with the best 5-year OS survival of 68 % corresponding to the low-risk risk group in the original study. However, the survival rates differences were not statistically significant ($p=0.06$).

The subgroup-specific differences in survival identified subsets of each of the 3 subgroup (high-risk, standard-risk and low-risk in SHH, were significant in Group 4 but not significant in SHH with p value of 0.06.

A possible explanation for the discrepancy in p -value, is that the cohort that was available for this study (NMB cohort) was smaller and differed demographically, e.g.

tumour from adult patients were not available (see 4.4.2.1), and contained significantly fewer high-risk tumours (3%) in comparison with the Shih cohort (8%).

Despite the above-mentioned issues, the DNA copy numbers were successfully identified by the methods developed in the chapter 3 using methylation arrays, as the copy number status of most samples was previously evaluated using other techniques and *a priori* known. Therefore, the suitability of both methylation arrays and the bioinformatic methods for CNA analysis was confirmed.

4.5.2. Application of CN analysis workflow and methods to standard-risk Group 3 and Group 4 medulloblastoma cohort validates previously reported risk biomarkers and prognostication scheme. Application of PNET4 prognostication scheme to a subset of NMB cohort processed on Illumina Human Methylation 450k arrays and comprised from Group 3 and Group 4 medulloblastomas using the methodology described in Chapter 3 showed survival rates consistent with the previously reported from non-WNT/non-SHH in HIT-SIOP PNET4 standard-risk cohort (Figures 4.8 and 4.9). This validation showed that the CNA methodology described in the previous chapter performed well thus enabling the next step, application of the PNET4 scheme to high-risk Group 3 and Group 4 combined medulloblastoma cohort to be made.

4.5.3. Application of the PNET4 prognostication scheme to high-risk medulloblastoma cohort shows its prognostic relevance to risk-stratification for disease with known risk factors. Application of PNET4 prognostication scheme to the high-risk subset of Group 3 and Group 4 medulloblastomas identifies favourable-risk and high-risk subsets of patients with very distinct and significant survival differences (87.5% vs 47%, $p=0.02$)

Analysis of the prognostic relevance of disease-associated clinical and molecular high-risk features and event-free survival and found significant association between metastatic disease at presentation and survival (Figure 4.7, B), the rest of the features were not associated with survival in the HR NMB Group 3 and Group 4 cohort (Figure 4.7 A, C-F).

Further analysis of tumour heterogeneity within separate subsets of the HR NMB Group 3 and 4 cohort, such as favourable-risk, high-risk, Group 3 only and Group 4

only subsets was not possible due to small number of samples and possible additional biomarkers remain to be investigated.

4.5.4. Clinical relevance of PNET4 prognostication scheme within high-risk medulloblastoma. Missing clinical information about some of the samples is an important limiting factor of brain tumour research and medulloblastoma studies in particular, as well as being a major obstacle for successful stratification of patients in clinical settings. The later is especially devastating in the developing countries where cutting edge diagnostic techniques are simply not available and everyone diagnosed with medulloblastoma will be treated the same, regardless of potential risk factors (Bouffet, 2019). Some of these unavailable features may be high-risk factors, and while in this particular cohort most of the samples with missing information are recently diagnosed and the treatment regiment decision is very time sensitive. Oftentimes, even if there is a chance that the missing information can be recovered at a later stage, however treatment regiments cannot be amended once commenced. And, in some cases, this will not be possible at all due to a lack of diagnostic facilities, moreover, due to various other factors e.g. insufficient tumour tissue available or inconclusive results.

As the application of the PNET4 risk stratification scheme to the HR cohort showed great prognostication potential, this scheme could benefit those newly diagnosed patients for whom clinical risk information is unavailable. This could be of particular importance in the developing countries, where relatively inexpensive and widespread techniques such as FISH could be used to detect whole chromosome CN status and thus deliver molecularly-directed treatment stratification..

4.5.5. Future work. The investigation of the functionality of PNET4 prognostication scheme within HR medulloblastoma cohort performed in this study has some limitations. Only a small number of HR disease medulloblastoma samples were identified as bearing PNET4 scheme biomarkers. Out of 100 samples analysed only 11 were positive for whole chromosome 7 gain and chromosomes 8 and 11 loss, 6 (54%) of which were metastatic disease and 1 (9%) patient had an event (relapse and subsequent death). The evaluation of influence of known clinical risk factors on survival in this group and identification additional prognostic factors was not possible due to small sample size.

It is still possible that survival rates within HR cohorts would still be smaller due to the chances that some of the known clinical high risk features may still affect the survival in the negative manner. In order to confirm or disprove this hypothesis, further analysis that would involve larger cohorts, is needed. In particular, further investigations would benefit from performing evaluation of event-free survival in the favourable-risk group of HR disease within Group 3 and Group 4 medulloblastoma together, as well as within Group 3 and Group 4 separately to identify potential subgroup-specific risk factors.

If the performance of the PNET4 scheme in HR disease is successfully validated in larger cohorts, PNET4 biomarkers positive samples still only represent 10% of the high-risk patients, and the remaining 90% of HR disease population could be further stratified.

4.5.6. Summary. The aims set in the introduction to this chapter were fulfilled. The previously published Shih stratification scheme (Shih et al., 2014) was verified using methylation arrays as opposed to genotyping SNP arrays used in the original study and thus, it was demonstrated that both methylation arrays and CNA methodology developed during this research project and described in the previous chapter are effective tools for copy number analysis. This chapter also identified that PNET4 prognostication scheme has potential for stratification of patients with unknown or unassessed classic high-risk features. This may be useful in the developing countries where it is still challenging to accurately assess the patients for all the high-risk features. While identified favourable-risk and high-risk subsets of the HR Group 3 and Group 4 cohorts had striking differences in the survival, there is also potential to improve the stratification of the favourable-risk patients by identifying additional prognostic features using larger cohorts.

Standard risk medulloblastoma: methylomic analysis

5.1. Introduction

Non-WNT/non-SHH medulloblastomas, consisting of Group 3 and Group 4 medulloblastomas, represent about 65% of all medulloblastoma cases with vastly heterogeneous molecular and clinical characteristics and survival outcomes. The later are partially associated with known high-risk disease factors such as metastatic status, large cell/anaplastic histology, *MYC* amplification and subtotal resection status (Pizer and Clifford, 2009). However, a significant number of so called standard risk non-WNT/non-SHH medulloblastomas patients, who do not exhibit any of the high-risk features still relapse (Taylor et al., 2012; Goschzik et al., 2018).

In the light of urgent necessity for improved disease to improve patient survival, the development of novel treatments and stratification of current treatments, three independent studies were recently conducted in order to gain better understanding the heterogeneity of these subtypes of medulloblastoma (Schwalbe et al., 2017b; Northcott et al., 2017; Cavalli et al., 2017), using different analytical approaches, cohort composition and parameter choice, which resulted in each of them reporting different number of identified subtypes. A consensus study was conducted by Sharma et al. (2019), which reported 8 subtypes of non-WNT/non-SHH medulloblastomas, thus significantly improving understanding of the nature of the disease. While these findings provide a comprehensive insight into origins and heterogeneity of the non-WNT/non-SHH medulloblastoma, the practical aspects of pre-treatment prognostication in every day clinical settings remain challenging.

This chapter concentrates on risk stratification of non-WNT/non-SHH medulloblastoma and addresses its potential application in routine practice.

5.2. Aims

- Identify significantly differentially methylated loci that correlate with survival in standard risk Group 3 and Group 4 medulloblastoma
- To determine whether methylation markers identified within standard risk disease are prognostic within high risk disease in Group 3 and Group 4 medulloblastoma
- Evaluate integration of identified prognostic methylation markers into an improved risk stratification scheme for non-WNT/non-SHH medulloblastoma
- Validate any findings within independent non-WNT/non-SHH medulloblastoma cohort

5.3. Materials and methods

5.3.1. NMB Standard Risk non-WNT/non-SHH cohort. The standard risk medulloblastoma (negative for conventional high-risk features such as metastatic disease, large-cell/anaplasia, *MYC/MYCN* amplification and subtotal resection) cohort described in the previous chapter (section 4.3.2.2) (n=75) was used as a discovery cohort to identify prognostic methylomic markers shown in the third column of the Tables 4.2 and 5.1.

Matching RNA-seq data for this cohort used for gene expression analysis (see 5.4.11) was provided Dr D. Williamson, RNA prepared by Ms Amanda Smith and sequencing was outsourced to Eurofins Scientific.

5.3.2. NMB High Risk non-WNT/non-SHH cohort and available clinico-pathological correlates. A high risk cohort (n=100) with entry criteria being the presence of at least one known high risk clinico-pathological correlate (such as metastatic disease, large-cell/anaplasia, *MYC/MYCN* amplification and subtotal resection) was used as a test cohort. The cohort was previously used and available clinico-pathological covariates were validated in the chapter 4 and described in section 4.3.2.3. Clinical data for this cohort is represented in the fourth column of Table 4.2 and 5.1.

5.3.3. Age-matched, independent non-WNT/non-SHH cohort with limited clinical risk information. An aged-matched cohort of 244 non-WNT/non-SHH samples with only histopathology and metastatic status available previously published in Cavalli et al. (2017) (referred as Cavalli cohort further in the text) was used to verify the prognostic potential of the identified methylomic biomarkers in uncertain conditions, that is, when some clinico-pathological risk markers are unavailable and therefore even the samples that are standard-risk on available features can still, in fact, potentially be high-risk. Clinical and demographic data for this cohort is represented in the first column of the Table 5.1.

5.3.3.1. *Validation of known clinico-pathological covariates.* The available clinico-pathological biomarkers, such as presence of isochromosome 17q, LCA pathology and metastatic status were assessed using log-rank test and univariable Cox models in order to validate their prognostic value in the Cavalli cohort.

	Non-WNT/non-SHH in Cavalli et al. dataset (n=244)	Non-WNT/non-SHH in SR NMB cohort (n=75)	Non-WNT/non-SHH in HR NMB cohort (n=100)	p-value Cavalli et al. dataset vs SR NMB cohort	p-value Cavalli et al. dataset vs HR NMB cohort
Sex					
Male	165 (68%)	55 (73%)	71 (71%)		
Female	79 (32%)	20 (27%)	29 (29%)	0.4	0.6
Male:female ratio	2:1	2.75:1	2.44:1		
Age at diagnosis (years)*	8.0 (3-16) [5.0-10.0]	7.46 (3.25-15.92) [5.28-9.81]	7.04 (3.47-16.43) [5.32-10.47]	0.7	0.55
Histology					
Classic	188 (77%)	68 (91%)	75 (75%)		
Desmoplastic/nodular	21 (9%)	7 (9%)	1 (1%)	0.001	0.003
Large-cell/anaplastic	35 (14%)	0	24 (24%)		
Resection					
Gross total resection	N/A	75 (100%)	69 (69%)	N/A	N/A
Subtotal resection	N/A	0	31 (31%)		
Follow-up (years)	4.4 (0.08-10)	3.97 (0.25-28.99)	3.06 (0.18-16.01)		
Molecular subgroup					
Group3	64 (26%)	16 (21%)	33 (33%)	0.44	0.2
Group4	180 (74%)	59 (79%)	67 (67%)		
MYC amplification					
Amplified	N/A	0	8 (8%)	N/A	N/A
Not amplified	N/A	75 (100%)	92 (92%)		
MYCN amplification					
Amplified	N/A	0	10 (10%)	N/A	N/A
Not amplified	N/A	75 (100%)	90 (90%)		
Metastasis					
M0	150 (61%)	75 (100%)	33 (33%)	0.001	0.001
M+	94 (39%)	0	67 (67%)		
Chromosome 17 (total assessed)	244	75	100		
iso17q	122 (50%)	35 (46%)	42 (42%)	0.7	0.2
Other	122 (50%)	40 (54%)	58 (58%)		

TABLE 5.1. Comparison of clinical and molecular characteristics of non-WNT/non-SHH (Cavalli et al., 2017) and NMB cohorts. SR - standard risk, HR - high risk, M0 - no evidence of metastasis, M+ - metastatic disease, N/A - unavailable data. The p-value shows results from testing for differences in correlate representation using Fisher's Exact test (gender, metastatic stage) and t-test (age). * Data are median (range)[IQR]

5.3.4. Identification of potentially prognostic methylation markers. In order to identify potentially prognostic methylation probes, all available probes underwent a two-step filtering process. Firstly, potential prognostic methylated probes were filtered, to ensure that only potentially useful markers were tested, using SD as described in section 2.13. Secondly, for the remaining 2000 probes Bimodality Index was calculated and the 402 most bimodal probes were selected as described in section 2.14.

5.3.5. GLM-net to identify significantly differentially methylated CpG sites within standard risk non-WNT/non-SHH medulloblastoma. The GLM-net algorithm (section 2.15) was used to identify prognostic methylation markers on a base Cox model consisting of Group 3 and Group 4 standard risk disease, using the same cohort of non-WNT/non-SHH standard risk medulloblastoma samples as in the previous chapter. This algorithm uses elastic-net regularisation to fit a Cox model. Cross-validation was performed using k-fold CV method with number of fold k and parameters of alpha 1 (lasso), 0 (ridge) and 0.5 (elastic-net).

The model was fitted using following parameters of the alpha: 1 (lasso), 0 (ridge) and 0.5 (elastic-net) and cross-validated in each case using 5 fold cross-validation for the Cox model. The value of lambda that gives minimum mean cross-validated error was calculated, the active coefficients corresponding to potential prognostic covariates were identified. subsequently, univariable and multivariable Cox models were built in order to identify significant covariates among the active coefficients of the model. The suitability of the identified markers was assessed using C-index (section 2.16).

5.3.6. Additional prognostic covariates and definition of improved standard risk non-WNT/non-SHH medulloblastoma stratification scheme. In order to investigate the potential utility of identified prognostic probes, univariable and multivariable Cox proportional hazard models (section 2.9.2) were constructed and C-indices (section 2.16) with 95% confidence intervals were calculated.

Univariate Cox proportional hazard models were build for using each identified prognostic covariate and a multivariate model was constructed using both identified prognostic probes.

The PNET4 signature (section 4.1.2) was introduced as an additional prognostic covariate in order to identify whether a combination of the PNET4 risk stratification

scheme and novel methylation markers offer improved prognostication for the standard-risk non-WNT/non-SHH medulloblastoma patients.

A multivariate Cox proportional hazard model was constructed using the identified prognostic loci and PNET4 signature, C-indices calculated and Kaplan-Meier curves (section 2.5) were plotted.

5.3.7. Investigation of the prognostic potential of the novel risk stratification methylation markers in high risk non-WNT/non-SHH medulloblastoma. Methylation markers identified using standard risk non-WNT/non-SHH medulloblastoma cohort were tested in high risk non-WNT/non-SHH cohort previously used to test the PNET4 prognostication scheme in Chapter 4.

Univariate Cox proportional hazard models were constructed for each methylation marker. Cox proportional hazard model was build incorporating both methylation markers and followed by model that involved both methylation markers combined with PNET4 signature (section 4.1.2).

5.3.8. Additional prognostic covariates in high risk non-WNT/non-SHH medulloblastoma. The high risk non-WNT/non-SHH cohort differed from its standard risk counterpart by containing samples with one or more clinical high risk features described in the previous chapter (see 4.1.2). The behaviour of previously-reported clinical risk markers was investigated in the NMB high-risk non-WNT/non-SHH cohort and reported in section 4.4.3.2. The metastatic status is one of the high-risk disease feature that was consistently found prognostic previously (Shih et al., 2014). Therefore, negative metastatic status (non-metastatic disease, denominated M- or M0) was chosen to be tested as an additional prognostic covariate. Additional four Cox proportional hazard models that combined each model described in the previous section with M0 metastatic status as a covariate were constructed and corresponding Kaplan-Meier curves plotted.

5.3.9. Validation of the identified methylation markers in an independent cohort with only some clinical risk factors available. Identified methylation markers were subsequently tested in the independent external cohort with the clinical risk information limited to LCA pathology, metastatic status and chromosome 17 status.

5.4. Results

Prognostic differentially methylated loci correlating with improved survival in standard risk non-WNT/non-SHH medulloblastoma were identified and investigated alongside of existing low-risk biomarkers in standard risk cohort. Subsequently, potential prognostic significance of these methylation markers was investigated in a high-risk non-WNT/non-SHH cohort and finally, the identified markers were tested in an independent external mixed-risk non-WNT/non-SHH medulloblastoma cohort.

5.4.1. Validation of the previously reported clinico-pathological correlates and cytogenetic markers in the high-risk non-WNT/non-SHH medulloblastoma cohort. Prognostic significance of previously reported risk markers such as metastatic and resection status, large cell/anaplastic pathology, *MYC/MYCN* amplification and isochromosome 17q, was investigated in the high-risk non-WNT/non-SHH medulloblastoma cohort and reported in the previous chapter.

5.4.2. Validation of the previously reported clinico-pathological correlates and cytogenetic markers in the age-matched external cohort with uncertain risk status . The behaviour of the available previously reported prognostic markers in the 244 member age-matched mixed-risk non-WNT/non-SHH medulloblastoma cohort was investigated. Large cell/anaplastic histology and isochromosome 17q were found significantly associated with adverse survival ($p < 0.001$ and $p = 0.02$ correspondently, Figure 5.1 C and D), metastatic status however was not significantly associated with the survival ($p=0.2$, Figure 5.1, B). PNET4 scheme previously described in the previous was able to distinguish high and favourable-risk cases in this cohort ($p = 0.007$, Figure 5.1 A).

5.4.3. Identification of significant methylation markers in the standard-risk non-WNT/non-SHH medulloblastoma cohort. The GLM-net with Cox regression algorithm was used to identify significant methylation markers in standard risk non-WNT/non-SHH medulloblastoma.

When using the lasso penalty, all the 402 coefficients of the model were found to be zero and when using ridge penalty all the coefficients appeared to be "active" i.e. non-zero. When using elastic-net penalty with parameter of alpha set to 0.5, 11 out 402

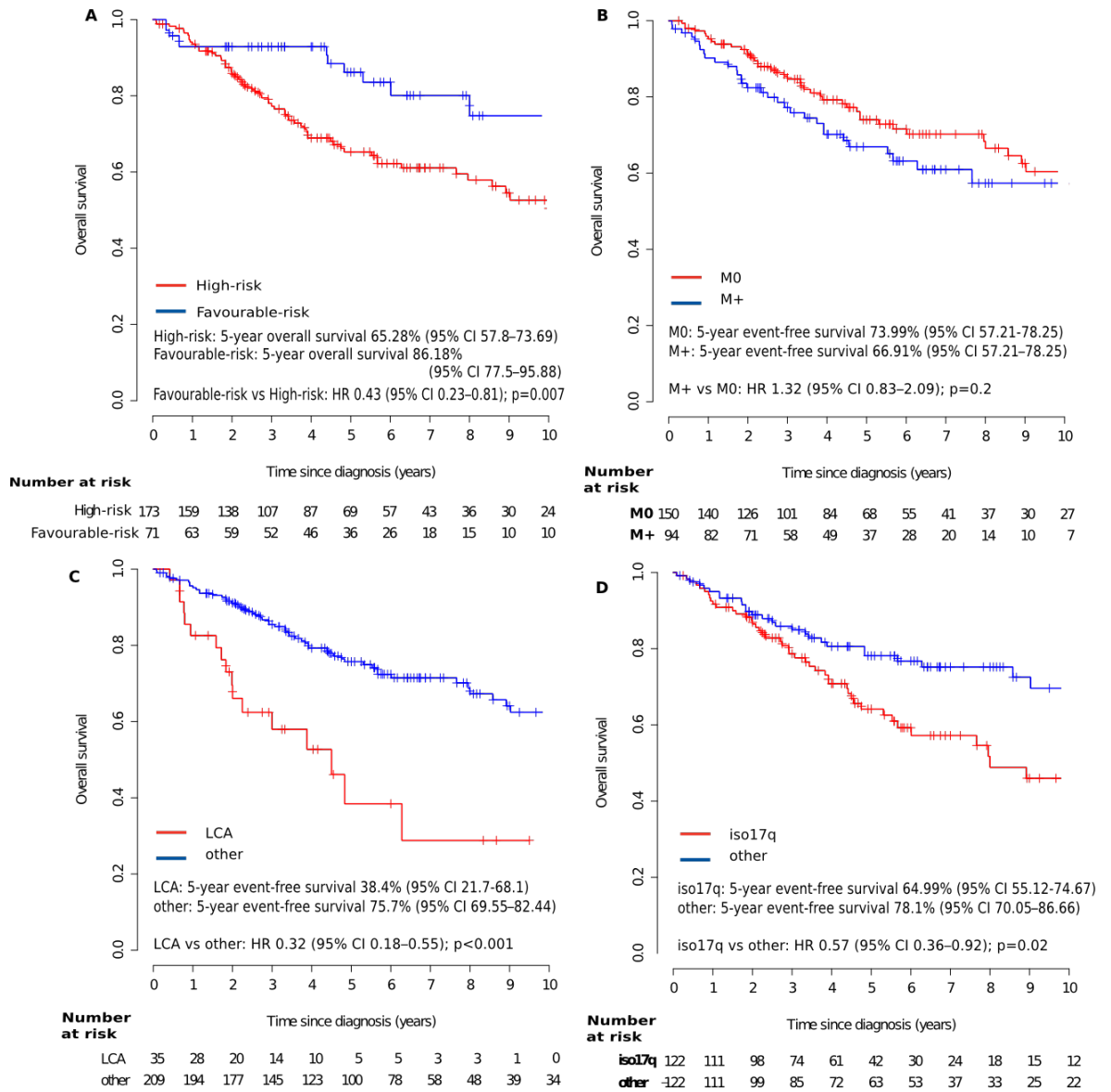


FIGURE 5.1. Kaplan-Meier plots of PNET4 signature (A), metastatic status (B), Large Cell/Anaplastic pathology (C) and presence of isochromosome 17q (D) in Cavalli dataset (Cavalli et al., 2017); .

coefficients were found to be "active", meaning the model proposed by the algorithm consisted from eleven potentially prognostic probes.

The initial Cox model was build using all eleven probes, however only three of those probes ("cg26107890" mapping to the gene *SLC12A8*, "cg08879470" mapping to the gene *MYO7A* and "cg00982958" mapping to the gene *TRIM72*) proved to be significant with the p-values of 0.004, 0.0025 and 0.004 respectively.

5.4.4. Assessment of prognostic potential of identified methylation probes and selection of the cut-off value for classification of methylated and unmethylated loci. In order to assess the suitability of the identified probes for further

Gene	Probe	Univariate			Multivariate		
		Hazard Ratio	p-value	C-index (concordance)	Hazard Ratio	p-value	C-index (concordance)
<i>SLC12A8</i>	cg26107890	0.1899, 95% CI=[0.036-0.985]	0.0479	0.573 (se = 0.057, 95% CI=[0.4612-0.6847])	0.03254	0.00153	0.737 (se = 0.05 95% CI=[0.639-0.835])
<i>MYO7A</i>	cg08879470	0.2256, 95% CI=[0.07704-0.6608]	0.00661	0.621 (se = 0.057, 95% CI=[0.50928-0.732])	0.1724	0.00928	0.737 (se = 0.05 95% CI=[0.639-0.835])
<i>TRIM72</i>	cg00982958	0.1911, 95% CI=[0.06941,0.5264]	0.00137	0.655 (se= 0.057, 95% CI=[0.54328- 0.76672])	0.16474	0.00222	0.737 (se = 0.05, 95% CI=[0.639-0.835])

TABLE 5.2. Assessment of prognostic methylation markers as continuous variables in univariate and multivariate analysis of standard risk non-WNT/non-SHH medulloblastoma cohort. Hazard ratios with 95% confidence interval, p-values and concordance indices with standard error and 95% confidence interval are shown for each methylation covariate.

analysis, univariate and multivariate Cox models were build and examined using concordance index (Table 5.2). All three potentially prognostic probes were significant ($p\text{-value} < 0.05$) in both univariate and multivariate analyses. However, for the probe corresponding to the *SLC12A8* gene, even though its concordance index exceeded 0.5 (C-Index=0.573), the 95% confidence interval included 0.5, indicating marginal significance, and therefore, this probe was removed from further analyses.

It is possible that in a routine clinical setting, prognostication involving methylated loci will be dichotomised for ease of use, rather than measured as a continuous variable.

Intensity values derived from Illumina Human Methylation 450k arrays allow to distinguish between methylated ($\beta \in [0.8 : 1]$), hemi-methylated ($\beta \in (0.2 : 0.8)$) and unmethylated ($\beta \in [0 : 0.2]$) status of loci (Bibikova et al., 2011). However, since samples with hemi-methylated prognostic loci showed survival similar to the ones with methylated loci in *MYO7A* (66.7% vs 66.1% respectively) and similar to unmethylated in *TRIM72* (25% vs 38%), for the simplicity of the model, it was decided that loci with $\beta \in [0 : 0.5]$ were unmethylated and loci with $\beta \in [0.5 : 1]$ were considered methylated.

Univariate Cox proportional hazard models were constructed at the cut-off $\beta = 0.5$ and the probes showed similar behaviour to the univariate model with the continuous methylation data: the *SLC12A8* gene appeared not to be significant and *MYO7A* and *TRIM72* proved to be significant with the p-values of 0.01 and 0.001 correspondingly (Table 5.3).

All p-values were adjusted for multiple testing.

Gene	Probe	Continuous			Cutoff		
		Hazard Ratio	p-value	C-index (concordance)	Hazard Ratio	p-value	C-index (concordance)
SLC12A8	cg26107890	0.1899, 95% CI=[0.036-0.985]	0.047	0.573 (se = 0.057, 95% CI=[0.4612-0.6847])	2.7, 95% CI=[0.82-9.15]	0.08	0.567 (se = 0.038 95% CI=[0.49-0.64])
MYO7A	cg08879470	0.2256, 95% CI=[0.07704-0.6608]	0.006	0.621 (se = 0.057, 95% CI=[0.50928-0.732])	2.87, 95% CI=[1.2-6.8]	0.01	0.596, (se = 0.044 95% CI=[0.51-0.682])
TRIM72	cg00982958	0.1911, 95% CI=[0.06941,0.5264]	0.001	0.655 (se= 0.057, 95% CI=[0.54328- 0.76672])	4.6, 95% CI=[2,1-9.9]	0.001	0.687 (se= 0.043, 95% CI=[0.6- 0.77])

TABLE 5.3. Assessment of prognostic methylation markers as binary variables at 0.5 cut-off in univariate analysis of standard risk non-WNT/non-SHH medulloblastoma cohort. For comparison, univariate Cox proportional hazard models are shown for the methylation scores as a continuous variable. Hazard ratios with 95% confidence interval, p-values are shown for each methylation covariate .

5.4.5. Methylation pattern of the selected probes. The distributions of β -values for the identified probes *MYO7A* and *TRIM72* were assessed by plotting combined histogram and density plots for both loci across non-WNT/non-SHH disease (Figure 5.2) and boxplots showing distribution of β -values for Group 3 and Group 4 subgroups of non-WNT/non-SHH medulloblastoma separately (Figure 5.3). Both probes show clear bi-modal distribution pattern with a majority of samples being methylated in standard risk non-WNT/non-SHH disease.

Distribution by subgroups indicates that Group 3 shows mixed methylation profile, while Group 4 shows predominantly methylated profile for both loci.

5.4.6. Prognostic methylation biomarkers identified lower-risk patients within standard risk non-WNT/non-SHH medulloblastoma . The methylation status of the identified probes in the discovery cohort, consisting of 75 standard risk non-WNT/non-SHH medulloblastomas, was assigned using the β -value cut-off (see 5.4.4) to either the methylated ($\beta \geq 0.5$) or unmethylated ($\beta < 0.5$) subset.

For both identified loci, methylated status was associated with favourable event-free survival. Cox proportional hazard models are shown in table 5.3 and the Kaplan-Meier plots build for these models are presented in Figure 5.4.

For the subset with a methylated status for the *MYO7A* gene, five year event-free survival was 65% (95% CI 54-79) vs 42 % (95% CI 22-82) in the unmethylated group, p=0.01.

The subset with a methylated *TRIM72* locus had a 75.8% (95% CI 64-89) 5-year event-free survival rate, as opposed to 26.5% (95% CI 12-57) in the unmethylated subset, p<0.001.

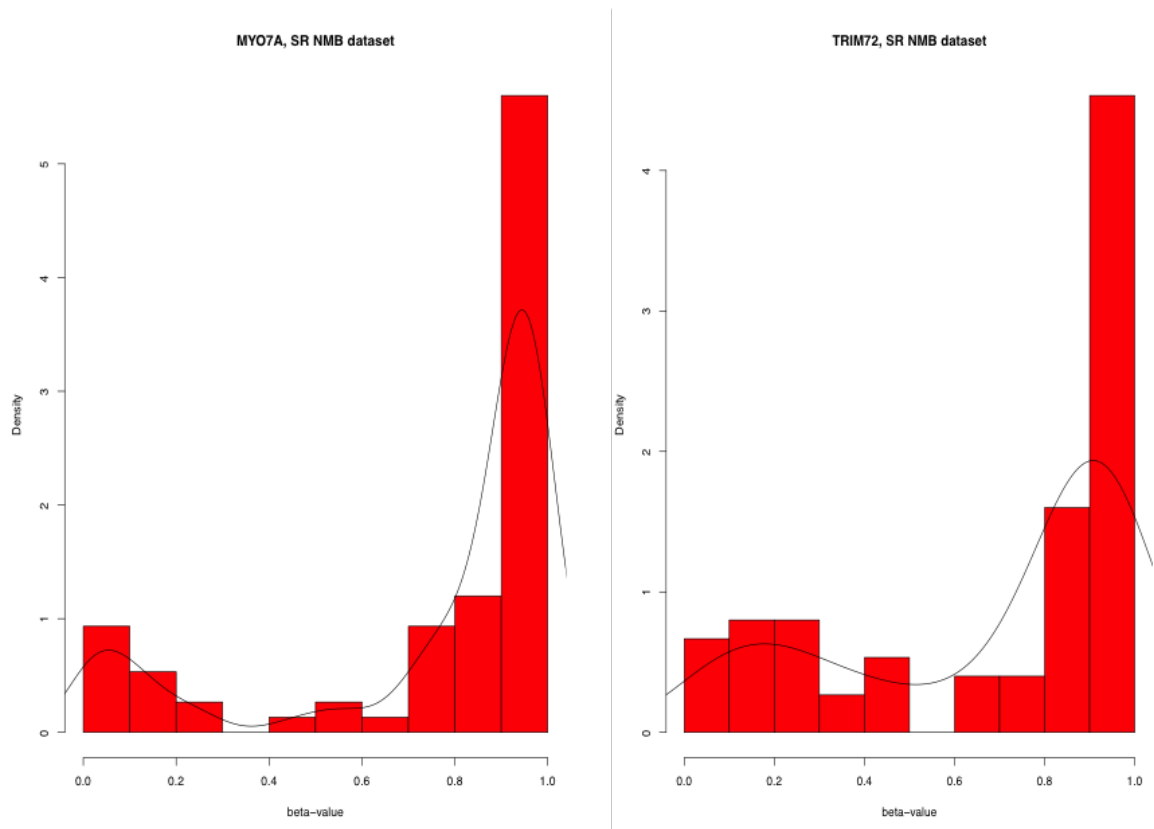


FIGURE 5.2. Methylation pattern of the identified prognostic methylation markers are shown for *MYO7A* and *TRIM72* genes.

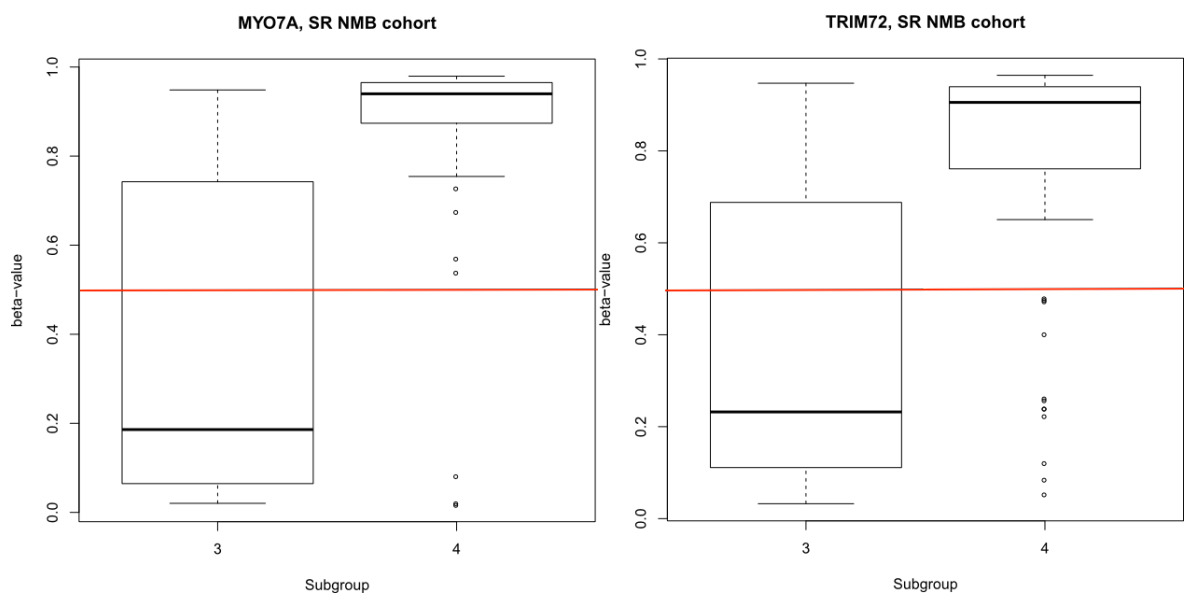


FIGURE 5.3. Distribution of β -scores of the identified prognostic methylation markers are shown as box-plots for *MYO7A* and *TRIM72* genes across non-WNT/non-SHH samples. Median score is shown as thick black horizontal line. Red line indicates selected β -values cut-off of 0.5. Subgroup 3 stands for Group 3, and subgroup 4 stands for Group 4 medulloblastoma subtypes.

Multivariate Cox proportional hazards model that takes into account methylation status of both loci simultaneously (Table 5.3), shows that for a subset of 47/75 samples, where the identified prognostic probes are both methylated, there was a favourable

outcome with 5-year survival rate of 76% (95% CI 65-91), and absence of one or the other, or both methylation markers conferred a poor prognosis (22%, 95% CI 7-70 and 33%, 95% CI 14-85 respectively) with $p < 0.001$, which is also demonstrated on the Kaplan-Meier plot (Figure 5.5).

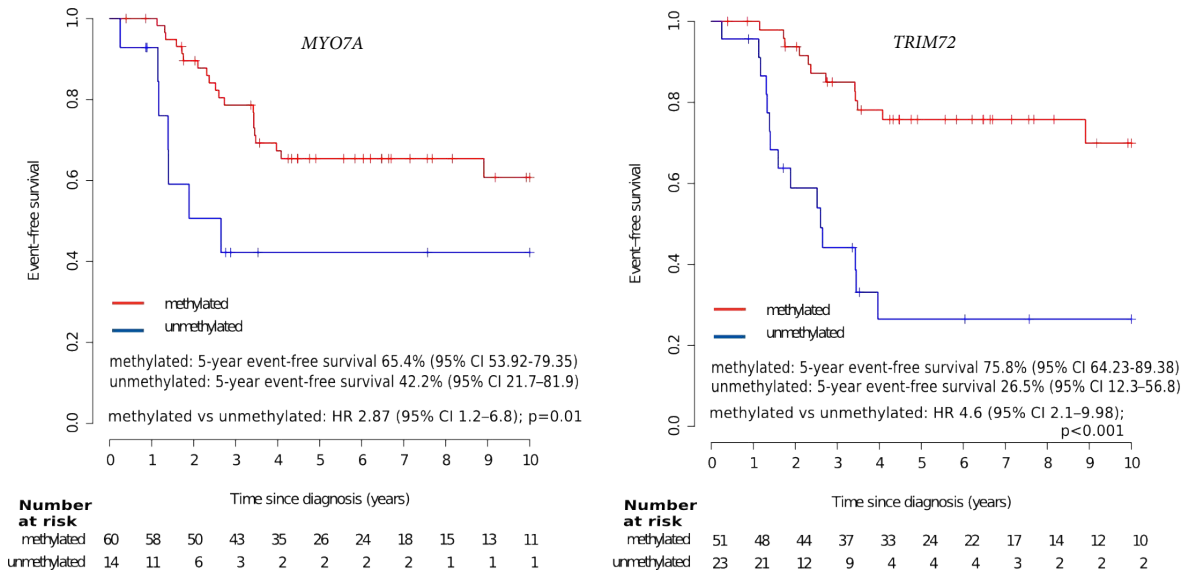


FIGURE 5.4. Kaplan-Meier plots of prognostic methylation markers in standard risk non-WNT/non-SHH medulloblastoma are shown for *MYO7A* and *TRIM72* genes; Red-methylated, blue - unmethylated.

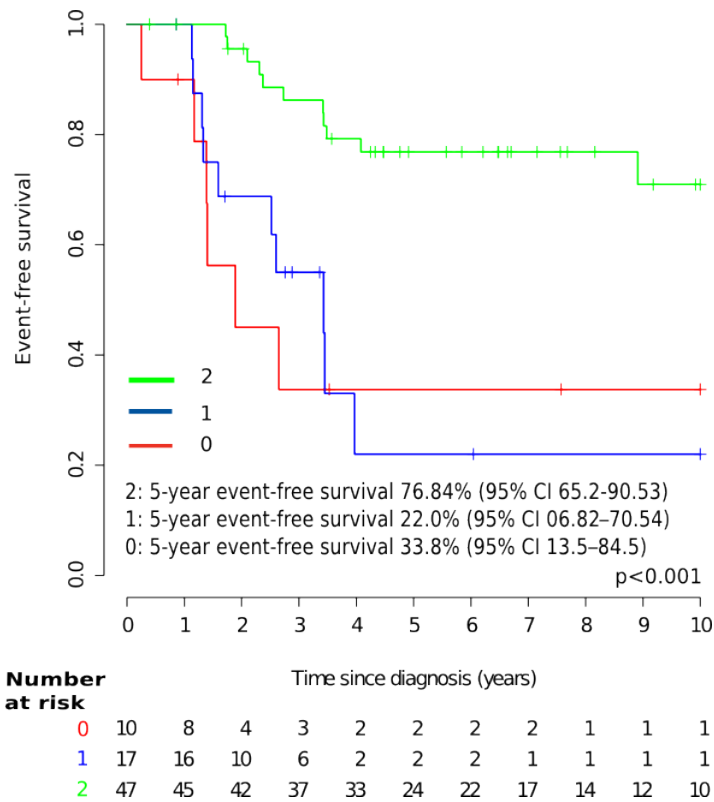


FIGURE 5.5. The number of prognostic methylation markers in relationship to survival in the standard risk non-WNT/non-SHH cohort. Kaplan-Meier plots are shown for each occurrence of prognostic methylation markers frequency: 0- red, 1 (either *MYO7A* or *TRIM72*)-blue, 2 (both simultaneously) - green.

5.4.7. Integration of previously reported survival markers with identified methylomic correlates . An extended Cox proportional hazard model, which incorporates both identifies prognostic loci and PNET4 prognostic signature (4.1.2) introduced as an additional covariate, and Kaplan-Meier curves (Figure 5.6) plotted for this model, show 100% (95% CI 100-100) 5-year survival rate for the subset of 6/75 patients positive for all three prognostic markers, and 73.45% (95% CI 60.57-89.06) for the subset of 42/75 patients positive for at list two out of three tested biomarkers, followed by 26% (95% CI 10.3-65.6) and 28.57% (95% CI 8.86-92.18) 5-year survival rates for the 18/75 patients positive for only one of the markers and the 8/75 subset with neither of the markers correspondingly.

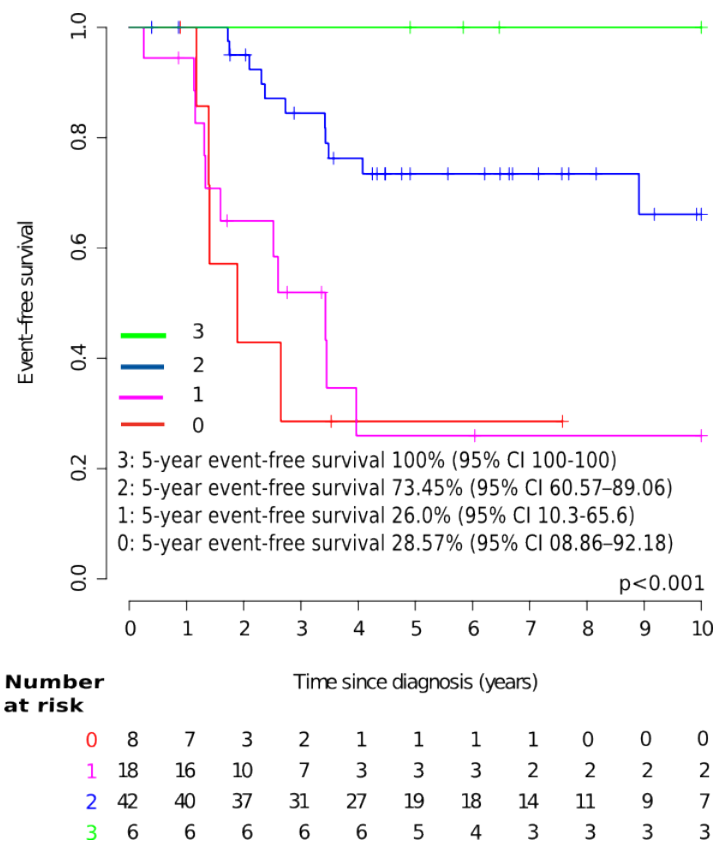


FIGURE 5.6. The number of prognostic methylation markers combined with PNET4 in relationship to survival in the standard risk non-WNT/non-SHH cohort. Kaplan-Meier plots are shown for each occurrence of prognostic methylation markers frequency: 0- red, 1 - magenta, 2 - blue, 3 - green.

5.4.8. Investigation of prognostic potential of identified methylomic markers in high-risk non-WNT/non-SHH medulloblastoma and integration of the additional prognostic covariates . The behaviour of methylomic markers, that were prognostic in the standard risk, non-WNT/non-SHH cohort, was investigated in

Gene	Probe	Univariate			Multivariate		
		Hazard Ratio	p-value	C-index (concordance)	Hazard Ratio	p-value	C-index (concordance)
<i>MYO7A</i>	cg08879470	1.55 95% CI=[0.84-2.8]	0.2	0.57 (se = 0.038, 95% CI=[0.49-0.64])	0.67 95% CI=[0.34-1.2]	0.23	0.59 (se = 0.044 95% CI=[0.503-0.67])
<i>TRIM72</i>	cg00982958	1.28, 95% CI=[0.7-2.3]	0.4	0.55 (se= 0.039, 95% CI=[0.47- 0.62])	0.9 95% CI=[0.47-1.72]	0.7	0.59 (se = 0.044 95% CI=[0.503-0.67])

TABLE 5.4. Assessment of prognostic methylation markers as binary variables at 0.5 cut-off in univariate and multivariate analysis of high risk non-WNT/non-SHH medulloblastoma cohort. Hazard ratios with 95% confidence interval, p-values are shown for each methylation covariate.

the high risk non-WNT/non-SHH cohort of 100 patients (see section 4.3.2.3). As before, the methylation status of both markers, was dichotomised in the same way as in the standard risk cohort, using a β -value cut-off of 0.5.

Log-rank tests did not identify any significant relationship between either of the loci and survival in the high risk cohort ($p=0.2$ for the *MYO7A* locus, and $p=0.4$ for the *TRIM72*). Neither there was any significant relationship observed between survival and methylation status of both loci, when taken into account simultaneously. These relationships are shown in Figure 5.7 and 5.8. 100% survival was observed for samples positive for all markers when combined with PNET4 scheme, however, either of them on their own or combination of any 2 did not sure any significant difference in the survival (Figure 5.8B). Cox proportional hazard models are presented in the Table 5.4

Five year event-free survival in the subset of 69/100 samples with a methylated *MYO7A* locus was 52.9% (95% CI 42-67) as opposed to the 49.2% (34-71) in the subset of 31/100 samples with unmethylated status of this gene, with hazard ratio of 1.55 (95% CI 0.84-2.8), $p=0.2$.

Five year event-free survival in the subset of 65/100 samples with a methylated *TRIM72* locus was 51.47% (95% CI 39.84-66.5) *vs* 50.79% (36.51-70.66) in the subset of 35/100 samples with unmethylated status of this gene, with hazard ratio of 1.28 (95% CI 0.7-2.3), $p=0.4$.

5.4.8.1. *Metastatic stage as a prognostic covariate* . As neither of the identified loci was found significantly prognostic in the high-risk non-WNT/non-SHH medulloblastoma but both markers were significant in the standard-risk cohort, it was likely that one of the high-risk features was playing role. As metastatic status was found prognostic in this cohort and in the previous studies (Chapter 4, Shih et al. (2014)) it was assumed to be the most likely candidate. Log-rank tests identified a significant relationship in non-metastatic cases between survival and methylation of the *MYO7A*

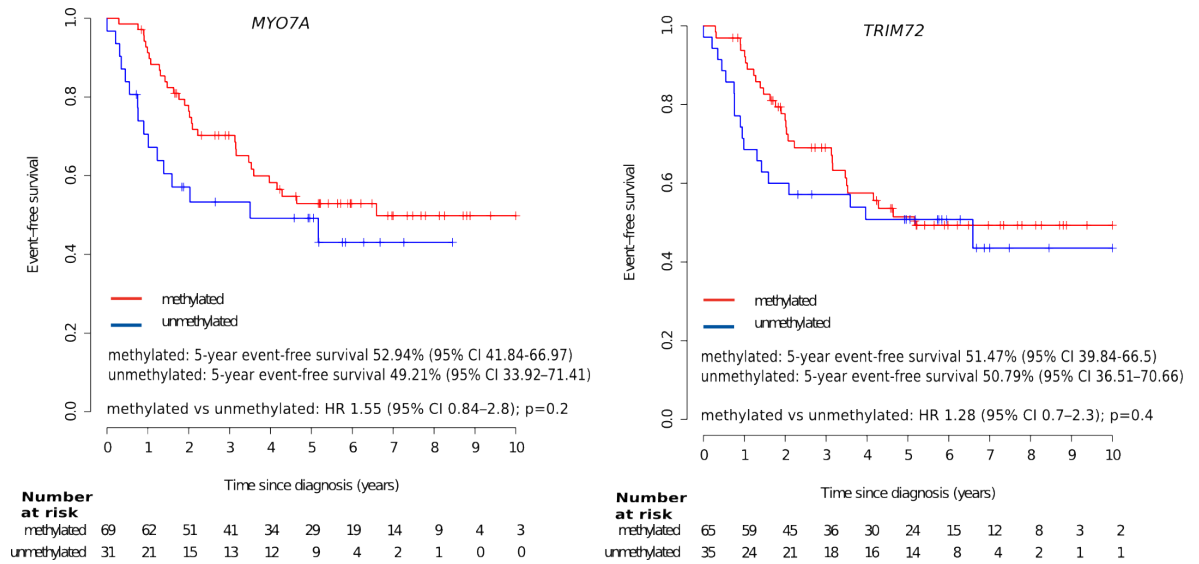


FIGURE 5.7. Kaplan-Meier plots of identified methylation markers are shown for *MYO7A* and *TRIM72* genes in high-risk non-WNT/non-SHH cohort; red-methylated, blue - unmethylated.

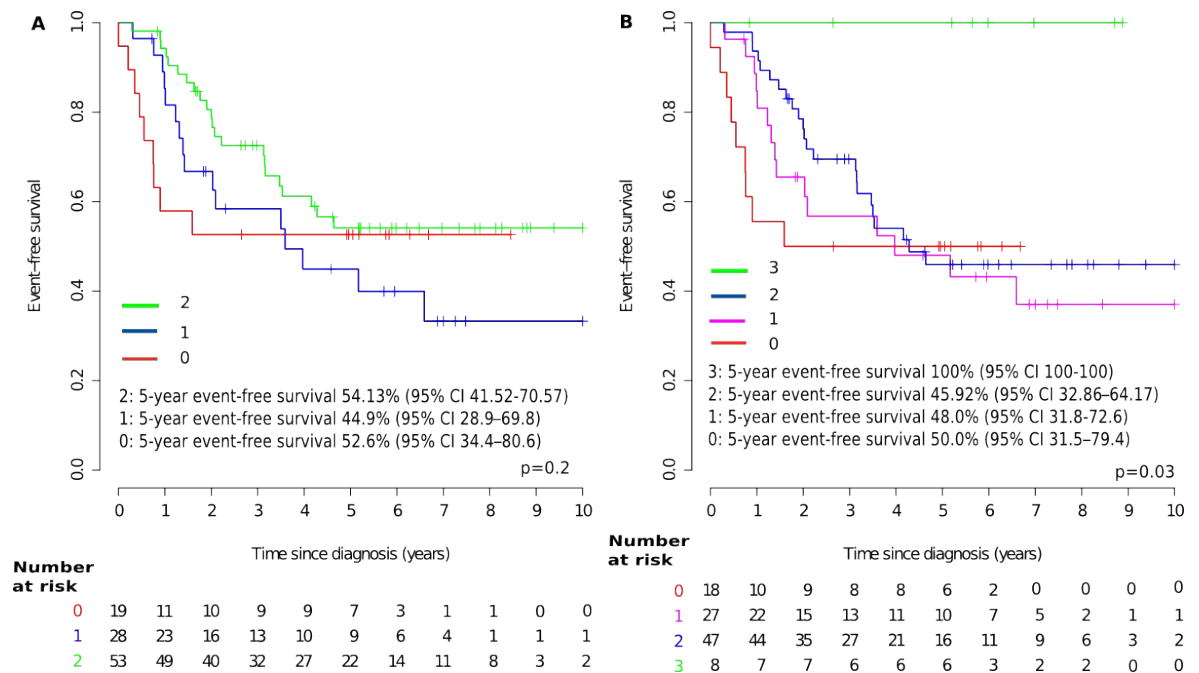


FIGURE 5.8. The number of prognostic methylation markers (A), and methylation markers combined with PNET4 scheme signature (B), in relationship to survival in the high risk non-WNT/non-SHH cohort. Kaplan-Meier plots are shown for each occurrence of prognostic methylation markers frequency: (A) 0 - red, 1 - blue, 2 - green. (B): 0 - red, 1 - magenta, 2 - blue, 3 - green.

locus (p-value=0.008, Figure 5.9 A), methylation of the *TRIM72* locus (p-value=0.03, Figure 5.9 B) and methylation of both loci *MYO7A* and *TRIM72* simultaneously (p-value=0.02, Figure 5.9 C). Moreover, survival was significantly associated with the methylation of both loci in combination with the previously described PNET4 cytogenetic prognostication signature (see section 4.2) (p-value=0.04, Figure 5.9 D).

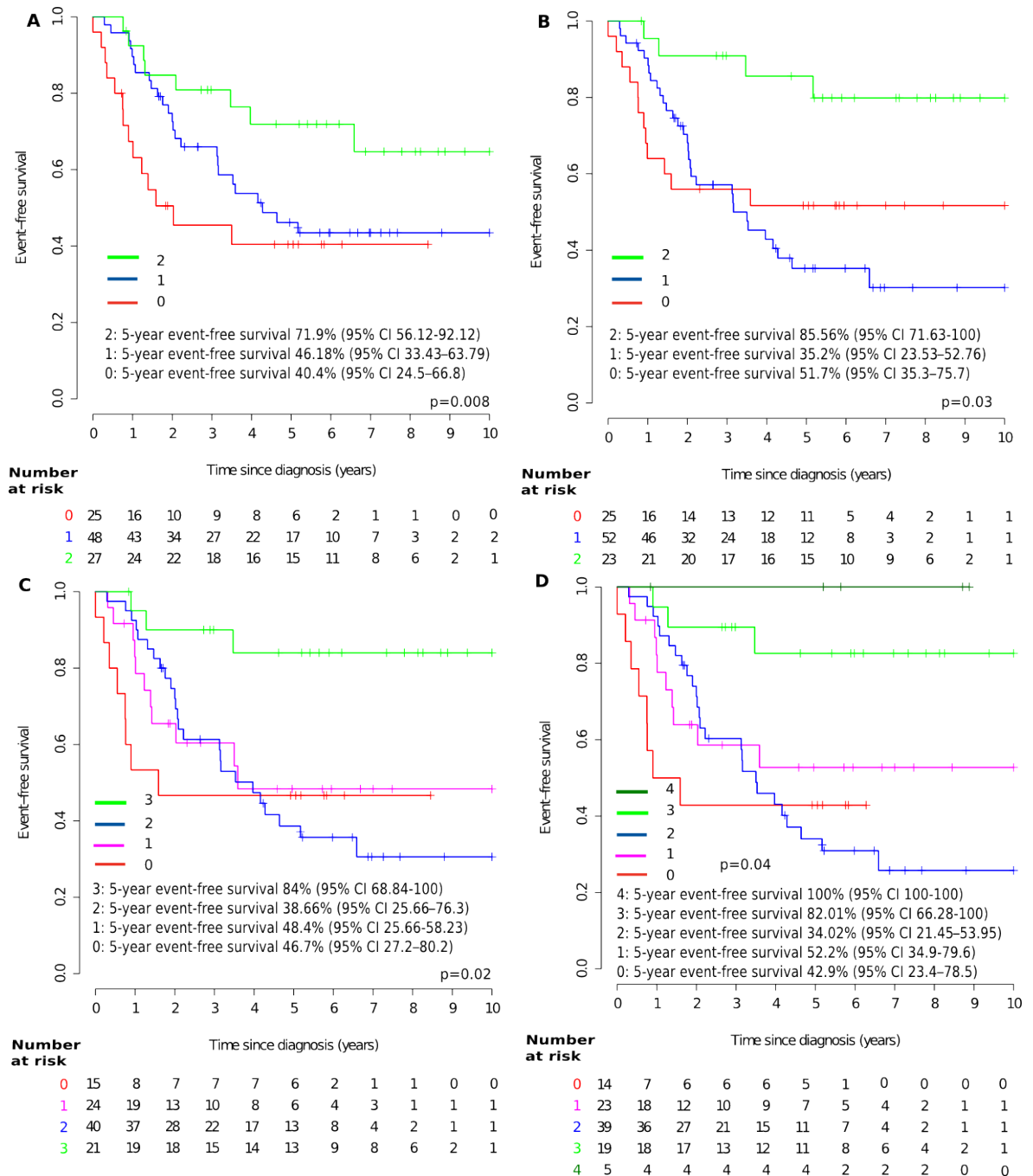


FIGURE 5.9. Kaplan-Meier plots of identified methylation markers are shown for *MYO7A* (A); and *TRIM72* (B) genes and M0 status as a covariate in high-risk non-WNT/non-SHH cohort. The number of prognostic methylation markers with metastatic status (M0) as additional covariate (C), and methylation markers combined with PNET4 scheme signature and metastatic status (M0) as additional covariate (D), in relationship to survival in the high risk non-WNT/non-SHH cohort. Kaplan-Meier plots are shown for each occurrence of prognostic methylation markers frequency:(A) 0 - red, 1 - blue, 2 - green. (B): 0 - red, 1 - blue, 2 - green. (C) 0 - red, 1 - magenta, 2 - blue, 3 - green. (D) 0 - red, 1 - magenta, 2 - blue, 3 - green, 4 - dark green.

5.4.9. Validation of the identified methylomic biomarkers in the external cohort with limited risk status information . The behaviour of the identified methylomic markers was investigated in the mixed-risk external age-matched cohort

(n=244) with limited risk-status information. Available high-risk markers in this cohort were assessed in section 5.4.2

Log-rank tests identified a significant relationship between overall survival and methylation of the *MYO7A* locus (p-value=0.002, Figure 5.10), methylation of both *MYO7A* and *TRIM72* loci simultaneously (p-value=0.01, Figure 5.11 A), and between both loci in combination with the PNET4 prognostication signature (p-value<0.001, Figure 5.11 B). However, the relationship between survival and the methylation status of *TRIM72* was not significant (p-value=0.3, Figure 5.10).

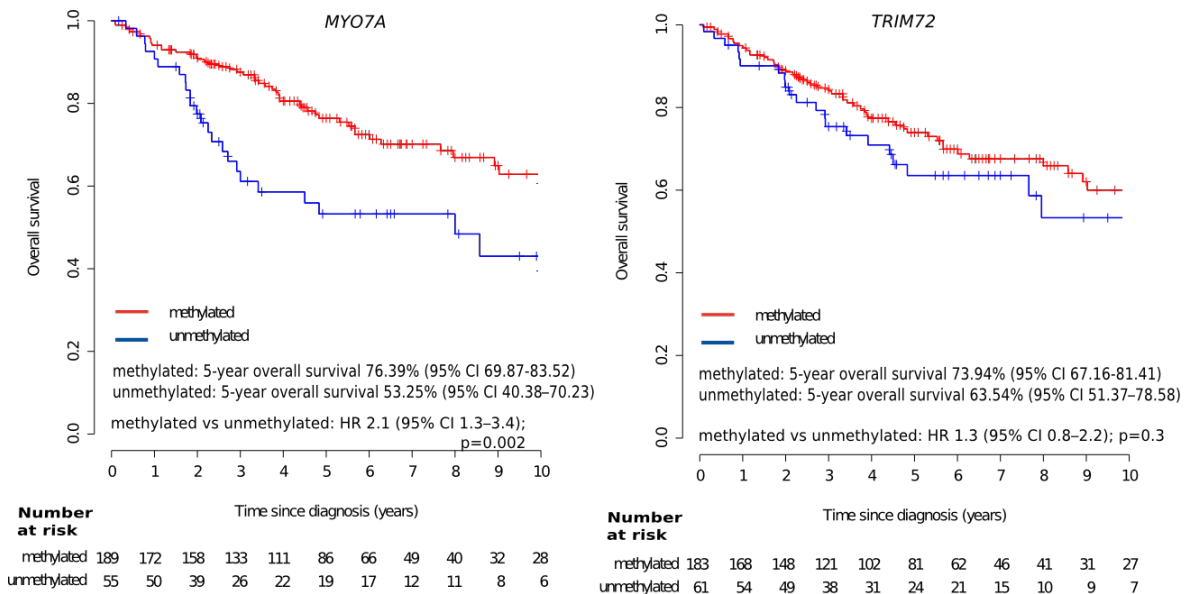


FIGURE 5.10. Kaplan-Meier plots of identified methylation markers are shown for *MYO7A* and *TRIM72* genes in mixed-risk external cohort (Cavalli et al., 2017); red-methylated, blue - unmethylated.

When metastatic stage was added as additional covariate, for the *MYO7A* locus, there was a significant difference between 5 year survival rates of patients with two low risk markers (i.e. non-metastatic and methylated) (80.1% 95% CI 72.3-88.), one marker (65%, CI 55.6-76) and with no low risk markers (54.6% 95% CI 36.4-82.1), p-value=0.004 (Figure 5.12 A); no significant difference for *TRIM72* with 75.98%, 69.3% and 57.9% for occurrences of 2, 1 and 0 low risk markers correspondingly, p-value=0.1 (Figure 5.12 B). When the methylation status of both loci were considered simultaneously alongside metastatic stage, there was no significant difference in 5 year survival between patients with 2 or 3 (77.9% and 77.3%) low risk markers, but the difference was significant for patients with 0 (62.5% 95% CI 36.5-100) or 1 (48%) marker, and patients with 2 or 3 markers, with p-value=0.005 (Figure 5.12 C), and when the PNET4 signature was added as an additional covariate: there was no significant

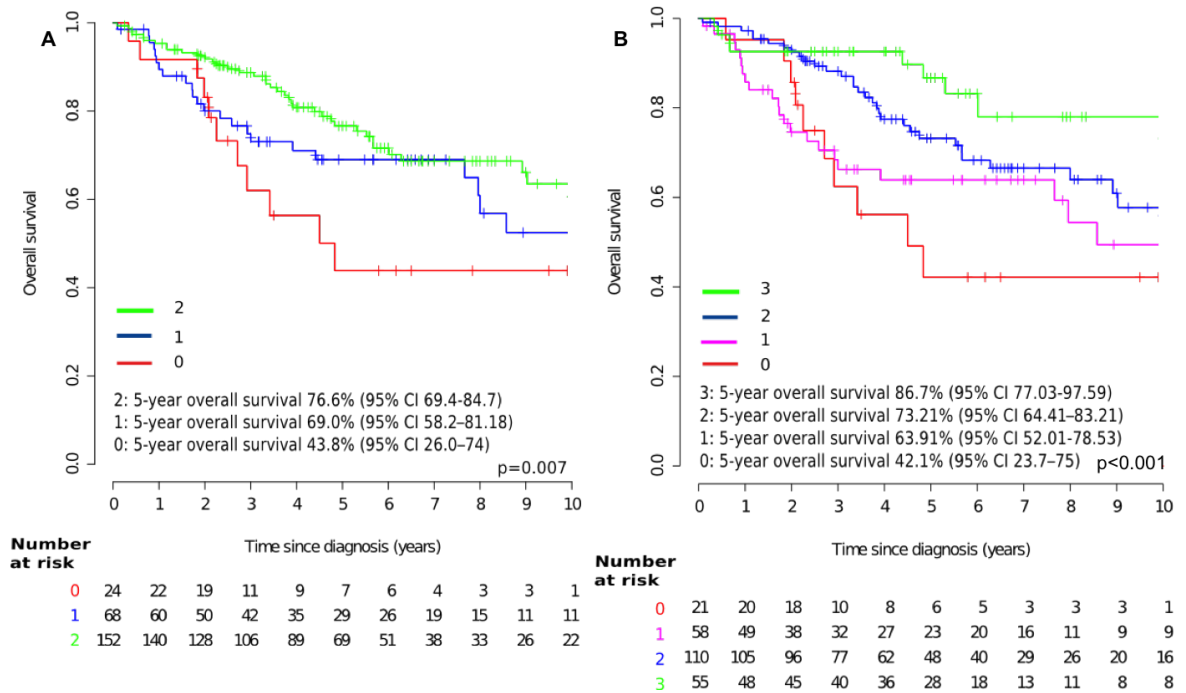


FIGURE 5.11. The number of prognostic methylation markers (A), and methylation markers combined with PNET4 scheme signature (B), in relationship to survival in mixed-risk external cohort (Cavalli et al., 2017). Kaplan-Meier plots are shown for each occurrence of prognostic methylation markers frequency: (A) 0 - red, 1 - blue, 2 - green. (B): 0 - red, 1 - magenta, 2 - blue, 3 - green.

difference in 5 year survival between patients with 2,3 or 4 low risk markers (70%, 81% and 82% respectively), but the difference was significant for patients with 0 (66.7% 95% 37.9-100) or 1 (42.5%, 95% CI 28.7-62.3) marker, and with 2,3 or 4 with p -value=0.0004 (Figure 5.12 D). This means that addition of metastatic status as a covariate did not significantly change the overall trend in survival of the patients in this cohort.

5.4.10. Formulation of extended testable survival model for risk stratification of patients with non-WNT/non-SHH medulloblastoma and its application to cohorts of variable risk status . A combined prognostication scheme for risk assessment in medulloblastoma was formulated by extending the existing PNET4 signature prognostication scheme (Goschzik et al., 2018). This scheme, which was initially developed for standard risk non-WNT/non-SHH medulloblastoma, was found to be prognostic in high risk disease earlier in this research project (Chapter 4.1.2). This scheme is presented on Figure 5.13 with the grey/dotted area indicating a proposed extension for all non-WNT/non-SHH tumours - this alternative pathway places metastatic patients as well as patients with 0-1 methylated loci and negative for the PNET4 signature into a high-risk group; non-metastatic patients with either both loci

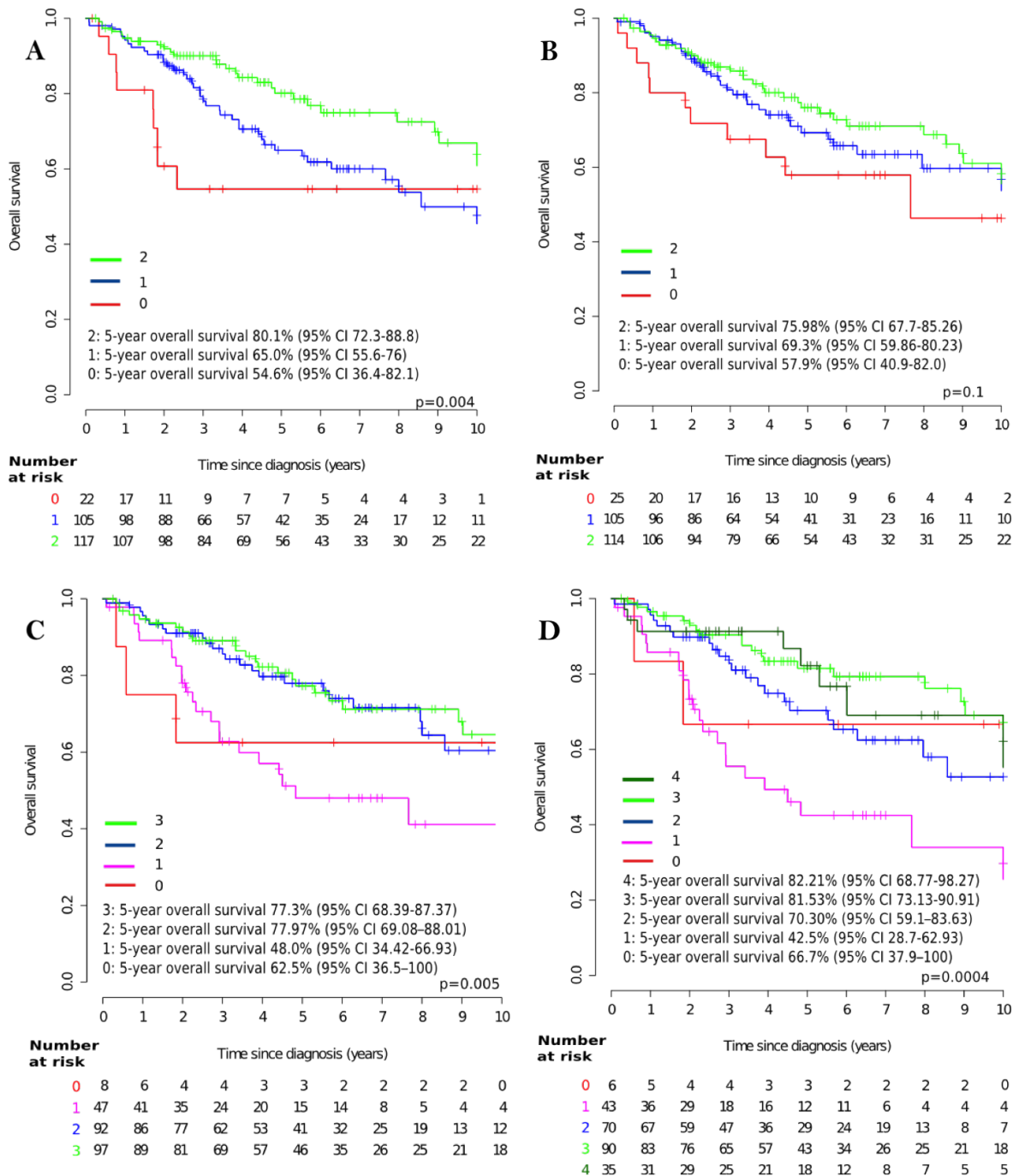


FIGURE 5.12. Kaplan-Meier plots of identified methylation markers are shown for *MYO7A* (A); and *TRIM72* (B) genes and M0 status as a covariate in the external non-WNT/non-SHH cohort (Cavalli et al., 2017). The number of prognostic methylation markers with metastatic status (M0) as additional covariate (C), and methylation markers combined with PNET4 scheme signature and metastatic status (M0) as additional covariate (D), in relationship to survival in the uncertain risk non-WNT/non-SHH cohort. Kaplan-Meier plots are shown for each occurrence of prognostic methylation markers frequency:(A) 0 - red, 1 - blue, 2 - green. (B): 0 - red, 1 - blue, 2 - green. (C) 0 - red, 1 - magenta, 2 - blue, 3 - green. (D) 0 - red, 1 - magenta, 2 - blue, 3 - green, 4 - dark green.

methyated loci or PNET4 signature positivity into standard-risk, and non-metastatic cases positive for PNET4 signature and both loci methyated into a low-risk group. The derivation of this scheme is reported below.

Survival models described in sections 5.4.6 and 5.4.7 shown that simultaneous methylation of both probes confer favourable prognosis (76% 5-year survival) and absence of methylation of one or the other, or both probes result in dismal outcome (22% 5-year survival) and that the methylation status of the probes was additionally prognostic when considered alongside the PNET4 signature.

Further investigation showed that the survival of the non-metastatic cases where both probes were methylated simultaneously indicated favourable prognosis and was comparable to the survival in the standard-risk cohort (72%, Figure 5.9 C *vs* 76% Figure 5.5) or either absence of methylation of one or the other, or both probes, or evidence of metastasis, resulted in dismal outcome (38-46% *vs* 22-33%). In the similar way, non metastatic cases with both probes methylated in combination with PNET4 signature showed survival identical to the standard-risk cohort (100% Figures 5.9 D and 5.6).

It was noted, that in stratification models in standard risk (Figure 5.5 and 5.6), non-metastatic high risk (Figure 5.9) and external (Figure 5.12) cohorts, survival increased with the increase of the total number of independently significant low risk markers (Table 5.7). Cases positive for PNET4 signature consistently showed better survival across all cohorts. Therefore, the high risk group is associated with 0-2 positive prognostic markers (despite some discrepancy shown on Figures 5.9C and D, and 5.12 C and D, which can be attributed to low numbers of the patients with 0 prognostic markers), standard risk with 2-3 positive prognostic markers and low risk with 3-4 positive prognostic markers.

The results of application of the scheme to these cohorts are shown as Kaplan-Meier plots on Figure 5.14 for NMB cohorts and Figure 5.15 for Cavalli cohort. Figure 5.14 offers a comparison of application of the scheme to the high risk (A) and standard risk (B) NMB non-WNT/non-SHH cohorts. The top panels of both figures show flowcharts for patient risk assignment and the bottom panels show the corresponding Kaplan-Meier plots. Both cohorts show similar behaviour and significant survival differences between low risk group (100% for both high risk and standard risk cohorts, labelled as LRisk on the plots), standard risk (79.5 (95% CI 61.1-100) for high risk NMB cohort and 72.5 (95% CI 59.8-87.8) for standard-risk NMB cohort correspondingly, labelled SRisk on the plots) and high risk group (43.2% (95% CI 33.06-56.38) and 23.6%

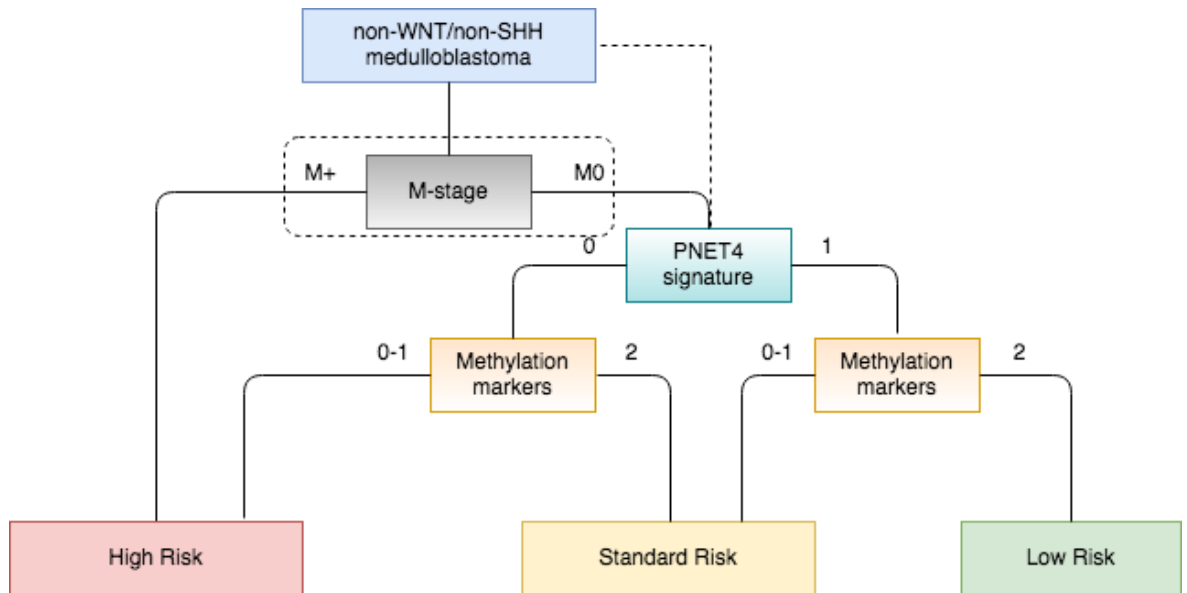


FIGURE 5.13. Decision tree for classification of non-WNT/non-SHH survival cohorts. Classification tree derived considering 4 potentially prognostic covariates: metastatic disease (M stage), PNET4 signature prognostication scheme, and methylation markers (*MYO7A* and *TRIM72*).

(95% CI 9.8-56.78) for high risk and standard risk NMB cohorts correspondingly) with p-value=0.01 for high risk NMB cohort and p-value=0.0003 for standard risk NMB cohort. As standard risk NMB cohort is non-metastatic by definition, both pathways of the scheme (with and without M-stage) produce the same result.

Figure 5.15 illustrates the application of the scheme to the Cavalli data-set of uncertain risk status. Figure 5.15 (A) shows the results of application of the scheme including consideration of metastatic status (M-stage): a flowchart for patient assignment is shown on the top panel and the corresponding Kaplan-Meier plot on the bottom. There was no significant survival difference in overall 5 years survival between low risk group (82.2%, 95% CI 68.7-98.27, labelled LRisk on the plot), standard risk (78.24%, 95% CI 68.3-89.63, labelled SRisk on the plot) and high risk group (64.9% with 95% CI 56.6-74.54, labelled HRisk) with p-value=0.1.

Figure 5.15 (B) shows results of application of the scheme's pathway to the Cavalli dataset when metastatic status is disregarded. There is a significant survival difference in overall 5 year survival between the low-risk group (86.7%, 95% CI 77.03-97.59, labelled LRisk on the plot), standard-risk (72.9%, 95% CI 64.12-82.3, labelled SRisk on the plot) and high-risk group (57.97% with 95% CI 46.98-71.54, labelled HRisk) with p-value=0.007.

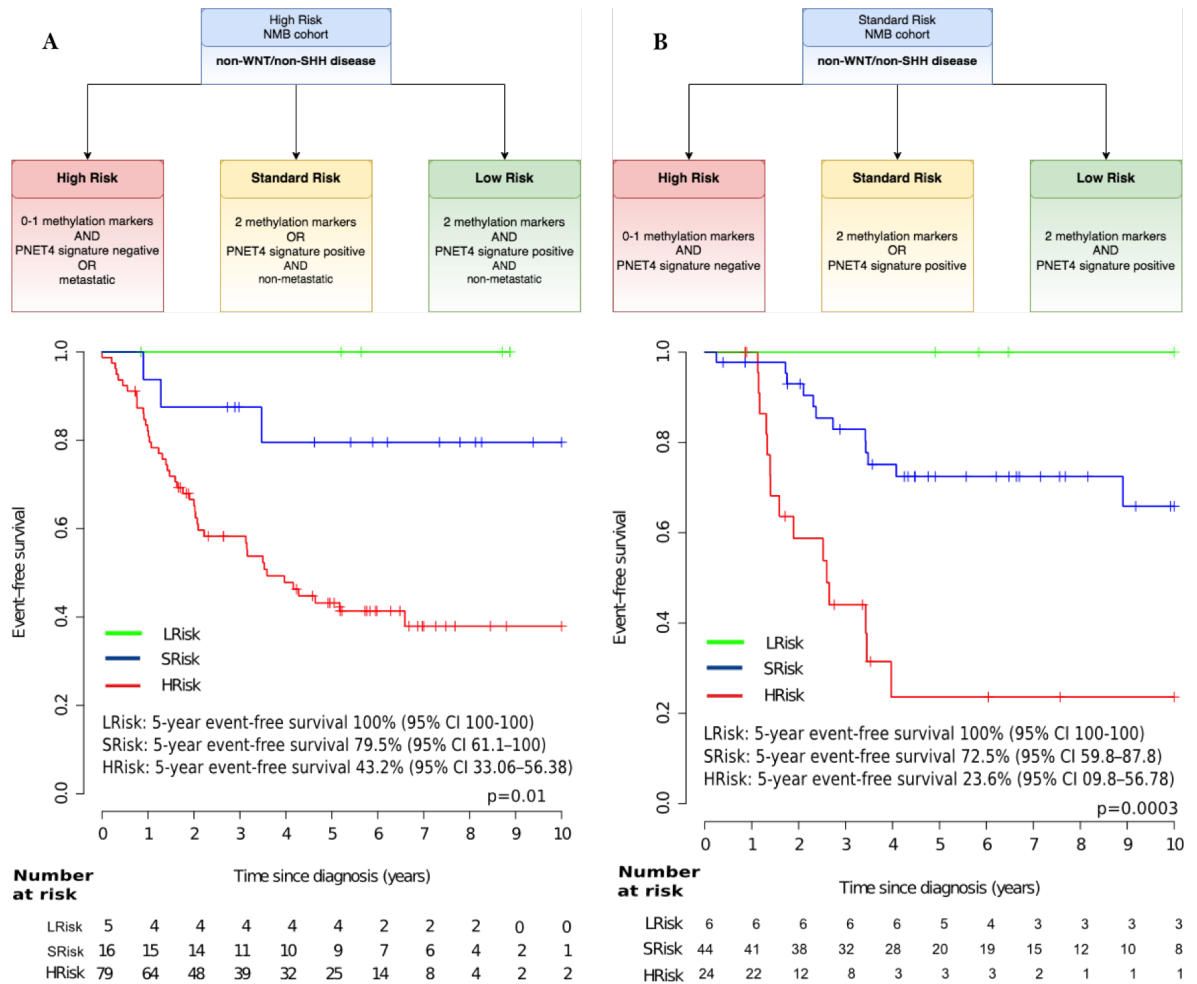


FIGURE 5.14. DNA methylation markers add prognostic value to PNET4 signature prognostication scheme in the high risk non-WNT/non-SHH cohort. (A) Top panel shows flowchart for risk assignment of the patients in the high risk non-WNT/non-SHH cohort. Bottom panel shows refinement of PNET4 signature prognostication scheme by addition of the methylomic markers applied to the high risk non-WNT/non-SHH cohort as Kaplan-Meier plots of event free survival; (B) Top panel shows flowchart for risk assignment of the patients in the standard risk non-WNT/non-SHH cohort. Bottom panel shows refinement of PNET4 signature prognostication scheme by addition of the methylomic markers applied to the standard risk non-WNT/non-SHH cohort as Kaplan-Meier plots of event free survival; Low risk is represented by green colour, standard risk by yellow and high risk by red colour.

Tables 5.5 and 5.6 show relationship between non-WNT/non-SHH medulloblastoma subtypes described by Northcott et al. (2017) and Schwalbe et al. (2017b) (both refined by Sharma et al. (2019)) and risk groups defined by the combined methylomic/PNET4 scheme. All cases defined as low-risk in HR NMB cohort and most in Cavalli dataset were associated with lower-risk subtypes by defined by both studies, however cases defined as high-risk and standard risk showed mixed association in both cohorts, indicating a need for further stratification in non-WNT/non-SHH medulloblastoma with clinicopathological high-risk features.

Cohort	HR-NMB			Cavalli		
Subtype	HR	SR	LR	HR	SR	LR
I	2	0	0	2	0	0
II	8	0	0	10	1	0
III	13	0	0	10	0	0
IV	4	0	1	3	3	0
V	7	3	0	8	2	0
VI	4	2	4	6	5	7
VII	12	2	0	26	14	19
VIII	13	5	0	17	26	2
MBNOS	3	1	0	7	8	1
p	0.002			<0.001		

TABLE 5.5. Cross-tabulation of combined methylomic/PNET4 scheme-defined risk groups and Group 3/4 disease subtypes described by Northcott et al. (2017) in High-Risk (HR) NMB and Cavalli cohorts.*MBNOS-Medulloblastoma not otherwise specified. HR-High-risk, SR-Standard-risk, LR - Low-risk.

Cohort	HR-NMB			Cavalli		
Subtype	HR	SR	LR	HR	SR	LR
Group3-HR1	8	0	0	10	1	0
Group3-HR2	3	0	0	3	1	0
Group3-LR1	10	0	0	5	3	1
Group3-LR2	8	0	1	8	0	0
Group4-HR	13	5	0	18	26	3
Group4-LR1-1	6	0	0	16	10	6
Group4-LR1-2	5	2	0	7	1	10
Group4-LR2-1	4	2	4	5	6	7
Group4-LR2-2	7	3	0	8	3	0
MBNOS*	2	1	0	9	8	3
p	0.01			0.001		

TABLE 5.6. Cross-tabulation of combined methylomic/PNET4 scheme-defined risk groups and Group 3/4 disease subtypes described by described by Schwalbe et al. (2017b) in HR NMB cohort.*MBNOS-Medulloblastoma not otherwise specified. HR-High-risk, SR-Standard-risk, LR - Low-risk.

	SR NMB cohort	HR NMB cohort	Cavalli cohort	Cavalli cohort (M-stage disregared)
LR	100% 95% CI (100-100)	100% 95% CI (100-100)	82.2% 95% CI (68.7-98.27)	86.7%, 95% CI (77.03-97.5)
SR	72.5% 95% CI (59.8-87.8)	79.5% 95% CI (61.1-100)	78.24% 95% CI (68.3-89.63)	72.9%, 95% CI (64.12-82.3)
HR	23.6% 95% CI (9.8-56.78)	43.2% 95% CI (33.06-56.38)	64.9% 95% CI (56.6-74.54)	57.97% 95% CI (46.98-71.5)
p	0.0003	0.01	0.1	0.007

TABLE 5.7. Combined methylomic/PNET4 prognostication scheme comparison table for 5 years survival in non-WNT/non-SHH medulloblastoma: standard risk NMB cohort *vs* high risk NMB cohort *vs* mixed-risk cohort and mixed-risk with M-stage disregared cohort (Cavalli dataset)

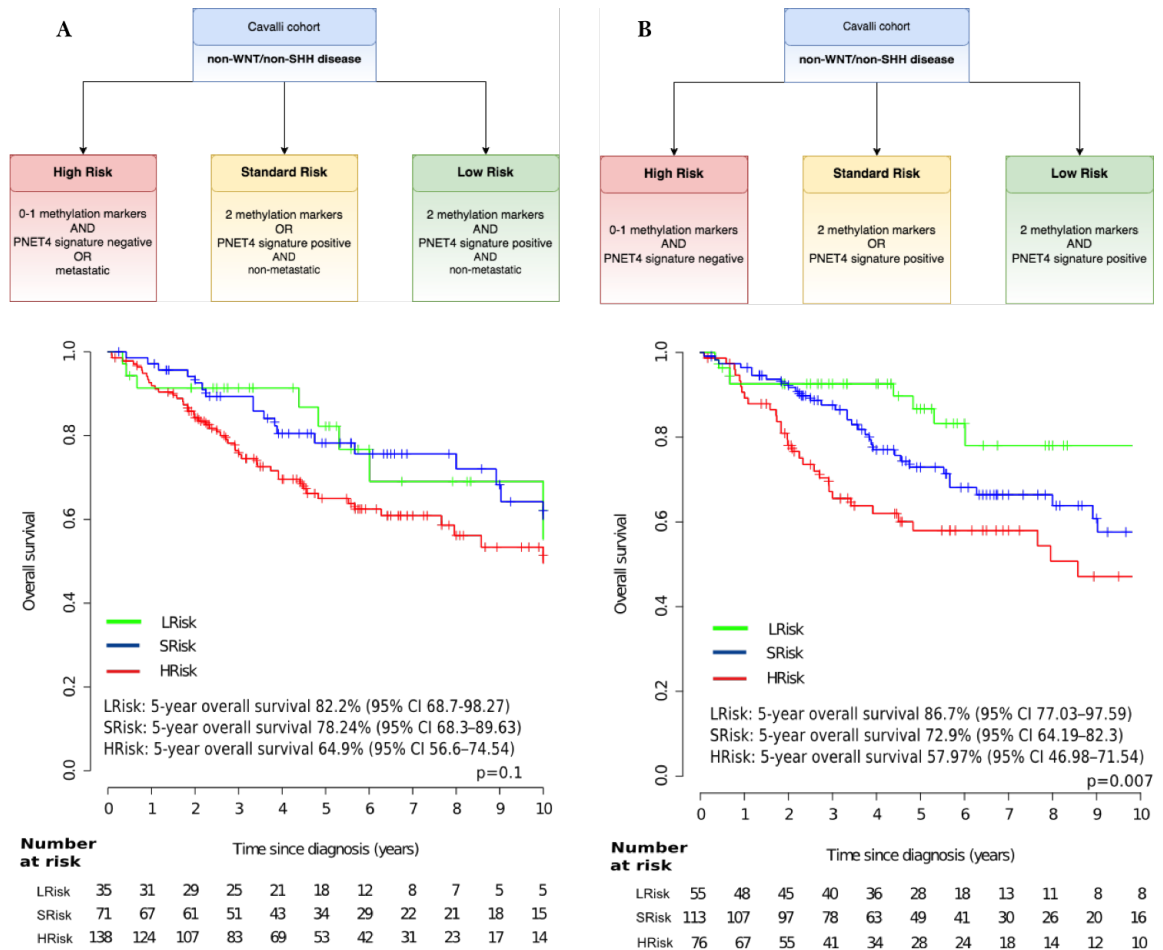


FIGURE 5.15. DNA methylation markers add prognostic value to PNET4 signature prognostication scheme in the mixed risk external cohort. (A) Top panel shows flowchart for risk assignment of the patients in the Cavalli dataset (Cavalli et al., 2017). Bottom panel shows refinement of PNET4 signature prognostication scheme by addition of the methylomic markers applied to the Cavalli dataset as Kaplan-Meier plots of overall survival; (B) Top panel shows flowchart for risk assignment of the patients in the Cavalli dataset (Cavalli et al., 2017) omitting metastatic status. Bottom panel shows refinement of PNET4 signature prognostication scheme by addition of the methylomic markers applied to the Cavalli dataset omitting metastatic status as Kaplan-Meier plots of overall survival; Low risk is represented by green colour, standard risk by yellow and high risk by red colour.

5.4.11. Relationship between methylation of identified loci and their gene expression . In the previous sections it was established that methylation status of the identified probes has effect on survival. In order to explore possible causes of how this effect could be manifested, and see if change in gene expression might be affected variable, scatter plots of the gene expression against beta-values for both loci were plotted. The Pearson's correlation coefficients in the standard risk NMB cohort was -0.0041 with p-value=0.983 for *MYO7A* (Figure 5.16 (A) top panel), 0.1804, p-value=0.3402, for *TRIM72* (Figure 5.16 (A) bottom panel); in high-risk NMB cohort: 0.1087, p-value=0.46621 for *MYO7A* (Figure 5.16 (B) top panel), 0.007, p-value=0.9624

for *TRIM72* (Figure 5.16 (B) bottom panel); and mixed (standard and high risk NMB cohorts together): for *MYO7A* 0.0637 with p-value=0.5796 (Figure 5.16 (C) top panel), 0.0801, p-value=0.4855 for *TRIM72* (Figure 5.16 (C) bottom panel). The scatter plots (Figure 5.16) of the relationship between methylation and gene expression levels of the *MYO7A* and *TRIM72* showed no correlation between methylation level of the probes and gene expression.

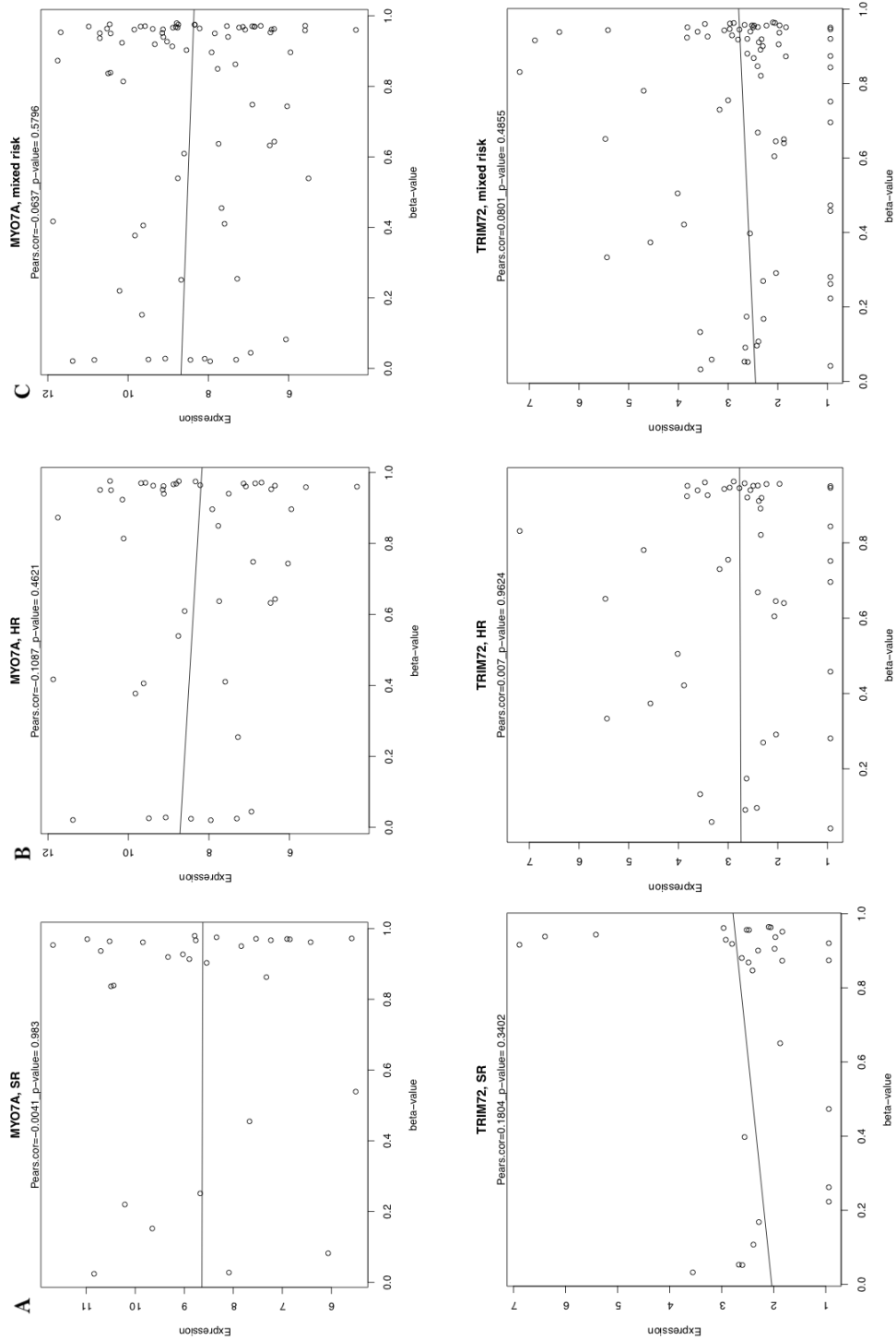


FIGURE 5.16. Relationship between methylation of identified loci and their gene expression. Scatter plot of the relationship between methylation and gene expression levels of the *MYO7A* on the top panel and *TRIM72* on the bottom in Standard risk NMB cohort (A), high risk NMB cohort (B) and mixed standard and high risk NMB cohorts together (C), with Pearson's correlation coefficient and corresponding p-value provided for each case. Dots represent patients samples and black line represents the regression line.

5.5. Summary and discussion

In chapter 4, it was demonstrated that the PNET4 signature scheme, which shows excellent prognostic results in the standard risk non-WNT/non-SHH medulloblastoma, has prognostic potential in high-risk, non-WNT/non-SHH disease. This chapter investigated the prognostic utility of methylomic biomarkers in cohorts with varied risk status, as determined by currently understood clinical and molecular markers, and the potential for their integration as additional prognostic markers with previously reported survival markers (Chapter 4.1.2, Goschzik et al. (2018)).

5.5.1. Mixed-risk survival cohort with some unavailable risk-factors to mimic the real world clinical settings. An age-matched independent cohort of 244 non-WNT/non-SHH medulloblastoma samples previously published in (Cavalli et al., 2017), was chosen as an external validation cohort, due its distinct features: this cohort contains both Group 3 and Group 4 samples and has some important risk factor statuses available, namely histology and metastatic status. However statuses of other risk factors were unknown. Therefore, this cohort was a mixture of high-risk samples and samples of standard risk on available features, i.e. the unknown risk factors could potentially be indicators of either standard or high-risk disease.

It is important to note that, contrary to what is understood in this disease, metastatic stage was not significantly associated with survival in this cohort. Metastatic status is a prominent and well-validated high-risk biomarker reported by Shih et al. (2014) in both Group 3 and Group 4 stratification schemes. In chapter 4, it was shown that metastatic status of samples in Group 3 and Group 4, when considered separately (section 4.4.2.1), or combined (section 4.4.3.2), was significantly associated with survival, confirming its prognostic value in non-WNT/non-SHH medulloblastoma.

5.5.2. The methylomic signature, consisting of both prognostic markers, identified favourable risk group in standard risk non-WNT/non-SHH medulloblastoma and is prognostic in conjunction with the PNET4 signature scheme. The methylation status of each of the identified prognostic markers was significantly associated with survival in both univariate and multivariate Cox proportional hazards models, however, the C-indices of each model indicated that the multivariate model had better predictive value than in the univariate models for both

MYO7A and *TRIM72* (Table 5.2). Kaplan-Meier plots revealed that despite showing significant survival differences separately, (Figure 5.4), both methylation probes have to be methylated simultaneously to confer a favourable outcome and unmethylated status of one or the other is associated with a poor prognosis (Figure 5.5). This observation could be explained by the fact that both loci are simultaneously methylated in the majority (62% or 47/75) of the cases while in only 17/75 (22%) cases were samples discordant for methylation at these two loci.

Addition of the identified methylation markers to the PNET4 cytogenetic prognostication scheme show additional predictive value in the standard-risk cohort, however, determining methylation status of specific loci is not currently a routinely applied clinical assay and moreover, the PNET4 scheme is capable of identifying a favourable risk subset of patients carrying the signature, with the survival rate between 93-100% (Chapter 4, Schwalbe et al. (2017b)) and identification of which is much more simpler and widespread enough to be used in clinical settings.

5.5.3. Prognostic potential of the methylomic markers in non-metastatic high-risk and uncertain risk non-WNT/non-SHH medulloblastoma. Neither of the methylation markers, considered together, or in isolation, were prognostic in the high risk non-WNT/non-SHH cohort. A potential reason for this could be one of the following possibilities: the presence of one or more of the high risk clinical markers could be the reason for the difference in behaviour of the methylation markers, since this is what defines the difference between the standard-risk and high-risk cohort by definition; alternatively, over-fitting occurred when these loci were identified in the standard risk cohort - the cohort size was relatively small (n=75), despite the fact that the GLM-net algorithm uses cross-validation to minimise a possibility of over-fitting.

Further analysis demonstrated that these methylation markers behaved in a similar manner in both standard-risk and non-metastatic high-risk cohorts. Therefore, it is likely that discrepancies in the behaviour of the methylation markers in these two survival cohorts could be attributed to the additional clinical risk factors, such as metastatic cases present in the high risk cohort rather than because of over-fitting. This finding demonstrates the possibility for use of these methylation markers for risk stratification of high-risk disease, and thus, to stratify patients for whom knowledge of

currently understood clinical risk factors is incomplete. Subsequently, a risk stratification scheme was formulated which employed newly developed methylation signature (both *MYO7A* and *TRIM72* loci methylated), PNET4 signature and M0 metastatic stage as low-risk markers (Section 5.4.10).

An external cohort of 244 uncertain-risk non-WNT/non-SHH medulloblastomas with only two clinical high risk factors available, M-stage and LCA pathology, was used to validate the behaviour of both identified methylomic markers, the combined methylomic signature and the formulated scheme 5.13 with mixed results. While methylation status of *MYO7A* locus was significantly associated with survival, methylation status of *TRIM72* showed no association with the survival. When both loci were combined and considered alongside the PNET4 signature, there was a significant association. Unsurprisingly, addition of metastatic status as a covariate did not improve risk stratification, as it was not found prognostic in this particular cohort. The newly formulated prognostic scheme 5.13 showed different behaviour to the high-risk NMB cohort (Figure 5.14 B): in this mixed cohort, the schema that did not take into account metastatic status performed better than the one that did 5.15. One possible explanation of this phenomena may be attributed to the unusual behaviour of metastatic stage in this cohort, which, as previously noted, was not prognostic. Also different in this cohort was that only overall survival information was available, rather than the event-free survival data considered in the other cohorts. It is important to note that in non-WNT/non-SHH medulloblastoma, relapse usually confers poor survival and records of time of relapse are important due to the tendency of this group to relapse later after diagnosis, sometimes around 5-years after diagnosis (Ramaswamy et al., 2013) with the points of event-free and overall survival being on the opposing sides of this common time cut-off. This means that there is a possibility that overall survival is not directly comparable with progression free survival outcomes.

5.5.4. Gene function of identified survival methylation biomarkers.

5.5.4.1. *MYO7A is associated with cancerous growth and metastasis.* The *MYO7A* (Myosin VIIA) gene, which spans 120 kb and has 49 exons, is located on chromosome 11q13.5 and is a member of the unconventional myosins gene family. Myosins have similar structure and help transport molecules within cells. *MYO7A* is expressed in the inner ear and in the retina. Myosins interact with actin, a protein important for cell

movement and shape (Williams et al., 2011). Mutations in *MYO7A* cause syndromic (USH1B) (Weil et al., 1997) and non-syndromic (DFNB2 and DFNA11) (Tamagawa et al., 2002) deafness in humans. Mutations in *MYO7A* are also responsible for recessive deafness in shaker-1 mice (Gibson et al., 1995; Mburu et al., 1997). In addition, a lack of myosin VIIA in the retina is associated with retinitis pigmentosa, a condition in which light-sensing cells of the retina gradually deteriorate, resulting in progressive vision loss.

Structurally, the tail of *MYO7A* contains two FERM domains (F for 4.1 protein, E for ezrin, R for radixin and M for moesin). Some FERM-containing proteins play important roles in tumor development. Some of these inhibit tumor motility, such as FRMD5 (Wang et al., 2012), while others promote tumor progression such as ERM proteins (Gautreau et al., 2002). *MYO7A* variant S1666C (rs2276288) is associated with increased risk of malignant melanoma in a case-control study, suggesting that *MYO7A* might play important roles in melanoma development (Fernández-Aceñero et al., 2019) and it was demonstrated that *MYO7A* interacts with integrin $\beta 5$ and selectively promotes integrin $\beta 5$ -mediated cell adhesion and migration (Liu et al., 2018). A recent study also showed that silencing *MYO7A* by means of RNAi inhibited melanoma cell growth through upregulation of cell cycle regulator p21 (also known as *CDKN1A*), suppressed melanoma cell migration and invasion through downregulation of *RhoGDI2* (also known as *ARHGDIB*) and *MMP9*. *MYO7A* depletion suppressed melanoma cell metastases to the lungs, kidneys and bones in mice. In contrast, overexpression of *MYO7A* promoted melanoma xenograft growth and lung metastasis (Liu et al., 2018).

5.5.4.2. *TRIM72*. The *TRIM72* (mitsugumin 53 (MG53)) gene is a six exon gene located on chromosome 16p11.2 and is a member of the tripartite motif family and it is considered to be involved in the cell membrane repair system, maintenance of myocardial cells and linked to several signalling pathways (Lemckert et al., 2016). Recent reports have also suggested TRIM can induce insulin resistance and metabolic syndrome due to its E3 ligase activity and may be involved in oncogenesis through its ubiquitin ligase function. Chen et al. (2018) showed that *TRIM72* levels were significantly lower in colon cancer tissue than in normal tissue and that disease progression is associated with reduction of *TRIM72* expression. Fernández-Aceñero et al. (2019) and colleagues also reported an association between lower blood levels of *TRIM72*

and recurrence and overall dismal prognosis in stage II colon carcinoma, attributing this finding to the ubiquitin ligase activity of TRIM72 and its relation with several signalling pathways involved in tumorigenesis.

5.5.4.3. *Potential function of methylation of identified loci.* Methylation status of the identified prognostic loci did not correlate with the expression of the corresponding genes. Both identified loci are located within the gene body. While gene promoter region methylation is more typically associated with transcriptional regulation, the function of gene body methylation is not well understood. Evidence suggests that it could regulate splicing and suppress the activity of intragenic transcriptional units (Maunakea et al., 2010; Lev Maor et al., 2015).

5.5.5. Future work. The aims set in the introduction to this chapter have been fulfilled, but many aspects of the findings require further investigation. In particular, the methylation loci were identified using a relatively small standard-risk cohort, and validation was performed in cohorts that included samples with high risk features, which by definition was different from the discovery cohort. Moreover, metastatic stage behaved unusually in the Cavalli cohort. However, this was the only external cohort available at the time of the analysis. Therefore, it is important to further validate the findings first using an independent standard-risk non-WNT/non-SHH cohort and then in another independent mixed risk cohort with known status of each clinical risk factor for each sample and event-free survival time available in addition to the overall survival.

As the scheme, consisting of identified methylation markers, does not outperform the PNET4 signature prognostication scheme in standard risk non-WNT/non-SHH medulloblastoma, but shows promising results in the high risk cohort, it would be interesting to ask the same questions and recapitulate the analysis using the high-risk cohort as a discovery cohort. Additionally, it would be interesting to use the PNET4 signature prognostication scheme as a base model in order to identify additionally prognostic methylation biomarkers.

There is suitable high-risk medulloblastoma clinical trial (NCT04696029, however it is currently in the recruiting stage with estimated completion date March 2029).

5.5.6. Summary. This chapter has identified biomarkers (*MYO7A* and *TRIM72* methylation) that correlate with survival in standard risk non-WNT/non-SHH medulloblastoma, and went on to investigate the prognostic potential of the identified markers in high-risk and mixed-risk disease on their own as well as in combination with previously reported PNET4 signature prognostication scheme and other high-risk features.

In this chapter, it was demonstrated that the identified methylomic biomarkers have prognostic potential in high risk and mixed risk disease and have additional prognostic value in combination with PNET4 signature prognostication scheme. However, the need for further validation in larger independent disease cohorts was also identified.

Chapter 6: Discussion and conclusion

6.1. Introduction

Medulloblastoma is the most common paediatric brain tumour, accounting for about 20% of all childhood brain malignancies and responsible for approximately 650 new cases a year in the European Union (Pizer and Clifford, 2009). Current risk stratification is based on clinico-pathological and molecular disease correlates. For instance, children younger than 16 years of age at diagnosis with WNT-activated medulloblastomas consistently show favourable outcomes (5 year event free survival $> 95\%$), (Laninger et al., 2012; Edelstein et al., 2011), whereas other disease features, including *MYC* or *MYCN* amplification, large cell/anaplastic histology, metastatic disease, or subtotal resection, indicate high risk disease (5 year event free survival $< 60\%$) (Palmer et al., 2010).

However, even though in general these features are considered high-risk, in some molecularly-defined groups of patients, the presence of these markers does not indicate a poor outcome and in others, absence of high-risk markers does not guarantee favourable outcomes. In fact, about 20% of standard risk patients, that is, in the absence of currently understood risk factors, still relapse and die, and most survivors are left with long-term disabilities as a consequence of their disease and treatment (Edelstein et al., 2011; Palmer et al., 2010), see also Section 1.3.13. It is thought that improved prognostication could enable therapy intensification in high-risk patients and therapy de-escalation to improve quality of survivorship in lower-risk patients. This means that among the groups of patients who are generally considered as high-risk, it may be possible to identify subsets of patients that are not actually high-risk and who may be eligible for therapy de-escalation and to determine additional, currently unknown, biomarkers within standard-risk disease that will lead to improved survival of these patients by intensifying their treatment where necessary.

Specific biomarkers found alongside classical clinical markers such as histopathology or extent of resection represent additional and important prognostic factors, since

biomarker identification and validation can lead to targeted therapy and more personalised cancer treatment.

There are now ongoing biomarker driven clinical trials (SIOP PNET 5 MB and SJMB12, also known as NCT02066220 and NCT01878617 trials respectively), which investigate the possibility of improving quality of life for favourable-risk patients by reducing intensity of therapies and improving survival rates by using risk-adapted therapies in the remaining patients. Another important example of the use of biomarkers in medulloblastoma management is that the prognosis of SHH-activated medulloblastoma is related to patient age; additionally, children with *TP53* mutation have poor outcomes (Northcott et al., 2011; Taylor et al., 2012; Zhukova et al., 2013). Ongoing clinical trials, such as NCT03434262, NCT01601184, NCT00939484 and NCT01239316, are aimed at evaluating SHH pathway inhibitors at diagnosis or for treating recurrent SHH-activated tumours or tumours refractory to conventional treatments.

Integration of genomic and epigenomic data is an increasingly widespread way of getting a better understanding of the complex mechanisms that underlie the development and progression of cancer. Methylation arrays allow the simultaneous measurement of both genomic and epigenomic information, reducing costs, processing times and amount of patient DNA material used in the analysis.

This project investigated the potential of using Illumina Human Methylation arrays to routinely retrieve genomic information, in particular to estimate genomic copy number of patient DNA samples, including historical, poor quality DNA samples archived as FFPE material. In chapter 3, a bioinformatics workflow was described that enabled integrated genomic and epigenomic data analysis from the same patient sample using a single platform that incorporated methods for assessing disease-specific copy number changes, both large-scale and focal, as well as methods for quality control that would enable robust copy number analysis using Human Methylation arrays for clinical material archived as FFPE tissue.

In chapter 4, the novel methods of copy number estimation developed in chapter 3 were applied to assess the prognostic value of a previously-published prognostication scheme defined by a combination of clinico-pathological and cytogenetic biomarkers (Shih et al., 2014) and developed using genotyping arrays (Affymetrix SNP6.0). In addition, a chromosomal signature for risk assignment derived from patients enrolled

in the PNET4 clinical trial of standard-risk medulloblastoma Goschzik et al. (2018) was validated and additionally assessed in a cohort of high-risk patients.

Finally, in chapter 5, the utility of DNA methylation biomarkers was investigated for prognostication within standard-risk non-WNT/non-SHH medulloblastoma. Using the GLMnet algorithm (Simon et al., 2011), two potentially prognostic methylomic biomarkers were identified and tested in a high-risk and external, mixed-risk cohort, in combination with each other and in combination with the PNET4 signature prognostication scheme validated in chapter 4 (Goschzik et al., 2018). A prognostication scheme was derived from the resulting Cox models that combined PNET4 signature scheme with identified methylomic markers and metastatic stage to stratify the patients into low, standard and high risk groups.

6.2. CN analysis workflow has potential for integrated array analysis

Chapter 3 demonstrated that Illumina 450k methylation microarrays are well suited to the estimation of genomic copy number changes in medulloblastoma.

A bioinformatics workflow was developed, containing two pipelines, focused on large-scale (arm-level) CNA and a focal, locus-specific pipeline. Each of the pipelines was extensible and could flexibly incorporate alternative methods for CNA analysis. In addition to the CNA analysis methods, the workflow includes methods for data pre-processing, sample quality control and copy number profile assembly.

6.2.0.1. *Advantages and limitations.* The workflow is modular, consisting of three main modules (stages) each of which can be executed separately. Each of the stages operates in a standalone manner and can therefore be extended with new bioinformatics tools and methods or adapted for use with new array platforms as technologies evolve.

The second stage consisted of two pipelines, designed for the identification of large-scale (arm-level) and focal locus-specific CNA pipelines, in addition to purposely developed in this project methods, was successfully extended with an alternative pipeline that uses the new bioinformatics tool *conumee* (see section 2.7). The pipeline is able to use the output of stage one and its output can be passed directly onto the third stage of the workflow.

Both the original and *conumee* based implementations of methods used in the large-scale pipeline allowed automatic detection of arm-level genomic copy number aberrations for both gains and losses.

In case of the focal, locus-specific pipeline, the original pipeline based on Method I operates only in manual mode - each sample has to be curated by inspection of the graphical output. However, the additional pipeline based on *conumee* is able to identify focal high-level amplifications in a semi-automatic manner, flagging up putative amplifications for manual curation. The major limitation of the focal detection of locus-specific changes pipeline is that it is unable to detect focal deletions due to issues with sensitivity.

In summary, the workflow provides an automated CNA analysis within modular stages that allows non-bioinformaticians with limited programming experience to assess aneuploidy and focal changes. The analysis was validated by assessing concordance between SNP6.0 and 450k-derived copy number calls, providing a proof of concept

that methylation arrays can be applied to CN analysis with high accuracy, which has potential to speed up classification and biomarker identification whilst reducing diagnostic costs, since methylation arrays are currently the gold standard for molecular subgroup assignment.

In addition, the methodology described in chapter 3 is not confined to analysis of medulloblastoma copy number aberrations. For example, it was successfully used for estimation of the copy number status of the *PAX5* gene in Acute Lymphoblastic Leukaemia samples. The results, including plots output produced using the methods developed for this project, were published in Gabriel et al. (2015), also attached in B.

6.3. Whole chromosomal aberration signature has utility for rapid disease risk stratification in clinical settings

The pre-treatment risk assessment of medulloblastoma patients is becoming fundamental to the assignment of treatment strategies. In chapter 4, the CN analysis workflow developed for DNA methylation microarrays was applied to assess two previously reported prognostication schemes based on clinico-pathological and cytogenetic markers.

The first part of chapter 4 addressed the validation of the prognostication scheme by Shih et al. (2014) originally developed using gold-standard genotyping arrays. The bioinformatics workflow developed in the chapter 3 was applied to three medulloblastoma cohorts consisting of single subgroup cohorts: SHH, Group 3 and Group 4, defined as per the consensus molecular subgrouping paper (Taylor et al., 2012). Samples were stratified according to the Shih et al. scheme. Results from the stratification of Group 3 and Group 4 were consistent with Shih et al's findings, however for the SHH subgroup, results did not show statistically significant differences in survival. Inconsistency in the results is likely due to a combination of factors, including the discrepancies in demographic differences between the published and validation cohorts, such as paucity of tumour samples with high-risk features and unavailability of samples from adult patients in the validation cohort.

In the second part of the chapter, a prognostication scheme, defined by PNET4 signature (at least two of chromosome 7 gain, chromosome 8 loss, and chromosome 11 loss (Goschzik et al., 2018)), was first validated in non-trial, standard-risk non-WNT/non-SHH disease. Then it was applied to the high-risk non-WNT/non-SHH

where it was able to identify a low-risk subset with survival rates significantly exceeding the rest of the cohort (87.5% vs 47% 5-years survival).

In summary, both schemes were successfully validated, and therefore, it was shown that DNA methylation microarrays and the methods developed in the chapter 3 are well suited for genomic CNA detection. Moreover, it was demonstrated that the process of CNA detection can be automated, which gives the methodology an advantage over commonly used methods such as FISH and MLPA, which do not provide genome-wide outputs and, moreover, are labour intensive.

6.4. Methyloomic signature has prognostic potential in non-WNT/non-SHH medulloblastoma with limited availability of known clinical risk markers

In chapter 5, the Illumina HumanMethylation450k microarray was used to identify prognostic methyloomic markers in standard-risk non-WNT/non-SHH medulloblastoma. A standard-risk medulloblastoma cohort (SR-NMB, n=75), the same cohort used in chapter 4 to validate the PNET4 cytogenetic prognostic signature) was tested. A two-step filtering process was applied to identify methylated loci with a large change in methylation and a bimodal distribution - after filtering, 402 potentially prognostic probes were identified as described in sections 2.13 and 2.14. The GLMnet algorithm (section 2.15) was applied to these loci, which identified three cross-validated potentially prognostic loci, of which two, *MYO7A* and *TRIM72* significantly correlated with survival in both univariable and multivariable Cox proportional hazard models. Further investigation revealed that despite showing significant survival differences when used in survival models separately, a favourable prognosis was associated with methylation of both probes, whereas an absence of methylation of one probe or the other conferred a dismal outcome.

Subsequently, the identified methylation markers were added to the PNET4-derived cytogenetic prognostic signature, described in chapter 4, as a covariate and their addition showed additional prognostic value.

The identified methylation markers were then applied to the high-risk non-WNT/non-SHH cohort (HR-NMB, n=100, as used in chapter 4) and to an external, independent, mixed-risk cohort (Cavalli et al., 2017).

The methylation markers were significantly associated with survival in the non-metastatic high-risk non-WNT/non-SHH and mixed-risk non-WNT/non-SHH medulloblastoma cohorts, but not for metastatic disease. These findings would require validation in larger independent cohorts, but nevertheless, this research demonstrates that methylation arrays are suitable for integrated genomic and epigenomic analysis.

6.5. Risk stratification in non-WNT/non-SHH disease

It is now widely accepted now that medulloblastoma is a highly heterogeneous disease with a wide range of survival outcomes. The 2012 international consensus on MB subgroups reported four distinct subgroups of the disease: WNT, SHH, Group 3 and Group 4 (Taylor et al., 2012; Ramaswamy et al., 2016b). Since publication of this consensus, the clinical and biological relevance of the medulloblastoma subgroups has been extensively studied. These advances recently led to recognition of medulloblastoma subgroups in WHO classification of CNS as four molecular variants of the disease: WNT, SHH-*TP53*-wild-type, SHH-*TP53*-mutant, and non-WNT/non-SHH. The non-WNT/non-SHH subgroup represents Group 3 and Group 4 of the consensus molecular variants (Louis et al., 2016; Ramaswamy et al., 2016b).

WNT and SHH subgroups are well defined having distinct activating pathways. WNT subgroup patients under 16 typically exhibit very good survival (over 90%), with less clear prognosis for the adults (Goschzik et al., 2018; Clifford et al., 2006; Ellison et al., 2011). SHH patients show more heterogeneous outcome associated with genetic features e.g. *TP53* mutations, histopathology and age at diagnosis (Schwalbe et al., 2017b; Goschzik et al., 2018; Zhukova et al., 2013). Group 3 and Group 4 represent the majority of all medulloblastoma cases - about 65% with heterogeneous clinicopathological features and widely differing survival outcomes which partially are associated with known clinical high-risk factors (Pizer and Clifford, 2009). Many patients of this subgroups relapse in the absence of these risk factors, which shows further molecular heterogeneity within these tumours (Goschzik et al., 2018; Taylor et al., 2012).

Recently, multiple independent research groups have reported further intra-subgroup biological and clinical heterogeneity of non-WNT/non-SHH subgroup with inconsistent definition and numbers of subtypes due to differences in analytical methods and patients cohorts. An international meta-analysis of Group 3 and Group 4 conducted by Sharma et al. (2019), that aimed to reconcile the discrepancies, most strongly supported

a definition of Group 3/Group 4 comprising eight subtypes (I-VIII, with evidence of subtype VII comprising additional subtypes VII-A and VII-B).

Interestingly, subtypes I, V and VII showed mixed Group 3 and Group 4 composition and therefore highlighted the interrelationship between the two, indicating that it might be beneficial taking into account their shared biology when developing novel prognostication schemes. Research conducted by Schwalbe et al. (2017b) showed that combining standard-risk subsets of the two subgroups led to a novel risk-stratification scheme that outperformed previous risk stratification models (Goschzik et al., 2018; Menyhárt et al., 2019). This scheme is defined by easily identifiable in clinical settings whole chromosome aberration signature: at least two of chromosome 7 gain, chromosome 8 loss and chromosome 11 loss. The signature identified favourable risk subset of patients with 100% 5 year PFS survival in the discovery cohort and 94.7% (95% CI 57.5-82.7) in the validation cohort. Research conducted in the chapter 4 not only confirmed these findings, but also showed that this cytogenetic signature was able to identify a favourable risk subset of patients in high-risk non-WNT/non-SHH medulloblastoma with 87.5% (95% CI 67.3-100) 5-years PFS survival; the survival of the remainder of the cohort was 47% (95% CI 37.7-59.7).

The work described in chapter 5 identified two methylomic markers that, combined together, were prognostic in standard-risk and non-metastatic high-risk non-WNT/non-SHH disease, that identified a standard-risk subset with 76.8% (95% CI 65.2-90.53) and 84% (95% CI 68.8-100) survival in standard-risk and high-risk cohorts respectively. However, neither of the methylomic markers were prognostic in the metastatic high-risk cohort.

Further investigation in chapter 5 showed that methylomic markers were prognostic when considered in conjunction with the PNET4 cytogenetic prognostic signature, with 5-year survival being 100% in the subset defined by combination of methylomic signature (both methylomic markers) and PNET4 signature as covariates in both standard-risk and non-metastatic high-risk NMB cohorts.

A novel risk stratification scheme based on the findings in chapters 4 and 5 was developed and described in chapter 5. This scheme was applicable to both standard- and high-risk disease, and, therefore, was suitable for mixed cohorts or for cohorts with incomplete clinical and molecular data. In this scheme, patients were assigned to

three risk groups: low-risk, standard-risk and high-risk. Patients with non-metastatic disease, and positive for both PNET4 and methylomic signatures were assigned to a low-risk subset (100%, 95% CI (100-10)), non-metastatic patients positive for only methylomic or PNET4 signature were assigned to a standard-risk subset (79.5%, 95% CI (61.1-100)) and patients who were negative for both signatures or metastatic were assigned as high-risk (43.2%, 95% CI (33.06-56.3)).

The proposed combined risk stratification scheme performed well in standard-risk and high-risk cohorts, outperforming both PNET4 scheme and the methylomic signature in the high-risk cohort. In fact, this scheme reclassified 21% of high-risk patients to a favourable-risk category, showing that for high-risk disease, addition of methylomic markers might play an important role. However, when validated in an external, independent, mixed-risk cohort (Cavalli et al., 2017), the scheme performed less well - it outperformed the methylomic signature in isolation, but demonstrated similar performance to the PNET4 scheme. This could be attributed to the unusual behaviour of metastatic disease in the non-WNT/non-SHH Cavalli cohort, where metastatic status was not associated with survival. This is unusual, since previous research showed that metastatic status is prognostic in non-WNT/non-SHH disease as a whole and in Group 3 and Group 4 separately (Shih et al., 2014).

6.6. Future work

The methodology described in chapter 3 offers a rapid, robust and cost-effective way to perform integrated genomic and epigenomic analysis on the same sample. It was developed for use with Illumina HumanMethylation 450K microarrays (containing 450,000 probes). These arrays have now been discontinued and supplanted with Illumina Human Infinium MethylationEPIC microarrays, which contain 850,000 probes. Therefore, additional pipelines that are applicable to current EPIC arrays would contemporise the workflow. Due to the modular structure of the workflow, the addition of EPIC array processing methods is straightforward. In fact, some amendments have been made since the initial analysis in the pre-processing stage as some Minfi package functions for 450K array processing have been deprecated and therefore had to be replaced with newer functions that are compatible with both 450K and EPIC arrays, in order to enable some additional analysis carried out at a later stage of this research project.

The modular workflow may serve as a prototype for a modular bioinformatics software suite, consisting of methods designed to process, analyse and display information extracted from methylation arrays, united by an application programming interface (API), that would define interactions between the modules and their constituent components. The software could potentially also have a graphic user interface for users who are unfamiliar with scripting. This would allow fully integrated "one-stop-shop"-style genomic and methylomic analysis based on methylation arrays.

Potential components of the software, apart from the above-mentioned separate CNA analysis pipelines for current EPIC arrays and previously-run 450K arrays, could include methylomic analysis pipelines and additional pipelines for non-human genome methylation analysis pipelines that would enable integrated analysis of non-human models of cancer. Recently, Illumina have introduced their new Infinium Mouse Methylation (Illumina, 2020), with over 285k markers across the methylome for high-resolution epigenetic analyses of diverse murine strains. It may be useful to add a murine methylation analysis pipeline, as well as both human and non-human methylation analysis modules that would utilise methylation sequencing platforms such as enzymatic methyl-sequencing (EM-seq). In particular, TrueSeq Methyl Capture EPIC kit (Illumina, 2021) would be a practical addition, as this is the indented replacement for the arrays used in the thesis. However, as TrueSeq Methyl Capture EPIC kit is unsuitable for FFPE samples by design (Illumina, 2021), bisulphite sequencing, suitable for FFPE archived material, could be introduced into the suite of methods offered by the software suite to further enhance its functionality. The later would enable detection of copy neutral loss-of-heterozygosity.

The role of the methylomic markers identified in chapter 5 in tumorigenesis in general and in medulloblastoma in particular is unknown. Further analysis showed that the methylation status of the identified prognostic loci did not correlate with the expression of the corresponding genes; both identified loci are located within the gene body. While gene promoter region methylation is more typically associated with transcriptional regulation, the function of gene body methylation is not well understood, it is suggested that it could regulate splicing and suppress the activity of intragenic transcriptional units (Lev Maor et al., 2015; Maunakea et al., 2010). It would be interesting to investigate this possibility.

While low-risk subsets in both HR-NMB and Cavalli cohorts defined by the scheme are associated with the lower-risk subtypes identified by both Northcott et al. (2017) and Schwalbe et al. (2017b) classifications, high-risk and standard-risk subsets contain a mixture of higher-risk and lower-risk subtypes. This implies that the scheme can be further refined, and that composition by subtypes should be considered. High-risk and standard-risk subsets are both enriched with high risk subtype VIII samples. This subtype mostly corresponds to Group4-HR as defined by Schwalbe et al. Subtype VIII shows a cytogenetically relatively balanced genome with the only common events being iso17q and loss of chromosome X. This could indicate that potential prognostic biomarkers may be present amongst differentially methylated loci.

An interesting observation was made that in the PNET4 signature positive subset of the high-risk NMB cohort (chapter 4), 5 out of 11 samples (45%) were metastatic. Also, the survival of the subset of high-risk NMB cohort positive for both PNET4 and methylomic signatures, that contained both metastatic and non-metastatic samples, reached 100% (Figure 5.8) despite the fact that methylomic signature was not prognostic on its own in this cohort. This might indicate that the PNET4 signature on its own is prognostic in metastatic non-WNT/non-SHH medulloblastoma, and therefore, metastatic samples that are PNET4 signature positive should be re-assigned to standard-risk rather than high-risk group in the novel prognostication scheme and the scheme would have to be updated to reflect that. However, the number of samples in this cohort is too small to make a definitive conclusion at this time and therefore, validation in a bigger cohort is needed.

It would be beneficial to perform methylomic analysis similar to the one conducted in the chapter 5 using a high-risk or mixed-risk cohort instead of a standard-risk cohort. This might lead to a discovery of prognostic methylomic markers that would be prognostic in high-risk medulloblastoma and particularly in metastatic disease. Another option could be perform a search for additional methylomic biomarkers using a base survival model that included M-stage and PNET4 signature as known prognostic covariates.

6.6.1. Conclusion. The research performed during this project demonstrates the potential for routine cytogenetic assessment concurrent with molecular sub-classification using DNA methylation microarrays. Additionally, the integrated genetic and epigenetic stratification from a single platform enabled a more refined prognostication and

the identification of a subset of patients, currently classified as high risk, who demonstrate improved outcomes and who may be eligible for reduced intensity treatments that would offer a better quality of life as brain tumour survivors.

APPENDIX A

Cohorts

SampleID	p17	q17	Subgroup	Sex	Age	M-Stage	Pathology
NMB109	Loss	gain	Grp4	M	8	M1	CLA
NMB110	Loss	gain	Grp4	F	10	M1	CLA
NMB111	Neutral	neutral	Grp4	M	11	M0	CLA
NMB126	Loss	gain	Grp4	F	8.4		CLA
NMB142	Neutral	gain	Grp4	M	12.06	M2	CLA
NMB144	Loss	neutral	Grp4	M	5.14	M3	CLA
NMB145	Neutral	neutral	Grp3	F	7.01	M0/1	LCA
NMB145p	Neutral	neutral			NA		
NMB146	Loss	gain	Grp4	F	5.6	M0	CLA
NMB148	Neutral	neutral	Grp3	M	1.54	M0/1	CLA
NMB149	Loss	gain	Grp4	M	6.33		LCA
NMB151	Loss	gain	Grp4	M	9.82	M3	CLA
NMB166	Loss	gain	Grp4	F	9.72	M0	CLA
NMB167	Gain	gain	Grp4	F	6.51	M0	CLA
NMB17	Neutral	neutral	Grp3	M	4.7	M0/1	CLA
NMB171	Neutral	neutral	Grp3	M	7.71	M0/1	CLA
NMB172	Gain	gain	Grp4	M	10.81		CLA
NMB173	Neutral	gain	Grp4	M	9.17	M0	CLA
NMB178	Loss	gain	Grp4	M	4.95	M3	CLA
NMB179	Neutral	neutral	Grp4	M	2.43	M0	CLA
NMB180	Gain	gain	Grp4	M	10.66	M0	DN
NMB183	Loss	gain	Grp4	F	11.6	M3	CLA
NMB185	Gain	gain	Grp4	M	9.65	M0	CLA
NMB186	Neutral	gain	Grp4	M	3.73	M1	CLA
NMB187	Loss	neutral	Grp4	M	3.7	M0	CLA
NMB188	Neutral	neutral	Grp3	F	8.56		CLA
NMB189	Neutral	neutral	Grp4	M	11.36	M0/1	CLA
NMB190	Loss	neutral	Grp4	M	11.71	M0	CLA
NMB227	Gain	gain	Grp4	M	4.58	M0	CLA
NMB227	Gain	gain	Grp4	M	4.58	M0	CLA

SampleID	p17	q17	Subgroup	Sex	Age	M-Stage	Pathology
NMB250	Neutral	gain	Grp4	M	4.8	M0	CLA
NMB255	Loss	gain	Grp4	M	10.26	M0/1	CLA
NMB257	Loss	gain	Grp4	M	10.14	M3	CLA
NMB259	Loss	gain	Grp4	M	13.32	M2	CLA
NMB262	Neutral	neutral	Grp4	F	13.47	M0	NOS
NMB266	Loss	gain	Grp4	M	7.89	M0	CLA
NMB267	Neutral	gain	Grp4	M	5.94	M0	LCA
NMB269	Neutral	gain	Grp3	M	7.61	M3	CLA
NMB273	Loss	gain	Grp3	M	5.72	M3	CLA
NMB277	Neutral	gain	Grp3	F	15.32	M3	NOS
NMB283	Neutral	gain	Grp4	M	7.62	M0	LCA
NMB316	Neutral	neutral	Grp4	M	9.61	M2	CLA
NMB318	Neutral	neutral	Grp3	M	5.91	M3	LCA
NMB346	Loss	gain	Grp4	M	7.75	M0	CLA
NMB357	Neutral	gain	Grp4	M	3.5	M1	CLA
NMB358	Loss	gain	Grp4	M	13.85	M0	CLA
NMB361	Loss	gain	Grp3	F	2.06	M3	LCA
NMB362	Loss	gain	Grp4	M	7.03	M0	DN
NMB365	Loss	gain	Grp4	M	4.71	M0	CLA
NMB366	Loss	gain	Grp4	M	6.28		DN
NMB368	Loss	gain	Grp4	M	12	M1	NOS
NMB369	Neutral	gain	Grp4	M	9.24	M0	NOS
NMB373	Neutral	neutral	Grp4	M	5.25	M2	CLA
NMB374	Neutral	neutral	Grp3	F	0.64	M0	CLA
NMB378	Neutral	neutral	Grp3	M	2.49	M0	CLA
NMB381	Gain	gain	Grp3	M	4.65	M3	NOS
NMB383	Loss	gain	Grp4	M	15.65	M0/1	NOS
NMB384	Neutral	neutral	Grp3	M	4.57	M0	NOS
NMB385	Gain	gain	Grp4	M	9.82	M0	CLA
NMB387	Neutral	gain	Grp4	M	6.76	M0	NOS

SampleID	p17	q17	Subgroup	Sex	Age	M-Stage	Pathology
NMB388	Neutral	gain	Grp3	M	11.92	M0	CLA
NMB393	Loss	gain	Grp4	M	11.51	M0	CLA
NMB398	Neutral	neutral	Grp4	M	5.49	M1	NOS
NMB400	Neutral	neutral			NA		
NMB401	Loss	gain	Grp4	M	5.8	M3	CLA
NMB403	Loss	gain	Grp4	M	6	M3	NOS
NMB405	Neutral	gain	Grp3	M	15.97	M1	LCA
NMB407	Neutral	gain	Grp4	F	6.34	M0	CLA
NMB410	Loss	gain	Grp4	M	15.82	M0	CLA
NMB412	Loss	gain			NA		
NMB415	Neutral	neutral	Grp4	F	7.93	M0	DN
NMB416	Gain	gain	Grp4	F	4.64	M0	CLA
NMB418	Neutral	gain	Grp4	M	6.44	M3	CLA
NMB419	Loss	gain	Grp4	F	6.01	M0	CLA
NMB420	Neutral	neutral	Grp3	M	1.43	M3	CLA
NMB422	Neutral	neutral	Grp3	M	3.74	M0	CLA
NMB43	Loss	gain	Grp4	M	9.96	M2	CLA
NMB433	Neutral	neutral	Grp3	M	1.89	M0	CLA
NMB438	Neutral	gain	Grp4	M	5.01	M0	CLA
NMB440	Neutral	neutral	Grp3	M	5.67	M3	LCA
NMB457	Gain	gain	Grp4	F	4.71	M0	CLA
NMB459	Neutral	neutral			NA		
NMB463	Neutral	gain	Grp4	M	7.08	M0	LCA
NMB51	Gain	gain	Grp4	M	6.76	M0	CLA
NMB52	Loss	gain	Grp4	F	8.58	M0	CLA
NMB529	Loss	gain	Grp4	F	8.11	M0	LCA
NMB535	Loss	gain	Grp3	M	1.97	M3	LCA
NMB536	Loss	neutral	Grp3	M	8.46	M0	CLA
NMB550	Loss	gain	Grp4	M	5.83	M0	CLA
NMB610	Neutral	neutral	Grp3	M	8.88	M0	CLA

SampleID	p17	q17	Subgroup	Sex	Age	M-Stage	Pathology
NMB648	Neutral	gain	WNT	F	6.08	M0	CLA
NMB69	Loss	gain	Grp4	M	7.75	M0/1	CLA
NMB70	Neutral	neutral	Grp3	F	3.29	M0/1	CLA
NMB725	Loss	gain	Grp4	M	5.75	M3	CLA
NMB733	Gain	gain	Grp4	F	6.67	M2	CLA
NMB734	Neutral	neutral	Grp4	M	4.24	M3	CLA
NMB735	Loss	gain	Grp4	M	15.92	M0	CLA
NMB736	Neutral	gain	Grp4	M	6.94	M2	CLA
NMB737	Loss	gain	Grp4	M	7.59	M3	CLA
NMB739	Loss	gain	Grp4	M	11.42	M2	CLA
NMB741	Loss	gain	Grp3	M	2.94	M0	LCA
NMB755	Neutral	neutral	Grp3	M	6.38	M2	CLA
NMB760	Loss	gain	Grp3	M	2.96	M3	CLA
NMB762	Loss	gain	Grp3	M	0.82	M3	CLA
NMB766	Neutral	neutral	Grp3	F	3.57	M3	NOS
NMB768	Neutral	neutral	Grp3	F	9.31	M1	CLA
NMB769	Neutral	gain	Grp3	M	4.01	M3	LCA
NMB77	Loss	gain	Grp4	F	8.54	M3	CLA
NMB771	Loss	gain	Grp4	M	15.43	M0	CLA
NMB774	Neutral	gain	Grp3	M	5.89	M0	CLA
NMB782	Neutral	gain	Grp4	M	5.44	M0	CLA
NMB785	Loss	gain	Grp4	M	8.44	M0	CLA
NMB786	Neutral	neutral	Grp3	M	10.06	M3	CLA
NMB787	Loss	gain	Grp4	F	13.54	M0	CLA
NMB795	Loss	gain	Grp3	M	5.43	M0	LCA
NMB799	Neutral	neutral	Grp3	F	2.46		CLA
NMB801	Loss	gain	Grp4	M	10.46		LCA
NMB804	Neutral	gain	Grp3	M	2.8	M1	CLA
NMB807	Loss	gain	Grp3	M	4.42	M0	CLA
NMB812	Loss	neutral	Grp3	M	4.12	M0	CLA

SampleID	p17	q17	Subgroup	Sex	Age	M-Stage	Pathology
NMB82	Loss	gain	Grp4	M	5.35	M0	CLA
NMB832	Neutral	gain	Grp4	M	7.39		CLA
NMB859	Neutral	neutral	Grp4	M	10.64	M1	CLA
NMB867	Neutral	gain	Grp4	M	6.56	M1	CLA
NMB868	Neutral	neutral	Grp3	F	2.74	M0	CLA
NMB872	Neutral	neutral	Grp3	M	9.08		NOS
NMB889	Neutral	gain	Grp4	F	8.82	M0	CLA
NMB890	Neutral	gain	Grp4	M	7.83		CLA
NMB892	Neutral	neutral	Grp3	M	4.94		LCA
NMB897	Loss	gain	Grp4	M	14.77	M0	CLA
NMB898-2	Neutral	gain	Grp4	M	8.7		CLA
NMB907-2	Neutral	neutral	Grp4	F	7.25		CLA
NMB908-2	Neutral	gain	Grp4	F	10.91		CLA
NMB910-2	Neutral	gain	Grp4	M	3.61		CLA

TABLE A.1. Clinical demographic of the arm-level pipeline test cohort used for chapter 3. Cohort consisted of 134 primary medulloblastoma tumours from Newcastle Medulloblastoma Archive. Molecular subgroup, gender, age, metastatic stage and pathology are shown. Histological subtype is coded as follows: CLAS- classic; LCA- large cell/anaplastic; DN- desmoplastic/nodular; MBEN- MB with extensive nodularity; NOS-not otherwise specified.

Sample ID	Subgroup	Sex	Age	Stage	Pathology
NMB109	Grp4	M	8	M1	CLA
NMB110	Grp4	F	10	M1	CLA
NMB111	Grp4	M	11	M0	CLA
NMB113	SHH	M	3	M0	CLA
NMB115	WNT	M	14.81	M1	CLA
NMB125	Grp4	F	5.99	M1	CLA
NMB126	Grp4	F	8.4		CLA
NMB127	Grp3	M	13.55	M1	LCA
NMB129			NA		
NMB130	Grp4	M	7.46	M0	CLA
NMB131	WNT	M	10.38	M0	CLA
NMB132	Grp4	M	9.34	M0	CLA
NMB138	SHH	M	3.3	M0	CLA
NMB139	WNT	M	12.68	M0	CLA
NMB141	SHH	F	11.71	M0	CLA
NMB145	Grp3	F	7.01	M0/1	LCA
NMB146	Grp4	F	5.6	M0	CLA
NMB149	Grp4	M	6.33		LCA
NMB153	Grp3	M	3.34		CLA
NMB160			NA		
NMB162			NA		
NMB165			NA		
NMB166	Grp4	F	9.72	M0	CLA
NMB167	Grp4	F	6.51	M0	CLA
NMB171	Grp3	M	7.71	M0/1	CLA
NMB172	Grp4	M	10.81		CLA
NMB173	Grp4	M	9.17	M0	CLA
NMB176	Grp3	F	4.31	M0	CLA
NMB180	Grp4	M	10.66	M0	DN
NMB181	SHH	M	8.41	M0	LCA

Sample ID	Subgroup	Sex	Age	Stage	Pathology
NMB182	MBNOS	M	9.64	M0	CLA
NMB184	Grp3	M	8.57	M0	NOS
NMB185	Grp4	M	9.65	M0	CLA
NMB186	Grp4	M	3.73	M1	CLA
NMB191	WNT	M	14.39	M0	CLA
NMB203	Grp4	M	6.24	M0	CLA
NMB212			NA		
NMB227	Grp4	M	4.58	M0	CLA
NMB250	Grp4	M	4.8	M0	CLA
NMB258	WNT	M	11.98	M0	CLA
NMB260	Grp4	F	11.44	M0	CLA
NMB261	SHH	M	5.25	M0	NOS
NMB262	Grp4	F	13.47	M0	NOS
NMB264	WNT	F	5.14	M1	CLA
NMB266	Grp4	M	7.89	M0	CLA
NMB267	Grp4	M	5.94	M0	LCA
NMB278	SHH	M	7.83	M0	DN
NMB282			NA		
NMB283	Grp4	M	7.62	M0	LCA
NMB284	Grp4	F	11.14	M0	DN
NMB285	Grp4	F	8.73	M0	NOS
NMB320	Grp4	M	3.82		CLA
NMB330	Grp3	M	4.72	M0/1	CLA
NMB332			NA		
NMB335	Grp3	M	4.27	M0/1	LCA
NMB346	Grp4	M	7.75	M0	CLA
NMB347	SHH	M	9.85	M0	CLA
NMB357	Grp4	M	3.5	M1	CLA
NMB358	Grp4	M	13.85	M0	CLA
NMB359	WNT	F	10.98	M0	CLA

Sample ID	Subgroup	Sex	Age	Stage	Pathology
NMB362	Grp4	M	7.03	M0	DN
NMB364	SHH	F	3.71	M0	DN
NMB365	Grp4	M	4.71	M0	CLA
NMB366	Grp4	M	6.28		DN
NMB367	WNT	M	9.97	M0	CLA
NMB368	Grp4	M	12	M1	NOS
NMB370	SHH	F	6.46	M0	CLA
NMB377	Grp4	F	8.96	M0	CLA
NMB380	WNT	M	7.86	M1	NOS
NMB382	WNT	M	11.71	M0	NOS
NMB383	Grp4	M	15.65	M0/1	NOS
NMB384	Grp3	M	4.57	M0	NOS
NMB386	WNT	M	4.72	M0	CLA
NMB387	Grp4	M	6.76	M0	NOS
NMB388	Grp3	M	11.92	M0	CLA
NMB392			NA		
NMB393	Grp4	M	11.51	M0	CLA
NMB394	Grp4	M	5.71	M0	NOS
NMB398	Grp4	M	5.49	M1	NOS
NMB405	Grp3	M	15.97	M1	LCA
NMB406	SHH	M	7.73	M1	DN
NMB407	Grp4	F	6.34	M0	CLA
NMB409	WNT	F	8.96	M0	LCA
NMB410	Grp4	M	15.82	M0	CLA
NMB416	Grp4	F	4.64	M0	CLA
NMB417	WNT	M	8.69	M0	CLA
NMB419	Grp4	F	6.01	M0	CLA
NMB422	Grp3	M	3.74	M0	CLA
NMB436	WNT	M	4.97	M0	CLA
NMB438	Grp4	M	5.01	M0	CLA

Sample ID	Subgroup	Sex	Age	Stage	Pathology
NMB439	SHH	F	3.53	M0	LCA
NMB441	Grp3	F	3.17	M0	NOS
NMB445	Grp4	M	7.95	M0	DN
NMB446			NA		
NMB45			NA		
NMB452			NA		
NMB457	Grp4	F	4.71	M0	CLA
NMB46			NA		
NMB460	SHH	M	3.91	M0	DN
NMB462	SHH	M	7.63	M0	DN
NMB503			NA		
NMB504			NA		
NMB51	Grp4	M	6.76	M0	CLA
NMB517			NA		
NMB519	Grp3	M	4.22	M1	LCA
NMB52	Grp4	F	8.58	M0	CLA
NMB529	Grp4	F	8.11	M0	LCA
NMB532	Grp4	M	7.69	M0	CLA
NMB533	Grp3	F	6.79	M0	LCA
NMB536	Grp3	M	8.46	M0	CLA
NMB542	Grp4	M	6.69	M0	CLA
NMB544			NA		
NMB548	SHH	M	14.69	M0	LCA
NMB549	SHH	F	5.68	M0	LCA
NMB550	Grp4	M	5.83	M0	CLA
NMB585			NA		
NMB59			NA		
NMB60	Grp3	M	5.01	M1	CLA
NMB610	Grp3	M	8.88	M0	CLA
NMB620	Grp4	F	8.85	M0	CLA

Sample ID	Subgroup	Sex	Age	Stage	Pathology
NMB626	Grp3	F	3.97	M0	CLA
NMB63	SHH	M	11.45	M0	CLA
NMB632	Grp4	M	4.83	M0/1	CLA
NMB637	Grp3	M	3.25	M0	CLA
NMB638			NA		
NMB640			NA		
NMB642			NA		
NMB645	SHH	M	12.44	M0	NOS
NMB649	SHH	F	1.94	M0	MBEN
NMB659			NA		
NMB68	Grp4	F	6.86	M0	CLA
NMB69	Grp4	M	7.75	M0/1	CLA
NMB70	Grp3	F	3.29	M0/1	CLA
NMB706	SHH	M	10.79	M0	DN
NMB707	WNT	M	15.3	M0	DN
NMB708	WNT	F	14.28	M0	NOS
NMB709	Grp3	F	8.48	M0/1	NOS
NMB710	Grp4	F	9.31	M2	DN
NMB713			NA		
NMB715	Grp4	F	4.59	M0	NOS
NMB718	Grp4	M	6.91	M0	CLA
NMB723	Grp4	F	3.65	M0	DN
NMB726	SHH	M	3.08	M0	DN
NMB729	Grp3	M	6.91	M0	CLA
NMB730	SHH	M	11.06	M0	CLA
NMB732	WNT	M	6.1	M0	CLA
NMB735	Grp4	M	15.92	M0	CLA
NMB740	WNT	F	8.52	M0	CLA
NMB743	SHH	M	14.31	M0	LCA
NMB746			NA		

Sample ID	Subgroup	Sex	Age	Stage	Pathology
NMB754	Grp3	M	3.78	M0	CLA
NMB756	SHH	F	3.27	M0	DN
NMB757	WNT	F	7.63	M0	CLA
NMB758	Grp3	M	5.21	M1	LCA
NMB761	Grp4	M	11.72	M0	CLA
NMB763	SHH	F	8.29	M0	LCA
NMB764	Grp3	M	8.9	M0	CLA
NMB765	WNT	F	10.15		LCA
NMB767	Grp4	M	6.95		LCA
NMB768	Grp3	F	9.31	M1	CLA
NMB771	Grp4	M	15.43	M0	CLA
NMB774	Grp3	M	5.89	M0	CLA
NMB782	Grp4	M	5.44	M0	CLA
NMB784	Grp3	M	12.87	M0	CLA
NMB785	Grp4	M	8.44	M0	CLA
NMB787	Grp4	F	13.54	M0	CLA
NMB788	SHH	M	10.51	M0	LCA
NMB791	SHH	M	11.37	M0	CLA
NMB795	Grp3	M	5.43	M0	LCA
NMB80	Grp4	F	8.02	M0	NOS
NMB800	Grp4	M	10.22		CLA
NMB807	Grp3	M	4.42	M0	CLA
NMB810	SHH	F	8.42	M0	LCA
NMB812	Grp3	M	4.12	M0	CLA
NMB816	Grp3	F	4.08	M0	CLA
NMB818			NA		
NMB82	Grp4	M	5.35	M0	CLA
NMB820			NA		
NMB824			NA		
NMB825	Grp4	F	10.24	M0	CLA

Sample ID	Subgroup	Sex	Age	Stage	Pathology
NMB828			NA		
NMB831	WNT	F	7.74	M0	CLA
NMB832	Grp4	M	7.39		CLA
NMB857			NA		
NMB858	WNT	F	6.43	M0	CLA
NMB859	Grp4	M	10.64	M1	CLA
NMB866	SHH	M	8.02	M0	CLA
NMB867	Grp4	M	6.56	M1	CLA
NMB869	SHH	M	13.22	M0	CLA
NMB870	WNT	F	11.14	M0	CLA
NMB874	SHH	M	13.17	M0	NOS
NMB889	Grp4	F	8.82	M0	CLA
NMB890	Grp4	M	7.83		CLA
NMB891	Grp3	F	7.39		CLA
NMB894	Grp4	F	6.82		LCA
NMB897	Grp4	M	14.77	M0	CLA
NMB899	Grp3	M	4.5		NOS
NMB906			NA		
NMB915			NA		
NMB916			NA		
NMB93	WNT	M	10.78	M0	CLA
NMB94	WNT	F	9.41	M0	CLA

TABLE A.2. Clinical demographic of the test cohort used for chapter 3. Cohort consisted of 203 primary medulloblastoma tumours from Newcastle Medulloblastoma Archive. Molecular subgroup, gender, age, metastatic stage and pathology are shown. Histological subtype is coded as follows: CLAS- classic; LCA- large cell/anaplastic; DN- desmoplastic/nodular; MBEN- MB with extensive nodularity; NOS-not otherwise specified.

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB110	Grp4	F	child	10	M1	CLA	neutral	loss	gain	0	0	1.64	0	1.64	0
NMB111	Grp4	M	child	11	M0	CLA	neutral	neutral	neutral	0	0	2.52	0	2.52	0
NMB118	Grp4	F	child	11.85	M0	LCA	neutral	gain	gain	0	0	7.79	0	7.79	0
NMB119	Grp4	M	child	14.37	M0/1	CLA	neutral	neutral	gain	0	0	3.42	1	6.08	1
NMB123	SHH	M	child	13.51	M1	LCA	loss	loss	loss	0	0	2.82	1	3.71	1
NMB125	Grp4	F	child	5.99	M1	CLA	neutral	loss	gain	0	0	3.15	1	4.56	1
NMB127	Grp3	M	child	13.55	M1	LCA	gain	neutral	neutral	0	0	6.28	0	6.28	0
NMB128	SHH	M	child	4.52	M3	CLA	neutral	neutral	neutral	0	0	1.09	1	1.09	1
NMB130	Grp4	M	child	7.46	M0	CLA	neutral	loss	gain	0	0	2.37	1	4.42	1
NMB132	Grp4	M	child	9.34	M0	CLA	neutral	neutral	neutral	0	0	10.93	0	10.93	0
NMB134	Grp4	M	child	4.93	M3	CLA	neutral	loss	gain	0	0	0.66	1	1.12	1
NMB136	Grp4	M	child	10.5	M3	CLA	gain	loss	gain	0	0	0.72	0	0.72	1
NMB138	SHH	M	child	3.3	M0	CLA	neutral	neutral	neutral	0	0	0.79	1	0.8	1
NMB140	Grp4	M	child	9.37	M0	CLA	neutral	loss	gain	0	0	8.26	0	8.26	0
NMB141	SHH	F	child	11.71	M0	CLA	neutral	neutral	neutral	0	0	6.96	0	6.96	0
NMB142	Grp4	M	child	12.06	M2	CLA	neutral	neutral	gain	0	0	4.23	0	4.23	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB143	SHH	M	infant	2.84	M0	LCA	neutral	neutral	neutral	0	0	0.77	1	1.13	1
NMB144	Grp4	M	child	5.14	M3	CLA	neutral	loss	neutral	0	0	2	1	2.23	1
NMB145	Grp3	F	child	7.01	M0/1	LCA	loss	neutral	neutral	0	0	5.17	1	5.77	1
NMB146	Grp4	F	child	5.6	M0	CLA	neutral	loss	gain	0	0	9.36	0	9.36	0
NMB148	Grp3	M	infant	1.54	M0/1	CLA	neutral	neutral	neutral	0	0	0.31	1	0.32	1
NMB151	Grp4	M	child	9.82	M3	CLA	neutral	loss	gain	0	0	1.64	1	1.68	1
NMB152	Grp4	M	infant	0.49	M0	CLA	neutral	neutral	neutral	0	0	5.05	1	5.11	1
NMB153	Grp3	M	child	3.34	M0/1	CLA	neutral	loss	gain	0	0	13.17	0	13.17	0
NMB156	SHH	M	child	7.26	M0/1	DN	loss	loss	neutral	0	0	13.14	0	13.14	0
NMB166	Grp4	F	child	9.72	M0	CLA	neutral	loss	gain	0	1	9.38	0	9.38	0
NMB167	Grp4	F	child	6.51	M0	CLA	neutral	gain	gain	0	0	5.64	0	5.64	0
NMB168	SHH	F	child	9.77	M0	LCA	neutral	neutral	neutral	0	0	7.02	0	7.02	0
NMB169	Grp3	M	child	8.91	M0	LCA	neutral	loss	neutral	0	1	0.27	1	0.29	1
NMB17	Grp3	M	child	4.7	M0/1	CLA	neutral	neutral	neutral	0	0	19.78	0	19.78	0
NMB171	Grp3	M	child	7.71	M0/1	CLA	neutral	neutral	neutral	0	0	1.17	1	1.35	1
NMB173	Grp4	M	child	9.17	M0	CLA	neutral	neutral	gain	0	0	18.68	0	18.68	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB175	Grp3	F	child	7.1	M0	CLA	gain	neutral	neutral	0	0	7.87	0	7.87	0
NMB176	Grp3	F	child	4.31	M0	CLA	neutral	loss	gain	0	0	1.26	1	2.68	1
NMB177	Grp4	F	child	8.52	M0/1	CLA	neutral	loss	gain	0	0	7.3	0	7.3	0
NMB178	Grp4	M	child	4.95	M3	CLA	neutral	loss	gain	0	0	6.48	0	6.48	0
NMB179	Grp4	M	infant	2.43	M0	CLA	gain	neutral	neutral	0	0	7.43	0	7.43	0
NMB180	Grp4	M	child	10.66	M0	DN	neutral	gain	gain	0	0	8.91	1	9.32	1
NMB181	SHH	M	child	8.41	M0	LCA	neutral	neutral	gain	0	0	0.14	1	0.69	1
NMB183	Grp4	F	child	11.6	M3	CLA	neutral	loss	gain	0	0	4.28	1	5.3	1
NMB184	Grp3	M	child	8.57	M0	NOS	neutral	neutral	neutral	0	0	9.01	0	9.01	0
NMB185	Grp4	M	child	9.65	M0	CLA	neutral	gain	gain	0	0	9.92	0	9.92	0
NMB186	Grp4	M	child	3.73	M1	CLA	neutral	neutral	gain	0	0	3.59	1	5.57	1
NMB187	Grp4	M	child	3.7	M0	CLA	neutral	loss	neutral	0	0	9.18	0	9.18	0
NMB189	Grp4	M	child	11.36	M0/1	CLA	neutral	neutral	neutral	0	0	5.65	1	9.26	1
NMB190	Grp4	M	child	11.71	M0	CLA	neutral	loss	neutral	0	0	6.59	1	8.08	1
NMB199	Grp4	M	child	9.83	M3	CLA	neutral	loss	neutral	0	0	1.33	1	1.51	1
NMB2	Grp3	M	child	5.71	M0/1/3	CLA	neutral	gain	gain	0	1	0.63	1	0.63	1

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB200	SHH	F	infant	1.28	M0	MBEN	neutral	neutral	neutral	0	0	9.46	0	9.46	0
NMB203	Grp4	M	child	6.24	M0	CLA	neutral	neutral	neutral	0	0	6.7	0	6.7	0
NMB227	Grp4	M	child	4.58	M0	CLA	neutral	gain	gain	0	0	1.28	1	1.85	1
NMB250	Grp4	M	child	4.8	M0	CLA	neutral	neutral	gain	0	0	4.19	0	4.19	0
NMB251	Grp4	F	child	3.36	M0	CLA	neutral	neutral	neutral	0	0	3.44	0	3.44	0
NMB253	SHH	M	infant	0.24	M1	CLA	neutral	neutral	neutral	0	0	1.22	0	1.22	0
NMB254	SHH	F	infant	1.38	M0	DN	neutral	neutral	neutral	0	0	4.62	0	4.62	0
NMB255	Grp4	M	child	10.26	M0/1	CLA	neutral	loss	gain	0	0	2.31	1	3.82	1
NMB257	Grp4	M	child	10.14	M3	CLA	neutral	loss	gain	0	0	1.47	1	2.66	1
NMB259	Grp4	M	child	13.32	M2	CLA	neutral	loss	gain	0	0	4.71	0	4.71	0
NMB260	Grp4	F	child	11.44	M0	CLA	gain	loss	gain	0	0	4.62	0	4.62	0
NMB261	SHH	M	child	5.25	M0	NOS	neutral	neutral	gain	0	0	1.01	1	2.31	1
NMB262	Grp4	F	child	13.47	M0	NOS	neutral	neutral	neutral	0	0	4.6	0	4.6	0
NMB263	Grp4	F	child	8.73	M3	NOS	gain	neutral	gain	0	0	4.99	0	4.99	0
NMB266	Grp4	M	child	7.89	M0	CLA	neutral	loss	gain	0	0	4.62	0	4.62	0
NMB267	Grp4	M	child	5.94	M0	LCA	neutral	neutral	gain	0	0	4.19	0	4.19	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB269	Grp3	M	child	7.61	M3	CLA	gain	neutral	gain	0	0	1.61	1	2.24	1
NMB272	SHH	M	infant	0.73	M3	MBEN	neutral	neutral	neutral	0	0	4.57	0	4.57	0
NMB273	Grp3	M	child	5.72	M3	CLA	neutral	loss	gain	0	1	0	1	0.81	1
NMB277	Grp3	F	child	15.32	M3	NOS	gain	neutral	gain	0	0	3.12	1	3.7	1
NMB278	SHH	M	child	7.83	M0	DN	neutral	neutral	neutral	0	0	6.49	0	6.49	0
NMB281	Grp3	F	child	3.7	M3	CLA	neutral	neutral	gain	0	0	6.74	0	6.74	0
NMB283	Grp4	M	child	7.62	M0	LCA	neutral	neutral	gain	0	0	3.75	0	3.75	0
NMB284	Grp4	F	child	11.14	M0	DN	neutral	loss	gain	0	0	7.56	0	7.56	0
NMB285	Grp4	F	child	8.73	M0	NOS	neutral	loss	gain	0	0	6.99	0	6.99	0
NMB314	Grp4	M	child	6.02	M0	DN	neutral	neutral	gain	0	0	3.48	0	3.48	0
NMB316	Grp4	M	child	9.61	M2	CLA	neutral	neutral	neutral	0	0	2.07	1	2.86	1
NMB318	Grp3	M	child	5.91	M3	LCA	neutral	neutral	neutral	0	0	0.9	1	0.95	1
NMB32	SHH	F	infant	1.49	M0/1	DN	neutral	neutral	neutral	0	0	22.18	0	22.18	0
NMB325	Grp3	M	child	6.79	M0/1/3	CLA	neutral	neutral	neutral	0	0	2.2	1	12.76	1
NMB326	Grp3	F	infant	2.82		CLA	neutral	gain	gain	0	0	4.91	0	4.91	0
NMB330	Grp3	M	child	4.72	M0/1	CLA	gain	neutral	neutral	0	0	25.77	0	25.77	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB331	Grp3	M	infant	1.64	M0/1	CLA	neutral	neutral	neutral	0	0	20.02	0	20.02	0
NMB333	Grp3	M	infant	2.71	M2/3	LCA	neutral	gain	gain	0	0	2.01	1	2.32	1
NMB335	Grp3	M	child	4.27	M0/1	LCA	neutral	neutral	neutral	0	0	0.45	1	0.58	1
NMB341	SHH	M	infant	0.71	M0	CLA	neutral	neutral	neutral	0	0	7.14	0	7.14	0
NMB344	Grp3	M	child	5.06	M3	CLA	gain	neutral	neutral	0	0	3.5	1	4.66	1
NMB345	SHH	F	infant	1.53	M0	CLA	neutral	neutral	neutral	0	0	6.24	0	6.24	0
NMB346	Grp4	M	child	7.75	M0	CLA	neutral	loss	gain	0	0	3.09	0	3.09	0
NMB347	SHH	M	child	9.85	M0	CLA	neutral	neutral	neutral	0	0	3.21	0	3.21	0
NMB357	Grp4	M	child	3.5	M1	CLA	neutral	neutral	gain	0	0	3.33	0	3.33	0
NMB358	Grp4	M	child	13.85	M0	CLA	neutral	loss	gain	0	0	1.59	1	2.54	1
NMB360	Grp4	M	child	11.1	M2	CLA	neutral	loss	gain	0	0	8.8	0	8.8	0
NMB361	Grp3	F	infant	2.06	M3	LCA	neutral	loss	gain	0	1	0.14	1	0.36	1
NMB362	Grp4	M	child	7.03	M0	DN	neutral	loss	gain	0	0	1.75	1	2.25	1
NMB363	SHH	M	infant	0.6	M0	MBEN	neutral	neutral	neutral	0	0	4.94	0	4.94	0
NMB364	SHH	F	child	3.71	M0	DN	neutral	neutral	neutral	0	0	4.22	0	4.22	0
NMB365	Grp4	M	child	4.71	M0	CLA	neutral	loss	gain	0	0	3.21	0	3.21	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB368	Grp4	M	child	12	M1	NOS	neutral	loss	gain	0	0	1.63	1	2.61	1
NMB369	Grp4	M	child	9.24	M0	NOS	neutral	neutral	gain	0	0	3.54	0	3.54	0
NMB370	SHH	F	child	6.46	M0	CLA	neutral	loss	gain	0	0	16.8	0	16.8	0
NMB371	SHH	M	infant	2.29	M0	DN	neutral	neutral	neutral	0	0	1.63	1	8.89	0
NMB373	Grp4	M	child	5.25	M2	CLA	neutral	neutral	neutral	0	0	2.22	1	8.03	1
NMB374	Grp3	F	infant	0.64	M0	CLA	neutral	neutral	neutral	0	0	1.5	1	1.52	1
NMB375	Grp3	M	infant	1.13	M3	CLA	neutral	neutral	neutral	0	0	7.93	0	7.93	0
NMB376	Grp3	M	child	3.85	M3	CLA	neutral	neutral	gain	0	0	1.39	1	2.92	1
NMB377	Grp4	F	child	8.96	M0	CLA	neutral	loss	gain	0	0	6.47	0	6.47	0
NMB378	Grp3	M	infant	2.49	M0	CLA	gain	neutral	neutral	0	0	6.24	0	6.24	0
NMB379	SHH	M	infant	1.45	M0	CLA	neutral	neutral	neutral	0	0	0.66	1	1.19	0
NMB381	Grp3	M	child	4.65	M3	NOS	neutral	gain	gain	0	0	5.23	0	5.23	0
NMB383	Grp4	M	child	15.65	M0/1	NOS	neutral	loss	gain	0	0	3.28	0	3.28	0
NMB384	Grp3	M	child	4.57	M0	NOS	neutral	neutral	neutral	0	0	1.5	1	2.13	1
NMB385	Grp4	M	child	9.82	M0	CLA	neutral	gain	gain	0	0	8.16	0	8.16	0
NMB387	Grp4	M	child	6.76	M0	NOS	neutral	neutral	gain	0	0	5.12	0	5.12	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB388	Grp3	M	child	11.92	M0	CLA	neutral	neutral	gain	0	0	7.57	0	7.57	0
NMB393	Grp4	M	child	11.51	M0	CLA	neutral	loss	gain	0	0	3.48	1	3.24	0
NMB394	Grp4	M	child	5.71	M0	NOS	neutral	neutral	neutral	0	0	5.7	0	5.7	0
NMB398	Grp4	M	child	5.49	M1	NOS	neutral	neutral	neutral	0	0	5.73	0	5.73	0
NMB399	Grp3	M	child	7.3	M3	CLA	neutral	loss	gain	0	0	5.11	0	5.11	0
NMB400_1	SHH	M	child	6.99	M3	CLA	neutral	neutral	neutral	0	0	4.53	0	4.53	0
NMB401	Grp4	M	child	5.8	M3	CLA	neutral	loss	gain	0	0	3.16	1	3.78	0
NMB403	Grp4	M	child	6	M3	NOS	neutral	loss	gain	0	0	2.89	0	2.89	0
NMB405	Grp3	M	child	15.97	M1	LCA	gain	neutral	gain	0	0	3.78	0	3.78	0
NMB406	SHH	M	child	7.73	M1	DN	neutral	loss	neutral	0	0	2.65	0	2.65	0
NMB407	Grp4	F	child	6.34	M0	CLA	neutral	neutral	gain	0	0	2.63	0	2.63	0
NMB41	Grp3	F	child	5.3	M3	NOS	neutral	gain	gain	0	0	1.05	1	1.18	1
NMB410	Grp4	M	child	15.82	M0	CLA	neutral	loss	gain	0	0	1.64	0	1.64	0
NMB413	SHH	F	child	11.4	M0	LCA	neutral	neutral	neutral	1	0	1.91	0	1.91	0
NMB415	Grp4	F	child	7.93	M0	DN	neutral	neutral	neutral	0	0	2.42	0	2.42	0
NMB416	Grp4	F	child	4.64	M0	CLA	neutral	gain	gain	0	0	1.42	0	4.33	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB418	Grp4	M	child	6.44	M3	CLA	neutral	neutral	gain	0	0	1.99	0	1.99	0
NMB419	Grp4	F	child	6.01	M0	CLA	neutral	loss	gain	0	0	1.8	0	4.25	0
NMB420	Grp3	M	infant	1.43	M3	CLA	neutral	neutral	neutral	0	0	1.64	1	2.77	0
NMB421	Grp4	F	child	12.01	M2	CLA	neutral	loss	gain	0	0	1.5	0	4.28	0
NMB422	Grp3	M	child	3.74	M0	CLA	neutral	neutral	neutral	0	0	1.32	0	1.32	0
NMB43	Grp4	M	child	9.96	M2	CLA	neutral	loss	gain	0	0	4.16	1	6.39	1
NMB433	Grp3	M	infant	1.89	M0	CLA	gain	neutral	neutral	0	0	8.72	0	8.72	0
NMB435	SHH	M	child	15	M3	DN	neutral	neutral	neutral	0	0	1.69	0	1.69	0
NMB437	SHH	F	child	14.38	M2	NOS	neutral	neutral	neutral	0	0	0.64	1	0.9	1
NMB438	Grp4	M	child	5.01	M0	CLA	neutral	neutral	gain	0	0	1.72	1	1.74	0
NMB439	SHH	F	child	3.53	M0	LCA	loss	loss	loss	0	0	0.94	1	1.96	1
NMB440	Grp3	M	child	5.67	M3	LCA	neutral	neutral	neutral	0	0	1.17	0	1.17	0
NMB441	Grp3	F	child	3.17	M0	NOS	neutral	neutral	neutral	0	0	1.29	0	1.29	0
NMB443	Grp4	F	infant	1.91	M0	CLA	neutral	neutral	neutral	0	0	4.69	0	4.69	0
NMB445	Grp4	M	child	7.95	M0	DN	neutral	gain	neutral	0	0	3.77	0	3.77	0
NMB456	Grp3	M	infant	2.15	M0	CLA	neutral	neutral	neutral	0	0	2.35	0	2.35	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB457	Grp4	F	child	4.71	M0	CLA	neutral	gain	gain	0	0	7.68	0	7.68	0
NMB458	Grp3	M	child	4.42	M3	CLA	neutral	neutral	neutral	0	0	5.83	0	5.83	0
NMB460	SHH	M	child	3.91	M0	DN	loss	neutral	neutral	0	0	6.12	0	6.12	0
NMB462	SHH	M	child	7.63	M0	DN	neutral	loss	neutral	0	0	2.05	1	2.25	1
NMB463	Grp4	M	child	7.08	M0	LCA	neutral	neutral	gain	0	0	5.2	0	5.2	0
NMB465	SHH	F	infant	1.52	M3	MBEN	neutral	neutral	neutral	0	0	4.57	0	4.57	0
NMB466	SHH	M	infant	0.95	M2	DN	neutral	neutral	neutral	0	0	13.57	0	13.57	0
NMB471	SHH	F	infant	0.92	M2	MBEN	neutral	neutral	neutral	0	1	13.7	0	13.7	0
NMB474	SHH	M	infant	2.82	M0	CLA	neutral	neutral	neutral	0	0	3.6	1	3.69	1
NMB476	Grp3	M	infant	2.83	M0	CLA	neutral	neutral	neutral	0	0	2.08	1	2.67	1
NMB477	SHH	M	infant	2.56	M0	MBEN	neutral	neutral	neutral	0	0	1.44	1	3.7	0
NMB479	SHH	M	infant	1.86	M0	DN	neutral	loss	neutral	0	0	1.19	1	1.23	1
NMB48	Grp4	M	child	4.12	M0/1	NOS	neutral	neutral	neutral	0	0	16.66	0	16.66	0
NMB482	SHH	M	infant	1.64	M0	CLA	neutral	neutral	neutral	0	0	12.78	0	12.78	0
NMB483	SHH	M	infant	1.01	M3	MBEN	neutral	neutral	neutral	0	0	12.13	0	12.13	0
NMB484	Grp3	M	infant	1.2	M0	CLA	neutral	neutral	neutral	0	0	10.45	0	10.45	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB485	SHH	F	infant	2.79	M0	NOS	neutral	neutral	neutral	0	0	0.79	1	1.53	1
NMB486	SHH	F	infant	0.94	M0	DN	neutral	neutral	neutral	0	0	0.6	1	0.94	1
NMB489	Grp3	F	infant	1.91	M3	CLA	neutral	neutral	neutral	0	0	4.16	0	4.16	0
NMB490	Grp3	F	infant	1.97	M3	LCA	neutral	neutral	neutral	0	0	0.98	1	1.16	1
NMB491	Grp3	F	infant	2.62	M0	CLA	neutral	neutral	neutral	0	0	8.88	0	8.88	0
NMB493	Grp3	M	infant	1.86	M0	CLA	gain	neutral	neutral	0	0	2.68	1	4.02	1
NMB495	SHH	M	infant	1.51	M0	LCA	neutral	neutral	neutral	0	0	0.83	1	1.43	1
NMB496	SHH	M	infant	1.9	M2	DN	neutral	neutral	neutral	0	0	11.45	0	11.45	0
NMB497	SHH	F	infant	2.36	M0	NOS	neutral	neutral	neutral	0	0	9.19	0	9.19	0
NMB498	SHH	M	infant	2.42	M0	LCA	neutral	neutral	neutral	0	0	14.1	0	14.1	0
NMB499	Grp3	M	infant	2.47	M3	LCA	neutral	loss	gain	0	1	0.33	1	0.34	1
NMB502	SHH	M	child	7.47	M0/1	DN	loss	neutral	neutral	0	0	1.14	1	1.47	1
NMB51	Grp4	M	child	6.76	M0	CLA	neutral	gain	gain	0	0	10.44	0	10.44	0
NMB519	Grp3	M	child	4.22	M1	LCA	loss	neutral	neutral	0	0	8.45	0	8.45	0
NMB52	Grp4	F	child	8.58	M0	CLA	neutral	loss	gain	0	0	11.2	0	11.2	0
NMB521	Grp3	M	child	6.05	M0	LCA	neutral	neutral	neutral	0	0	7.26	0	7.26	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB529	Grp4	F	child	8.11	M0	LCA	neutral	loss	gain	0	0	6.68	0	6.68	0
NMB530	Grp3	M	child	8.05	M0	NOS	neutral	loss	gain	0	0	1.72	1	2.08	1
NMB531	Grp4	F	infant	2.69	M0	CLA	neutral	neutral	neutral	0	0	7.59	1	8.29	0
NMB532	Grp4	M	child	7.69	M0	CLA	neutral	neutral	neutral	0	0	4.08	1	5.06	1
NMB533	Grp3	F	child	6.79	M0	LCA	neutral	neutral	neutral	0	0	2.09	1	2.61	1
NMB534	SHH	F	infant	0.48	M0	NOS	neutral	neutral	neutral	0	0	0.36	1	0.37	1
NMB535	Grp3	M	infant	1.97	M3	LCA	neutral	loss	gain	0	1	0.4	1	0.41	1
NMB536	Grp3	M	child	8.46	M0	CLA	neutral	loss	neutral	0	0	6.87	0	6.87	0
NMB537	SHH	F	infant	0.54	M0	DN	neutral	neutral	neutral	0	0	1.78	0	1.78	0
NMB542	Grp4	M	child	6.69	M0	CLA	neutral	neutral	neutral	0	0	3.43	1	7.74	1
NMB546	Grp4	M	child	11.65	M2	LCA	neutral	loss	gain	0	0	4.64	1	5.13	1
NMB548	SHH	M	child	14.69	M0	LCA	neutral	neutral	neutral	0	0	6.28	0	6.28	0
NMB549	SHH	F	child	5.68	M0	LCA	neutral	loss	neutral	1	0	2.2	1	2.76	1
NMB550	Grp4	M	child	5.83	M0	CLA	neutral	loss	gain	0	0	6.25	0	6.25	0
NMB553	SHH	F	infant	1.07	M1	DN	neutral	neutral	neutral	0	0	2.58	1	18.91	0
NMB554	SHH	M	infant	1.76	M3	DN	neutral	neutral	neutral	0	0	1.72	1	15.38	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB555	SHH	F	infant	0.3	M3	CLA	neutral	neutral	neutral	0	0	3.43	1	4.49	1
NMB560	SHH	F	child	11.68	M2	LCA	neutral	neutral	neutral	0	0	0.74	1	0.78	1
NMB568	Grp3	F	child	5.03	M2	CLA	neutral	loss	gain	0	0	2.03	1	2.91	1
NMB569	Grp3	M	infant	2.64	M2	CLA	neutral	neutral	neutral	0	0	1.83	1	7.15	0
NMB580	SHH	F	infant	1.88	M0	DN	neutral	neutral	neutral	0	1	0.96	1	0.98	1
NMB581	Grp4	F	child	7.07	M0	CLA	neutral	loss	loss	0	0	2.52	1	3.64	1
NMB582	Grp4	M	child	8.75	M1	CLA	neutral	neutral	gain	0	0	1.63	1	3.18	1
NMB583	Grp4	F	child	13.88	M0	CLA	neutral	loss	gain	0	0	2.73	1	3.48	1
NMB584	SHH	M	child	6.58	M0	LCA	neutral	loss	gain	0	0	0.62	1	1.11	1
NMB591	Grp3	M	child	7.01	M0	CLA	neutral	neutral	neutral	0	0	1.4	1	1.82	1
NMB592	Grp4	F	child	10.46	M0	LCA	neutral	neutral	gain	0	0	1.31	1	2.38	1
NMB593	SHH	M	infant	0.72	M0	DN	neutral	loss	gain	0	0	1.11	0	1.11	0
NMB60	Grp3	M	child	5.01	M1	CLA	neutral	neutral	gain	0	0	1.01	1	1.02	1
NMB600	Grp3	M	child	5.82	M3	CLA	neutral	neutral	neutral	0	0	5.76	0	5.76	0
NMB606	Grp4	M	child	13.81	M0	CLA	neutral	neutral	neutral	0	0	4.91	0	4.91	0
NMB608	SHH	M	infant	1.24	M0	MBEN	neutral	neutral	neutral	0	0	6.47	0	6.47	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB609	Grp3	M	infant	2.82	M0/1	DN	gain	neutral	neutral	0	0	12.6	0	12.6	0
NMB610	Grp3	M	child	8.88	M0	CLA	neutral	neutral	neutral	0	0	0.54	0	0.54	0
NMB612	SHH	F	infant	2.47	M0/1	CLA	neutral	neutral	neutral	0	0	0.57	1	2.46	1
NMB615	Grp3	M	child	3.77	M2	CLA	neutral	neutral	neutral	0	0	5.98	0	5.98	0
NMB618	Grp3	M	child	3.89	M0	CLA	gain	neutral	neutral	0	0	2.65	1	18.26	0
NMB619	SHH	M	child	9.3	M0	NOS	neutral	neutral	neutral	0	0	9.17	0	9.17	0
NMB620	Grp4	F	child	8.85	M0	CLA	neutral	loss	gain	0	0	9.31	0	9.31	0
NMB624	Grp4	M	child	9.69	M0	CLA	gain	loss	gain	0	0	4.91	0	4.91	0
NMB625	SHH	F	infant	0.3	M1	CLA	neutral	neutral	neutral	0	0	0.34	1	0.41	1
NMB626	Grp3	F	child	3.97	M0	CLA	neutral	neutral	gain	0	0	3.62	0	5.84	0
NMB629	Grp3	F	child	12.16	M1	CLA	neutral	loss	neutral	0	0	1.59	1	2.82	1
NMB63	SHH	M	child	11.45	M0	CLA	neutral	neutral	neutral	0	0	11.36	0	11.36	0
NMB632	Grp4	M	child	4.83	M0/1	CLA	neutral	loss	gain	0	0	2.1	1	2.14	1
NMB633	Grp3	F	child	4.54	M1	CLA	neutral	neutral	neutral	0	0	0.99	1	1	1
NMB637	Grp3	M	child	3.25	M0	CLA	gain	gain	gain	0	0	0.25	1	0.26	1
NMB64	SHH	F	infant	1.43	M0	MBEN	neutral	neutral	neutral	0	0	1.57	1	2.03	1

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB645	SHH	M	child	12.44	M0	NOS	neutral	neutral	neutral	0	0	0.94	0	0.94	0
NMB646	Grp4	M	child	13.23	M0	CLA	neutral	loss	gain	0	0	1.49	0	1.49	0
NMB647	Grp3	M	infant	1.7	M3	CLA	gain	gain	gain	attention_req	0	1.23	0	1.23	0
NMB649	SHH	F	infant	1.94	M0	MBEN	neutral	neutral	neutral	0	0	1.75	1	9.31	0
NMB658	Grp3	M	infant	1.56	M3	LCA	neutral	loss	gain	0	1	0	1	0.16	1
NMB662	SHH	F	infant	0.83	M0	CLA	neutral	neutral	gain	0	0	0.18	1	0.32	1
NMB666	Grp3	M	infant	2.24	M3	CLA	gain	neutral	gain	0	0	0.39	1	1.21	1
NMB667	SHH	F	infant	2.62	M0	LCA	loss	neutral	neutral	0	0	0.47	1	0.85	1
NMB669	Grp3	F	infant	2.56	M0	CLA	neutral	loss	gain	0	1	0.52	1	0.65	1
NMB670	SHH	M	infant	0.99	M3	CLA	neutral	neutral	neutral	0	0	1.25	1	1.28	1
NMB671	Grp4	F	infant	2.22	M0	CLA	neutral	neutral	neutral	0	0	15.87	0	15.87	0
NMB673	SHH	M	infant	1.14	M0	CLA	neutral	neutral	neutral	0	0	14.9	0	14.9	0
NMB674	SHH	F	infant	2.63	M0	DN	neutral	neutral	neutral	0	0	13.52	0	13.52	0
NMB675	SHH	F	infant	2.72	M0	CLA	neutral	neutral	neutral	0	0	14.39	0	14.39	0
NMB676	SHH	M	infant	0.96	M0	CLA	neutral	neutral	neutral	0	0	13.97	0	13.97	0
NMB679	Grp3	M	infant	2.5	M3	LCA	neutral	neutral	neutral	0	0	15.5	0	15.5	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB68	Grp4	F	child	6.86	M0	CLA	neutral	loss	gain	0	0	10.31	0	10.31	0
NMB684	Grp3	F	infant	2.59	M3	CLA	neutral	gain	gain	0	0	13.4	0	13.4	0
NMB69	Grp4	M	child	7.75	M0/1	CLA	neutral	loss	gain	0	0	8.19	0	8.19	0
NMB690	SHH	M	infant	2.57	M3	DN	neutral	neutral	neutral	0	0	0.08	1	9.1	1
NMB70	Grp3	F	child	3.29	M0/1	CLA	gain	neutral	neutral	0	0	7	0	7	0
NMB706	SHH	M	child	10.79	M0	DN	neutral	neutral	neutral	0	0	6.55	0	6.55	0
NMB709	Grp3	F	child	8.48	M0/1	NOS	neutral	neutral	gain	0	0	5.32	0	5.32	0
NMB710	Grp4	F	child	9.31	M2	DN	neutral	loss	gain	0	0	0.33	1	5.13	0
NMB711	SHH	F	infant	1.77	M0	DN	neutral	loss	gain	0	0	4.99	0	4.99	0
NMB712	Grp4	F	child	4.93	M0	NOS	neutral	neutral	neutral	0	0	4.91	0	4.91	0
NMB715	Grp4	F	child	4.59	M0	NOS	neutral	neutral	neutral	0	0	4.03	0	4.03	0
NMB716	Grp4	F	infant	2.98	M0	NOS	neutral	loss	gain	0	0	1.78	1	2.68	0
NMB717	Grp3	M	infant	0.24	M1	DN	neutral	loss	gain	0	0	4.85	0	4.85	0
NMB718	Grp4	M	child	6.91	M0	CLA	neutral	loss	gain	0	0	0.85	0	3.57	0
NMB719	SHH	M	child	7.46	M2	LCA	neutral	neutral	neutral	0	0	1.05	0	1.05	0
NMB723	Grp4	F	child	3.65	M0	DN	neutral	neutral	neutral	0	0	1.11	0	1.11	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB724	SHH	F	infant	1.21	M0	CLA	neutral	neutral	neutral	0	0	0.69	1	1.03	1
NMB725	Grp4	M	child	5.75	M3	CLA	neutral	loss	gain	0	0	0.44	0	0.44	0
NMB726	SHH	M	child	3.08	M0	DN	neutral	neutral	neutral	0	0	0.47	0	0.47	0
NMB727	Grp4	M	child	6.62	M3	CLA	neutral	loss	gain	0	0	0	1	1.92	0
NMB729	Grp3	M	child	6.91	M0	CLA	neutral	neutral	neutral	0	0	1.13	1	1.29	1
NMB730	SHH	M	child	11.06	M0	CLA	neutral	neutral	neutral	0	0	1.98	0	3.09	0
NMB733	Grp4	F	child	6.67	M2	CLA	neutral	gain	gain	0	0	1.69	0	3.05	0
NMB734	Grp4	M	child	4.24	M3	CLA	neutral	neutral	neutral	0	0	0.99	0	0.99	0
NMB735	Grp4	M	child	15.92	M0	CLA	neutral	loss	gain	0	0	0.61	0	0.61	0
NMB736	Grp4	M	child	6.94	M2	CLA	neutral	neutral	gain	0	0	1.09	0	1.09	0
NMB737	Grp4	M	child	7.59	M3	CLA	neutral	loss	gain	0	0	1.4	0	1.4	0
NMB738	SHH	F	child	9.82	M3	CLA	neutral	neutral	neutral	0	0	1.33	1	1.49	0
NMB739	Grp4	M	child	11.42	M2	CLA	neutral	loss	gain	0	0	1.7	0	1.7	0
NMB741	Grp3	M	infant	2.94	M0	LCA	neutral	loss	gain	0	1	0.23	1	0.45	1
NMB742	Grp3	M	infant	1.91	M3	CLA	neutral	neutral	neutral	0	0	0.86	0	0.86	0
NMB743	SHH	M	child	14.31	M0	LCA	loss	loss	neutral	0	0	0.1	1	0.63	1

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB745	SHH	M	child	14.82	M0	DN	neutral	neutral	neutral	0	0	2.75	0	2.75	0
NMB754	Grp3	M	child	3.78	M0	CLA	neutral	neutral	neutral	0	0	1.23	0	1.23	0
NMB755	Grp3	M	child	6.38	M2	CLA	gain	neutral	neutral	0	0	1.29	0	1.29	0
NMB756	SHH	F	child	3.27	M0	DN	loss	neutral	neutral	0	0	0.7	0	0.7	0
NMB758	Grp3	M	child	5.21	M1	LCA	neutral	neutral	gain	0	1	0.76	1	0.85	1
NMB759	SHH	M	infant	1.18	M3	CLA	neutral	neutral	neutral	0	0	0.48	0	0.48	0
NMB760	Grp3	M	infant	2.96	M3	CLA	neutral	loss	gain	0	1	0.51	1	0.82	1
NMB761	Grp4	M	child	11.72	M0	CLA	neutral	loss	gain	0	0	1.11	0	1.11	0
NMB762	Grp3	M	infant	0.82	M3	CLA	gain	loss	gain	0	0	0.47	0	0.47	0
NMB763	SHH	F	child	8.29	M0	LCA	neutral	neutral	neutral	0	0	0.98	0	0.98	0
NMB764	Grp3	M	child	8.9	M0	CLA	neutral	neutral	neutral	0	0	1.01	0	1.01	0
NMB766	Grp3	F	child	3.57	M3	NOS	neutral	neutral	neutral	0	0	1.28	1	1.53	1
NMB768	Grp3	F	child	9.31	M1	CLA	neutral	neutral	neutral	0	0	0.95	1	0.98	1
NMB769	Grp3	M	child	4.01	M3	LCA	neutral	neutral	gain	0	1	0	1	0.18	1
NMB77	Grp4	F	child	8.54	M3	CLA	neutral	loss	gain	0	0	10.61	0	10.61	0
NMB770	Grp4	M	child	7.98	M3	CLA	neutral	loss	gain	0	0	0.91	1	0.97	1

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB771	Grp4	M	child	15.43	M0	CLA	neutral	loss	gain	0	0	1.16	0	1.16	0
NMB772	Grp4	M	child	7.93	M3	CLA	neutral	loss	gain	0	0	0.74	0	0.74	0
NMB773	Grp4	M	child	15.3	M3	DN	neutral	loss	gain	0	0	0.76	0	0.76	0
NMB774	Grp3	M	child	5.89	M0	CLA	neutral	neutral	gain	0	0	0.86	0	0.86	0
NMB78	Grp4	M	child	5.48	M2	CLA	neutral	neutral	gain	0	0	0.29	1	2.05	1
NMB782	Grp4	M	child	5.44	M0	CLA	neutral	neutral	gain	0	0	0.58	0	0.58	0
NMB783	Grp3	F	child	3.47	M3	CLA	neutral	neutral	neutral	0	0	0.83	0	1.99	0
NMB784	Grp3	M	child	12.87	M0	CLA	gain	neutral	neutral	0	0	0.81	0	1.89	0
NMB785	Grp4	M	child	8.44	M0	CLA	neutral	loss	gain	0	0	0.6	0	0.6	0
NMB786	Grp3	M	child	10.06	M3	CLA	gain	neutral	neutral	0	0	0.48	1	0.82	0
NMB787	Grp4	F	child	13.54	M0	CLA	neutral	loss	gain	0	0	0.86	0	0.86	0
NMB788	SHH	M	child	10.51	M0	LCA	neutral	loss	neutral	0	0	0.85	1	0.89	0
NMB790	SHH	F	child	3.44	M3	DN	neutral	neutral	neutral	0	0	8.84	0	8.84	0
NMB791	SHH	M	child	11.37	M0	CLA	neutral	neutral	neutral	0	0	0.65	0	0.65	0
NMB795	Grp3	M	child	5.43	M0	LCA	neutral	loss	gain	0	1	0.7	0	0.7	0
NMB797	Grp3	F	child	10.7	M3	CLA	neutral	neutral	neutral	0	0	1.68	0	1.68	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS_ Time	PFS_ Status	OS_ Time	OS_ Status
NMB798	SHH	M	infant	1.87	M0	DN	neutral	neutral	neutral	0	0	0.43	0	1.65	0
NMB80	Grp4	F	child	8.02	M0	NOS	neutral	neutral	gain	0	0	8.03	0	8.03	0
NMB81	SHH	F	child	15.06	M3	DN	neutral	neutral	neutral	0	0	4.44	1	7.01	1
NMB82	Grp4	M	child	5.35	M0	CLA	neutral	loss	gain	0	1	8.71	0	8.71	0
NMB859	Grp4	M	child	10.64	M1	CLA	neutral	neutral	neutral	0	0	0.99	0	0.99	0
NMB867	Grp4	M	child	6.56	M1	CLA	neutral	neutral	gain	0	0	0.91	0	0.91	0
NMB868	Grp3	F	infant	2.74	M0	CLA	neutral	neutral	neutral	0	0	0.24	0	0.24	0
NMB869	SHH	M	child	13.22	M0	CLA	neutral	neutral	neutral	0	0	0.73	0	0.73	0
NMB89	Grp3	F	child	4.51	M3	CLA	neutral	loss	gain	0	0	16.02	0	16.02	0
NMB92	Grp3	M	child	4.23	M3	CLA	gain	loss	gain	0	0	0.31	1	0.78	1
NMB947	Grp4	M	child	12	M0/1	CLA	neutral	neutral	neutral	0	1	0.84	1	1.8	1
NMB948	SHH	M	child	15	M0/1	CLA	loss	neutral	neutral	0	0	0.73	1	1.3	1
NMB949	Grp4	M	child	9	M0/1	CLA	loss	neutral	neutral	0	0	9.95	0	9.95	0
NMB958	SHH	M	child	10	M0/1	CLA	neutral	neutral	neutral	0	0	0.95	0	10.2	0
NMB960	Grp4	M	child	5	M0/1	CLA	neutral	neutral	neutral	0	0	1.5	1	2.2	1
NMB961	Grp3	M	child	3.6	M0/1	LCA	neutral	neutral	neutral	0	0	10.4	0	10.4	0

Sample ID	Sub-group	Sex	Age	Age2	Stage	Pathology	q14	p17	q17	<i>GLI2</i> status	<i>MYC</i> status	PFS- Time	PFS- Status	OS- Time	OS- Status
NMB962	Grp3	M	child	6	M0/1	CLA	neutral	loss	gain	0	0	1.44	1	1.8	1
NMB964	Grp3	M	child	9	M0/1	CLA	neutral	loss	gain	0	0	1.45	1	1.5	1
NMB966	Grp3	M	child	6	M0/1	CLA	neutral	loss	gain	0	0	12.37	0	12.37	0
NMB967	Grp4	F	child	11	M0/1	CLA	neutral	gain	gain	0	0	11.07	0	11.07	0
NMB969	SHH	F	child	7	M0/1	LCA	neutral	neutral	neutral	0	0	8.73	0	8.73	0
NMB970	Grp3	M	child	5	M0/1	LCA	neutral	neutral	neutral	0	1	0.27	1	0.5	1
NMB972	Grp4	F	child	4	M2/3	CLA	neutral	neutral	neutral	0	0	0.75	1	0.8	1
NMB973	SHH	F	child	13	M2/3	LCA	neutral	neutral	neutral	0	0	2.2	1	2.9	1
NMB975	SHH	M	child	4	M2/3	CLA	neutral	neutral	neutral	0	0	1.71	1	1.8	1
NMB978	Grp4	M	child	7	M2/3	CLA	neutral	neutral	neutral	0	0	2.25	0	2.25	0

TABLE A.3. Clinical demographic of Shih biomarkers validation cohort used for chapter 4. Cohort consisted of 338 primary medulloblastoma tumours from Newcastle Medulloblastoma Archive. Molecular subgroup, gender, age, metastatic stage and pathology are shown. Histological subtype is coded as follows: CLAS- classic; LCA- large cell/anaplastic; DN- desmoplastic/nodular; MBEN- MB with extensive nodularity; NOS-not otherwise specified.

NMB NO	Subgroup	CLA	DN	MBEN	LCA	Sex	Age
NMB119	4	1	0	0	0	M	14.37
NMB130	4	1	0	0	0	M	7.46
NMB132	4	1	0	0	0	M	9.34
NMB162	4	1	0	0	0	M	5.14
NMB17	3	1	0	0	0	M	4.7
NMB171	3	1	0	0	0	M	7.71
NMB180	4	0	1	0	0	M	10.66
NMB185	4	1	0	0	0	M	9.65
NMB187	4	1	0	0	0	M	3.7
NMB203	4	1	0	0	0	M	6.24
NMB24	4	1	0	0	0	M	6.69
NMB250	4	1	0	0	0	M	4.8
NMB251	4	1	0	0	0	F	3.36
NMB255	4	1	0	0	0	M	10.26
NMB284	4	0	1	0	0	F	11.14
NMB30	3	1	0	0	0	M	5.42
NMB330	3	1	0	0	0	M	4.72
NMB346	4	1	0	0	0	M	7.75
NMB358	4	1	0	0	0	M	13.85
NMB362	4	0	1	0	0	M	7.03
NMB365	4	1	0	0	0	M	4.71
NMB377	4	1	0	0	0	F	8.96
NMB385	4	1	0	0	0	M	9.82
NMB388	3	1	0	0	0	M	11.92
NMB393	4	1	0	0	0	M	11.51
NMB407	4	1	0	0	0	F	6.34
NMB410	4	1	0	0	0	M	15.82
NMB415	4	0	1	0	0	F	7.93
NMB416	4	1	0	0	0	F	4.64
NMB419	4	1	0	0	0	F	6.01

NMB NO	Subgroup	CLA	DN	MBEN	LCA	Sex	Age
NMB422	3	1	0	0	0	M	3.74
NMB438	4	1	0	0	0	M	5.01
NMB445	4	0	1	0	0	M	7.95
NMB45	4	1	0	0	0	M	12.63
NMB457	4	1	0	0	0	F	4.71
NMB46	4	1	0	0	0	M	5.14
NMB49	4	1	0	0	0	M	5.88
NMB509	4	1	0	0	0	F	9.77
NMB51	4	1	0	0	0	M	6.76
NMB52	4	1	0	0	0	F	8.58
NMB532	4	1	0	0	0	M	7.69
NMB542	4	1	0	0	0	M	6.69
NMB582	4	1	0	0	0	F	7.07
NMB585	4	1	0	0	0	F	13.88
NMB591	3	1	0	0	0	M	7.01
NMB595	4	1	0	0	0	F	11.08
NMB601	4	1	0	0	0	M	5.45
NMB606	4	1	0	0	0	M	13.81
NMB610	3	1	0	0	0	M	8.88
NMB614	4	1	0	0	0	F	9.8
NMB618	3	1	0	0	0	M	3.89
NMB624	4	1	0	0	0	M	9.69
NMB626	3	1	0	0	0	F	3.97
NMB632	4	1	0	0	0	M	4.83
NMB637	3	1	0	0	0	M	3.25
NMB68	4	1	0	0	0	F	6.86
NMB710	4	0	1	0	0	M	8.55
NMB718	4	1	0	0	0	M	6.91
NMB723	4	0	1	0	0	F	3.65
NMB729	4	1	0	0	0	M	6.91

NMB NO	Subgroup	CLA	DN	MBEN	LCA	Sex	Age
NMB735	4	1	0	0	0	M	15.92
NMB748	3	1	0	0	0	M	9.31
NMB754	3	1	0	0	0	M	3.78
NMB764	4	1	0	0	0	M	8.9
NMB782	4	1	0	0	0	M	5.44
NMB784	3	1	0	0	0	M	12.87
NMB785	4	1	0	0	0	M	8.44
NMB787	4	1	0	0	0	F	13.54
NMB800	4	1	0	0	0	M	10.22
NMB812	3	1	0	0	0	M	4.12
NMB825	4	1	0	0	0	F	10.24
NMB832	4	1	0	0	0	M	7.39
NMB890	4	1	0	0	0	M	7.83
NMB891	3	1	0	0	0	F	7.39
NMB897	4	1	0	0	0	M	14.77

TABLE A.4. Clinical demographics of standard-risk NMB cohort used for chapters 4 and 5. Cohort consisted of 75 primary high-risk non-WNT/non-SHH medulloblastoma tumours from Newcastle Medulloblastoma Archive. Molecular subgroup, gender, age, metastatic stage and pathology are shown. 0 - negative status, 1 - positive status

Sample ID	MYC	Resection	Metastat	LCA	Pathology	Sex	Age	Subgroup
NMB118	0	0	0	1	1	F	11.85	4
NMB121	0	0	1	0	0	F	12.57	4
NMB125	0	0	1	0	0	F	5.99	4
NMB127	0	0	1	1	1	M	13.55	3
NMB129	0	0	1	1	1	M	5.14	3
NMB136	0	1	1	0	0	M	10.5	4
NMB140	0	1	0	0	0	M	9.37	4
NMB142	0	1	1	0	0	M	12.06	4
NMB144	0	1	1	0	0	M	5.14	4
NMB145	0	0	0	1	1	F	7.01	3
NMB165	0	1	0	0	0	M	12.71	4
NMB166	1	0	0	0	0	F	9.72	4
NMB167	0	1	0	0	0	F	6.51	4
NMB178	0	0	1	0	0	M	4.95	4
NMB183	0	1	1	0	0	F	11.6	4
NMB186	0	1	1	0	0	M	3.73	4
NMB190	0	1	0	0	0	M	11.71	4
NMB227	0	1	0	0	0	M	4.58	4
NMB257	0	0	1	0	0	M	10.14	4
NMB265	0	0	1	1	1	M	16.43	4
NMB266	0	1	0	0	0	M	7.89	4
NMB316	0	0	1	0	0	M	9.61	4
NMB318	0	0	1	1	1	M	5.91	3
NMB335	0	0	0	1	1	M	4.27	3
NMB344	0	1	1	0	0	M	5.06	3
NMB357	0	0	1	0	0	M	3.5	4
NMB360	0	0	1	0	0	M	11.1	4
NMB365	0	0	0	0	0	M	4.71	4
NMB373	0	0	1	0	0	M	5.25	4
NMB376	0	0	1	0	0	M	3.85	4

Sample ID	MYC	Resection	Metastat	LCA	Pathology	Sex	Age	Subgroup
NMB391	0	0	1	0	0	M	10.53	4
NMB401	0	1	1	0	0	M	5.8	4
NMB405	0	0	1	1	1	M	15.97	4
NMB407	0	0	0	0	0	F	6.34	4
NMB411	1	0	1	0	0	M	16.06	3
NMB418	0	1	1	0	0	M	6.44	4
NMB421	0	1	1	0	0	F	12.01	4
NMB43	0	0	1	0	0	M	9.96	4
NMB440	0	0	1	1	1	M	5.67	3
NMB459	0	0	1	0	0	M	4.42	3
NMB463	0	0	0	1	1	M	7.08	4
NMB519	0	0	1	1	1	M	4.22	3
NMB521	0	0	0	1	1	M	6.05	3
NMB529	0	1	0	1	1	F	8.11	4
NMB533	0	0	0	1	1	F	6.79	3
NMB536	0	1	0	0	0	M	8.46	3
NMB546	0	1	1	1	1	M	11.65	4
NMB561	0	1	1	1	1	M	11.47	4
NMB568	0	0	1	0	0	F	5.03	3
NMB583	0	0	1	0	0	M	8.75	4
NMB592	0	0	0	1	1	F	10.46	4
NMB593	0	0	1	0	0	F	16.21	4
NMB595	0	0	0	0	0	F	11.08	4
NMB60	0	0	1	0	0	M	5.01	3
NMB600	0	0	1	0	0	M	5.82	3
NMB615	0	0	1	0	0	M	3.77	3
NMB624	0	0	0	0	0	M	9.69	4
NMB629	0	0	1	0	0	F	12.16	3
NMB630	0	0	1	0	0	M	7.24	3
NMB633	0	0	1	0	0	F	4.54	3

Sample ID	MYC	Resection	Metastat	LCA	Pathology	Sex	Age	Subgroup
NMB638	1	0	0	1	1	M	6.78	3
NMB646	0	1	0	0	0	M	13.23	4
NMB718	0	0	0	0	0	M	6.91	4
NMB725	0	0	1	0	0	M	5.75	4
NMB727	0	0	1	0	0	M	6.62	4
NMB733	0	0	1	0	0	F	6.67	4
NMB734	0	0	1	0	0	M	4.24	4
NMB736	0	0	1	0	0	M	6.94	4
NMB737	0	1	1	0	0	M	7.59	4
NMB739	0	0	1	0	0	M	11.42	4
NMB749	0	0	1	0	0	F	7.97	4
NMB755	0	0	1	0	0	M	6.38	3
NMB758	1	0	1	1	1	M	5.21	3
NMB761	0	1	0	0	0	M	11.72	4
NMB767	0	0	0	1	1	M	6.95	4
NMB768	0	1	1	0	0	F	9.31	3
NMB769	1	0	1	1	1	M	4.01	3
NMB77	0	0	1	0	0	F	8.54	4
NMB770	0	0	1	0	0	M	7.98	4
NMB771	0	1	0	0	0	M	15.43	4
NMB772	0	1	1	0	0	M	7.93	4
NMB773	0	1	1	0	0	M	15.3	4
NMB774	0	1	0	0	0	M	5.89	3
NMB78	0	1	1	0	0	M	5.48	4
NMB783	0	0	1	0	0	F	3.47	3
NMB793	0	1	1	0	0	F	8.85	4
NMB795	1	0	0	1	1	M	5.43	3
NMB797	0	0	1	0	0	M	8.29	3
NMB801	0	0	1	1	1	M	10.46	4
NMB805	0	0	1	0	0	M	4.83	4

Sample ID	MYC	Resection	Metastat	LCA	Pathology	Sex	Age	Subgroup
NMB816	0	1	0	0	0	F	4.08	3
NMB82	1	0	0	0	0	M	5.35	4
NMB859	0	0	1	0	0	M	10.64	4
NMB867	0	0	1	0	0	M	6.56	4
NMB889	0	1	0	0	0	F	8.82	4
NMB89	0	0	1	0	0	F	4.51	3
NMB894	0	0	0	1	1	F	6.82	4
NMB915	0	0	1	0	0	F	10.4	4
NMB916	1	0	0	0	0	F	9.14	3
NMB92	0	0	1	0	0	M	4.23	3

TABLE A.5. Clinical demographics of high-risk NMB cohort used for chapters 4 and 5. Cohort consisted of 100 primary high-risk non-WNT/non-SHH medulloblastoma tumours from Newcastle Medulloblastoma Archive. Molecular subgroup, gender, age, metastatic stage and pathology are shown. 0 - negative status, 1 - positive status

AccessionID	Age	Gender	Histology	Metastasis	Subgroup
GSM2260746	3.65	M	Classic	1	Group3
GSM2260747	5.46	F	Classic	0	Group4
GSM2260748	6.55	F	Classic	0	Group4
GSM2260752	11.05	M	LCA	0	Group3
GSM2260753	3.19	M	Classic	0	Group4
GSM2260764	4.82	F	Classic	1	Group4
GSM2260766	7	M	Classic	0	Group3
GSM2260768	5	M	Classic	1	Group3
GSM2260769	11	M	Classic	0	Group4
GSM2260770	5.83	M	Classic	1	Group4
GSM2260771	3.29	M	Desmoplastic	0	Group4
GSM2260773	4.48	M	Classic	0	Group3
GSM2260782	13	F	Classic	1	Group4
GSM2260783	12	M	Classic	1	Group4
GSM2260784	6	M	Classic	1	Group3
GSM2260785	13	M	Classic	0	Group4
GSM2260788	9	F	Classic	1	Group3
GSM2260789	11	M	Classic	0	Group4
GSM2260790	7	F	Classic	0	Group4
GSM2260799	4.82	M	Classic	1	Group3
GSM2260812	8	F	Classic	0	Group3
GSM2260814	4	M	Classic	1	Group4
GSM2260815	10	M	LCA	1	Group4
GSM2260819	4	F	Classic	1	Group3
GSM2260820	5	M	LCA	1	Group3
GSM2260832	4.92	M	Classic	0	Group3
GSM2260833	6.61	F	Classic	0	Group4
GSM2260834	12.19	M	Classic	0	Group4
GSM2260837	5.62	F	Classic	0	Group4
GSM2260838	5.8	M	Classic	1	Group4

AccessionID	Age	Gender	Histology	Metastasis	Subgroup
GSM2260839	10.59	F	Classic	0	Group4
GSM2260842	4	M	Classic	1	Group3
GSM2260844	8	M	Classic	1	Group4
GSM2260845	4	F	Classic	0	Group4
GSM2260847	5	F	LCA	1	Group4
GSM2260849	4	F	LCA	0	Group3
GSM2260850	8	M	Classic	0	Group3
GSM2260852	8	M	Classic	0	Group3
GSM2260853	12	F	Classic	1	Group3
GSM2260859	14	M	Classic	1	Group4
GSM2260860	13	M	Classic	0	Group4
GSM2260861	10	M	Classic	0	Group4
GSM2260863	4	M	Classic	1	Group4
GSM2260866	8	M	Classic	1	Group4
GSM2260867	5	F	Classic	0	Group4
GSM2260869	9	M	Classic	1	Group4
GSM2260870	8	M	Classic	1	Group4
GSM2260871	10	F	Classic	1	Group4
GSM2260872	5	M	Classic	1	Group4
GSM2260873	4	M	Classic	1	Group4
GSM2260874	11	M	Classic	0	Group3
GSM2260875	10	M	Classic	0	Group3
GSM2260876	7	M	Classic	1	Group4
GSM2260877	6	F	Classic	0	Group4
GSM2260878	6	M	Classic	0	Group4
GSM2260879	3	F	Classic	0	Group4
GSM2260883	3	M	Classic	1	Group3
GSM2260884	10	M	Classic	0	Group4
GSM2260885	6	F	Classic	0	Group3
GSM2260886	14	F	Classic	1	Group4

AccessionID	Age	Gender	Histology	Metastasis	Subgroup
GSM2260888	9	M	Classic	1	Group4
GSM2260890	6	M	Classic	0	Group4
GSM2260891	4	M	Classic	1	Group4
GSM2260892	5	M	Classic	0	Group4
GSM2260893	10	M	Classic	0	Group4
GSM2260895	11	M	Classic	0	Group4
GSM2260897	4	M	Classic	1	Group4
GSM2260899	4	M	Classic	1	Group4
GSM2260900	8	M	Classic	0	Group4
GSM2260901	4	F	Classic	1	Group4
GSM2260902	8	F	Classic	1	Group4
GSM2260903	12	M	Classic	0	Group4
GSM2260904	7	F	Classic	1	Group4
GSM2260906	4	M	Classic	0	Group4
GSM2260907	14	M	Classic	0	Group3
GSM2260908	4	M	Classic	0	Group3
GSM2260910	8	M	Classic	0	Group4
GSM2260927	7.4	M	Classic	0	Group3
GSM2260928	12.5	M	Classic	0	Group4
GSM2260929	11.7	F	Classic	1	Group3
GSM2260930	13.4	M	Classic	0	Group4
GSM2260931	6.8	F	Classic	0	Group4
GSM2260934	9.19	M	Classic	0	Group4
GSM2260935	3.53	M	LCA	1	Group4
GSM2260962	9.09	M	LCA	0	Group4
GSM2260966	13	F	Classic	0	Group4
GSM2260968	9	F	Classic	1	Group4
GSM2260970	8	F	Classic	1	Group3
GSM2260974	4	M	Classic	0	Group4
GSM2260976	5	M	Classic	1	Group4

AccessionID	Age	Gender	Histology	Metastasis	Subgroup
GSM2260977	9	M	LCA	0	Group4
GSM2260979	7	M	Classic	1	Group4
GSM2260980	11	M	Classic	1	Group4
GSM2260984	10	F	Classic	1	Group4
GSM2260985	4	M	Classic	1	Group4
GSM2260986	3	F	Classic	0	Group4
GSM2260988	10	M	Classic	0	Group4
GSM2260990	5	M	Classic	1	Group4
GSM2260991	8	F	Classic	0	Group4
GSM2260993	10	F	Classic	0	Group4
GSM2260994	8	M	Classic	0	Group4
GSM2260996	6	F	Classic	1	Group4
GSM2260997	5	F	Classic	0	Group4
GSM2260999	15	F	Classic	1	Group4
GSM2261000	8	F	Desmoplastic	0	Group3
GSM2261001	8	M	Classic	0	Group3
GSM2261002	9	M	LCA	0	Group4
GSM2261003	4	F	LCA	0	Group4
GSM2261004	7.72	M	Classic	0	Group3
GSM2261006	7	M	LCA	0	Group4
GSM2261011	5	M	MBEN	1	Group4
GSM2261013	9	F	Desmoplastic	0	Group4
GSM2261015	11	M	Classic	0	Group4
GSM2261016	8	M	LCA	0	Group4
GSM2261017	13	F	Classic	0	Group4
GSM2261018	9	M	Classic	1	Group4
GSM2261020	4	M	LCA	1	Group4
GSM2261021	10	F	Classic	0	Group4
GSM2261025	5.91	M	Classic	0	Group3
GSM2261026	5.16	F	Desmoplastic	1	Group4

AccessionID	Age	Gender	Histology	Metastasis	Subgroup
GSM2261027	5.16	F	Classic	0	Group3
GSM2261032	12	M	Classic	0	Group4
GSM2261034	13	F	Classic	0	Group4
GSM2261036	7	M	Classic	0	Group4
GSM2261038	8	F	LCA	0	Group4
GSM2261040	10	M	Classic	0	Group4
GSM2261050	4	M	LCA	1	Group3
GSM2261053	14.36	M	Classic	0	Group4
GSM2261054	12.57	M	Classic	1	Group4
GSM2261055	5.99	F	Classic	1	Group4
GSM2261057	8.48	F	Desmoplastic	0	Group3
GSM2261059	4.92	F	Classic	1	Group4
GSM2261071	11	M	Classic	0	Group4
GSM2261076	5	M	LCA	1	Group3
GSM2261077	10	M	Classic	1	Group4
GSM2261078	8	M	Classic	0	Group4
GSM2261080	7	M	Classic	0	Group4
GSM2261081	6	M	Classic	1	Group3
GSM2261086	12	M	LCA	1	Group4
GSM2261087	9.3	F	Desmoplastic	1	Group4
GSM2261088	10.1	F	Classic	1	Group3
GSM2261089	13.54	M	Classic	1	Group3
GSM2261090	5.13	M	Classic	1	Group3
GSM2261094	7.1	M	Classic	0	Group4
GSM2261099	9	M	Classic	1	Group4
GSM2261104	7	F	LCA	1	Group4
GSM2261105	4	M	Classic	0	Group4
GSM2261106	6	M	Classic	0	Group4
GSM2261107	9	M	Classic	0	Group3
GSM2261108	8	M	Classic	0	Group4

AccessionID	Age	Gender	Histology	Metastasis	Subgroup
GSM2261110	3	M	Classic	1	Group4
GSM2261111	6	M	Classic	0	Group4
GSM2261113	8	M	Desmoplastic	0	Group4
GSM2261115	4	F	Classic	1	Group4
GSM2261116	12	F	Classic	0	Group4
GSM2261117	8	M	Classic	0	Group4
GSM2261123	9	M	MBEN	0	Group4
GSM2261124	7.4	M	Classic	0	Group4
GSM2261125	7.3	M	MBEN	0	Group4
GSM2261126	5.7	M	Classic	1	Group3
GSM2261127	6.9	F	Classic	0	Group4
GSM2261128	12.8	M	Classic	1	Group4
GSM2261129	7.8	F	Desmoplastic	0	Group4
GSM2261130	8.7	F	Classic	0	Group4
GSM2261132	7.74	M	LCA	0	Group4
GSM2261133	6.96	F	LCA	0	Group3
GSM2261134	9	F	Classic	0	Group3
GSM2261135	8.44	M	Classic	1	Group4
GSM2261138	3	M	Desmoplastic	1	Group4
GSM2261140	9	M	Classic	0	Group4
GSM2261142	7	M	LCA	0	Group3
GSM2261143	11	F	Classic	0	Group4
GSM2261146	4	F	Classic	0	Group4
GSM2261147	5	F	LCA	0	Group4
GSM2261153	6	F	Classic	1	Group3
GSM2261157	5	M	Classic	0	Group4
GSM2261158	8	M	Classic	1	Group4
GSM2261160	4	F	Classic	0	Group4
GSM2261161	10	M	Classic	0	Group4
GSM2261175	3.92	M	Classic	0	Group3

AccessionID	Age	Gender	Histology	Metastasis	Subgroup
GSM2261179	16	M	Classic	0	Group4
GSM2261182	12	M	Classic	1	Group4
GSM2261183	10	M	Classic	0	Group4
GSM2261185	8	F	Classic	0	Group4
GSM2261186	5	M	LCA	0	Group4
GSM2261192	8	M	Classic	0	Group4
GSM2261195	8	M	Classic	0	Group4
GSM2261197	8	M	Desmoplastic	1	Group4
GSM2261198	9	M	Classic	1	Group4
GSM2261207	9	F	Classic	1	Group4
GSM2261209	5	F	Classic	0	Group3
GSM2261210	7	M	Classic	0	Group3
GSM2261211	12	F	Classic	0	Group4
GSM2261212	10	F	Classic	0	Group3
GSM2261214	15	M	Classic	1	Group4
GSM2261215	13	M	Classic	0	Group3
GSM2261220	9.5	F	Classic	0	Group4
GSM2261224	4.18	M	Classic	1	Group4
GSM2261225	11.43	M	Classic	0	Group4
GSM2261226	10.74	M	LCA	0	Group3
GSM2261227	12.99	M	Classic	1	Group4
GSM2261228	5.79	M	Classic	0	Group4
GSM2261230	3.03	M	Classic	0	Group3
GSM2261233	3.08	M	Classic	1	Group3
GSM2261238	8.79	M	LCA	0	Group4
GSM2261240	7.26	M	Classic	0	Group4
GSM2261242	7.71	F	Desmoplastic	0	Group3
GSM2261243	14.75	F	Classic	0	Group4
GSM2261245	4.3	F	Classic	0	Group4
GSM2261258	5	M	Classic	0	Group4

AccessionID	Age	Gender	Histology	Metastasis	Subgroup
GSM2261288	13.5	M	Desmoplastic	0	Group4
GSM2261292	7.8	M	LCA	0	Group4
GSM2261303	5.1	M	Classic	1	Group3
GSM2261304	4.4	M	LCA	0	Group3
GSM2261306	5.3	M	LCA	0	Group3
GSM2261313	7.4	M	Classic	1	Group4
GSM2261327	11.6	M	LCA	0	Group3
GSM2261347	6.9	F	Classic	0	Group3
GSM2261351	5.5	M	LCA	1	Group3
GSM2261362	9	M	Desmoplastic	1	Group4
GSM2261367	4	F	Classic	1	Group4
GSM2261369	16	M	LCA	0	Group4
GSM2261370	11	F	Classic	1	Group4
GSM2261371	8	F	Classic	0	Group4
GSM2261373	11	M	Classic	1	Group4
GSM2261374	5	M	Classic	0	Group4
GSM2261377	6	M	Classic	0	Group4
GSM2261379	11	M	Classic	1	Group4
GSM2261380	8	M	MBEN	0	Group4
GSM2261385	11	M	MBEN	0	Group4
GSM2261390	16	M	Classic	0	Group4
GSM2261394	4	M	Classic	0	Group4
GSM2261402	4	M	LCA	0	Group3
GSM2261405	10	F	Desmoplastic	1	Group4
GSM2261448	4.25	M	LCA	0	Group3
GSM2261450	11	F	Classic	0	Group4
GSM2261452	10	M	Desmoplastic	0	Group4
GSM2261453	3	M	Classic	1	Group3
GSM2261455	4	M	LCA	0	Group3
GSM2261457	6	F	LCA	0	Group3

AccessionID	Age	Gender	Histology	Metastasis	Subgroup
GSM2261458	10	M	Desmoplastic	1	Group4
GSM2261459	13	M	Classic	0	Group4
GSM2261460	6	M	Classic	1	Group4
GSM2261497	6.07	M	Classic	0	Group4

TABLE A.6. Clinical demographics of the Cavalli dataset used for chapter 4. Cohort consisted of 244 primary medulloblastoma tumours from Cavalli et al. (2017) study. Molecular subgroup, gender, age, metastatic stage and pathology are shown. Histological subtype is coded as follows: LCA- large cell/anaplastic; Desmoplastic- desmoplastic/nodular; MBEN- MB with extensive nodularity

APPENDIX B

Publications

Epigenetic landscape correlates with genetic subtype but does not predict outcome in childhood acute lymphoblastic leukemia

Alem S Gabriel^{†1}, Fadhel M Lafta^{†1}, Edward C Schwalbe^{†1,2}, Sirintra Nakjang^{1,3}, Simon J Cockell³, Alice Iliasova^{1,2}, Amir Enshaei¹, Claire Schwab¹, Vikki Rand¹, Steven C Clifford¹, Sally E Kinsey⁴, Chris D Mitchell⁵, Ajay Vora⁵, Christine J Harrison¹, Anthony V Moorman¹, and Gordon Strathdee^{1,6,*}

¹Northern Institute for Cancer Research; Faculty of Medical Sciences; Newcastle University; Newcastle upon Tyne, UK; ²Department of Applied Sciences; Northumbria University; Newcastle upon Tyne, UK; ³Bioinformatics Support Unit; Faculty of Medical Sciences; Newcastle University; Newcastle upon Tyne, UK; ⁴Department of Pediatric Haematology and Oncology; Leeds General Infirmary; Leeds, UK; ⁵Department of Pediatric Oncology; John Radcliffe Hospital; Oxford, UK; ⁶Department of Haematology; Great Ormond Street Hospital; London, UK

[†]Contributed equally to the manuscript

Keywords: 450K, ALL, biomarker, childhood, methylation, relapse

Although children with acute lymphoblastic leukemia (ALL) generally have a good outcome, some patients do relapse and survival following relapse is poor. Altered DNA methylation is highly prevalent in ALL and raises the possibility that DNA methylation-based biomarkers could predict patient outcome. In this study, genome-wide methylation analysis, using the Illumina Infinium HumanMethylation450 BeadChip platform, was carried out on 52 diagnostic patient samples from 4 genetic subtypes [*ETV6-RUNX1*, high hyperdiploidy (HeH), *TCF3-PBX1* and *dic(9;20)* (*p11-13;q11*)] in a 1:1 case-control design with patients who went on to relapse (as cases) and patients achieving long-term remission (as controls). Pyrosequencing assays for selected loci were used to confirm the array-generated data. Non-negative matrix factorization consensus clustering readily clustered samples according to genetic subgroups and gene enrichment pathway analysis suggested that this is in part driven by epigenetic disruption of subtype specific signaling pathways. Multiple bioinformatics approaches (including bump hunting and individual locus analysis) were used to identify CpG sites or regions associated with outcome. However, no associations with relapse were identified. Our data revealed that *ETV6-RUNX1* and *dic(9;20)* subtypes were mostly associated with hypermethylation; conversely, *TCF3-PBX1* and HeH were associated with hypomethylation. We observed significant enrichment of the neuroactive ligand-receptor interaction pathway in *TCF3-PBX1* as well as an enrichment of genes involved in immunity and infection pathways in *ETV6-RUNX1* subtype. Taken together, our results suggest that altered DNA methylation may have differential impacts in distinct ALL genetic subtypes.

Introduction

Acute lymphoblastic leukemia (ALL) is the most common form of childhood cancer, representing more than 80% of diagnosed childhood leukemia cases in the UK each year, with a gradually increasing incidence.¹ It has long been established that chromosomal abnormalities are major drivers of ALL. Current treatment involves risk stratification guided by age and white blood cell count (WBC), karyotype, and treatment response.^{2,3} Although risk stratification and multi-agent chemotherapy have achieved around 90% survival, about 10% of patients relapse.² Increasing evidence supports the

inclusion of additional genomic signatures generated by transcriptome, as well as copy number changes and mutations in risk stratification.⁴⁻⁶ Current efforts are focused on identifying biomarkers for predicting relapse or reducing late complications of intensified therapy.

DNA methylation is a key epigenetic modification, which occurs primarily at CpG dinucleotide sequences.⁷ CpG sites are underrepresented throughout the genome, with the exception of short stretches of DNA known as CpG islands, which are often associated with gene promoter regions.⁸ The development of cancer, including ALL, is associated with dramatic shifts in genomic DNA methylation, involving both genome wide hypomethylation

© Alem S Gabriel, Fadhel M Lafta, Edward C Schwalbe, Sirintra Nakjang, Simon J Cockell, Alice Iliasova, Amir Enshaei, Claire Schwab, Vikki Rand, Steven C Clifford, Sally E Kinsey, Chris D Mitchell, Ajay Vora, Christine J Harrison, Anthony V Moorman, and Gordon Strathdee

*Correspondence to: Gordon Strathdee; Email: gordon.strathdee@newcastle.ac.uk

Submitted: 05/01/2015; Revised: 05/29/2015; Accepted: 06/05/2015

<http://dx.doi.org/10.1080/15592294.2015.1061174>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

and localized hypermethylation of promoter-associated CpG islands.⁹ Hypermethylation of promoter-associated CpG islands leads to gene inactivation and many important tumor suppressor genes are known to be inactivated by this mechanism.¹⁰ Furthermore, the comparative ease of detection and tumor specificity of CpG island hypermethylation has led to considerable interest in their potential as novel prognostic biomarkers.¹¹ Such methylation based markers may help direct current therapies, such as methylation of the DNA repair gene, *MGMT*, which predicts response to therapy in glioblastoma patients,¹² or improve patient stratification, as we have demonstrated in chronic lymphocytic leukemia.¹³

While our understanding of methylation in cancer is improving, little is known about the role of methylation changes in the development and progression of childhood ALL. However, a number of studies have provided preliminary evidence that altered patterns of DNA methylation may be associated with outcome in ALL.¹⁴⁻¹⁹ Recent advances in whole genome screening technologies have facilitated the screening of CpG sites at a genomic level, generating a more thorough view of the methylation landscape^{15,20-23} and raised the possibility of using such technologies to identify methylation based biomarkers that could be used to further improve risk stratification in ALL patients.

In the present study we applied Illumina Infinium HumanMethylation450 genome-wide methylation arrays that cover > 485,000 methylation sites, including 99% of Refseq genes as well as 96% of CpG islands²⁴ to a cohort of 52 diagnostic ALL samples in a 1:1 case-control design (26 cases who subsequently relapsed and 26 controls in continuous remission) to identify novel CpG sites that may be associated with relapse in 4 major cytogenetic subgroups in BCP-ALL.

Results

Genome-wide patterns of DNA methylation are strongly associated with cytogenetic subgroups

Genome-wide methylation data derived from the Infinium HumanMethylation450 BeadChip arrays was compared across the 52 diagnostic ALL samples and compared with control normal cells (B lymphocytes (CD19) and monocytes (CD14) derived from healthy volunteers). The majority of samples (n = 40) were derived from the 2 most common cytogenetic subtypes (*ETV6-RUNX1*, n = 20 ; HeH, n = 20), which despite having a good outcome still account for ~40% of relapse cases due to their prevalence (Moorman et al. 2010). Additional samples from less common cytogenetic subtypes [*TCF3-PBX1*, n = 6; dic(9;20)(p11-13;q11), n = 6] were also included to determine if any relapse-associated methylation changes were independent of genetic subgroup and also to allow clearer definition of subgroup specificity of DNA methylation in childhood ALL. Non-negative matrix factorization (NMF) was used to reduce the dimensionality of the data from 10,000 probes to a few metagenes. For each factorization rank from 3 to 7, we assessed stability of factorization by cophenetic coefficient (Fig. S1) and silhouette scores of consensus subgroup assignments after 100 iterations (Fig. 1A). NMF separated the samples

into 4 groups which corresponded very closely to the subgroups (Fig. 1), with only a single HeH sample and a single dic(9;20) sample failing to cluster with their genetic counterparts (we have excluded the possibility of cryptic *TCF3-PBX1* translocation via interphase FISH). The identified subgroups are characterized by a positive silhouette score (Fig. 1B), indicative of samples being placed into the correct cluster, and show clear separation by principal component analysis (PCA) of the methylation data (Fig. 1C). This association with cytogenetic subtype is consistent with previous reports.^{15,20-23} The differentially methylated CpG sites that demonstrated subtype specificity are listed for each subtype in supplementary Table 1. Furthermore, there were clear differences in the genomic locations and directionality of methylation changes between the subtypes; for example, the *ETV6-RUNX1* and dic(9;20) subtypes exhibited relatively more hypermethylation than hypomethylation, while changes in the HeH and *TCF3-PBX1* subtypes were predominantly hypomethylated (Fig. S2). We have performed a gene enrichment pathway analysis of these subtype specific differentially methylated cytosines (DMCs) after correcting for probe distribution and multiple testing in KEGG database. Interestingly, this analysis identified 10 pathways that were significantly over-represented in the *ETV6-RUNX1* DMCs (after correction for multiple testing) and 8/10 of these pathways were related to immune function. While no pathways were identified as over-represented in the HeH and dic(9;20) subtypes, a highly significant association ($P = 5.7 \times 10^{-17}$, after correction) was found with neuroactive ligand-receptor interaction in the *TCF3-PBX1* subtype (online supplementary Table 2). The association between subtype specific DMCs and specific pathways further implicates the differential methylation in different biological behavior of the cytogenetic subtypes.

All samples in the cohort had also previously been analyzed using Multiplex Ligation-dependent Probe Amplification (MLPA)²⁵ to identify copy number alterations in genes with a known role in ALL development (Fig. 2A). It is possible that the complexity of data derived from the HumanMethylation450 BeadChip arrays was masking changes at such key leukemia associated genes or that DNA methylation changes may be acting as a second hit at sites of heterozygous deletions. Therefore, methylation at all CpG sites associated with the 8 loci (*CRLF2*, *IL3RA*, and *CSF2RA*, considered as *PARI* locus) covered by the MLPA analysis were extracted from the array data. As illustrated for *PAX5* in Figure 2B, methylation levels were similar across the sample set. Raw methylation data could also be used to assess potential changes in copy number.²⁶ This analysis confirmed the MLPA data showing *PAX5* deletions in 5 of the patients (Fig. S3). However, methylation of *PAX5* did not vary in the samples with confirmed deletions (Fig. 2B), suggesting that altered methylation was not functioning as a second hit, at least for *PAX5*.

Genome-wide patterns of DNA methylation are not significantly different in patients that subsequently relapsed

The development of relapse is a crucial determinant of outcome, as survival rates following relapse are much lower. Identification of patients at the time of diagnosis who will subsequently

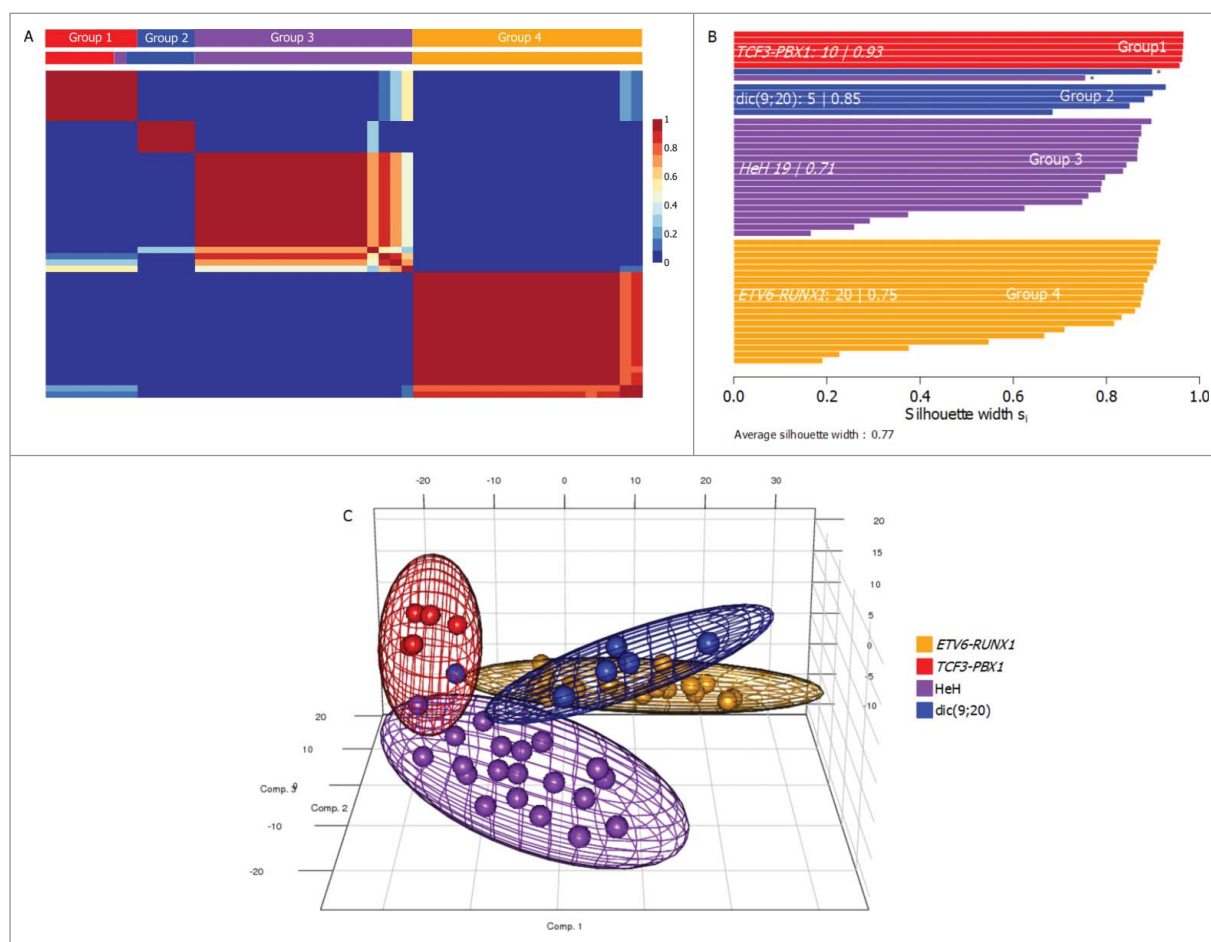


Figure 1. Methylation patterns identify cytogenetic groups in BCP-ALL. (A) Consensus clustering of DNA methylation patterns in 52 BCP-ALL samples. NMF using standard methods was carried out over 100 runs for 3–6 metagenes, with the cophenetic coefficient supporting 4 groups (metagenes). Colored squares above each column indicate the cytogenetic subgroup for each samples, showing the single *ETV6-RUNX1* samples and single *dic(9;20)* sample that clustered with the *TCF3-PBX1* group (group 1). (B) Silhouette plot by sample type and cytogenetic groups. Silhouette plots of consensus NMF subgroups demonstrate close relationships between cytogenetic subgroup and methylation subgroup assignment. For each subgroup, the number of members, the percentage of cluster members and average silhouette (s_i) width are shown. Samples marked with an asterisk indicate outlier cytogenetic cases that do not cluster with patients of the same cytogenetic subtype. (C) Principal component analysis based on 10000 most variable DMCs, labeled by methylation subtype. The first 3 principal component scores are shown for each sample. For each group, covariance spheroids, colored by the pre-dominating cytogenetic subgroup, are plotted along the 95% confidence intervals.

relapse and, indeed, those unlikely to relapse, is extremely valuable to optimize their treatment. Consequently, genome-wide methylation data was analyzed, using a number of different methods, to attempt to discover methylation-based signatures in diagnostic samples that were predictive of subsequent relapse. As described above, unsupervised NMF consensus clustering separated the samples by underlying genetic subgroups (Fig. 1). However, within each subgroup, there was no evidence of clustering of samples based on eventual outcome. In case the strong association with cytogenetics was masking a relapse signature, this analysis was repeated following the removal of data from the probes that were differentially methylated between subgroups; however, this analysis also showed no evidence of clustering according to outcome and no individual CpG site exhibited a statistically significant correlation with outcome (data not shown). Furthermore, analysis of individual CpG sites [differential CpG

sites were classed as having a difference in mean β value > 0.2 and an adjusted P -value < 0.01 (Nordlund-Backlin et al. 2013)] did not identify any individual CpG sites that was significantly associated with relapse. Genetic subtype specific analysis, for the *ETV6-RUNX1* and HeH subgroups, also yielded no individual probes significantly associated with subsequent relapse.

Altered DNA methylation often occurs coordinately across genomic regions, such as CpG islands. To determine whether any such regions were differentially methylated between samples from patients who subsequently relapsed and those who did not, a bump-hunting algorithm²⁷ within the Bumhunter package was utilized, with 1000 permutations. Probes were clustered into a region based on distance: all differentially methylated probes that were located within 300 bp of another differentially methylated probe were placed into the same cluster group, so that window widths were flexible and defined by

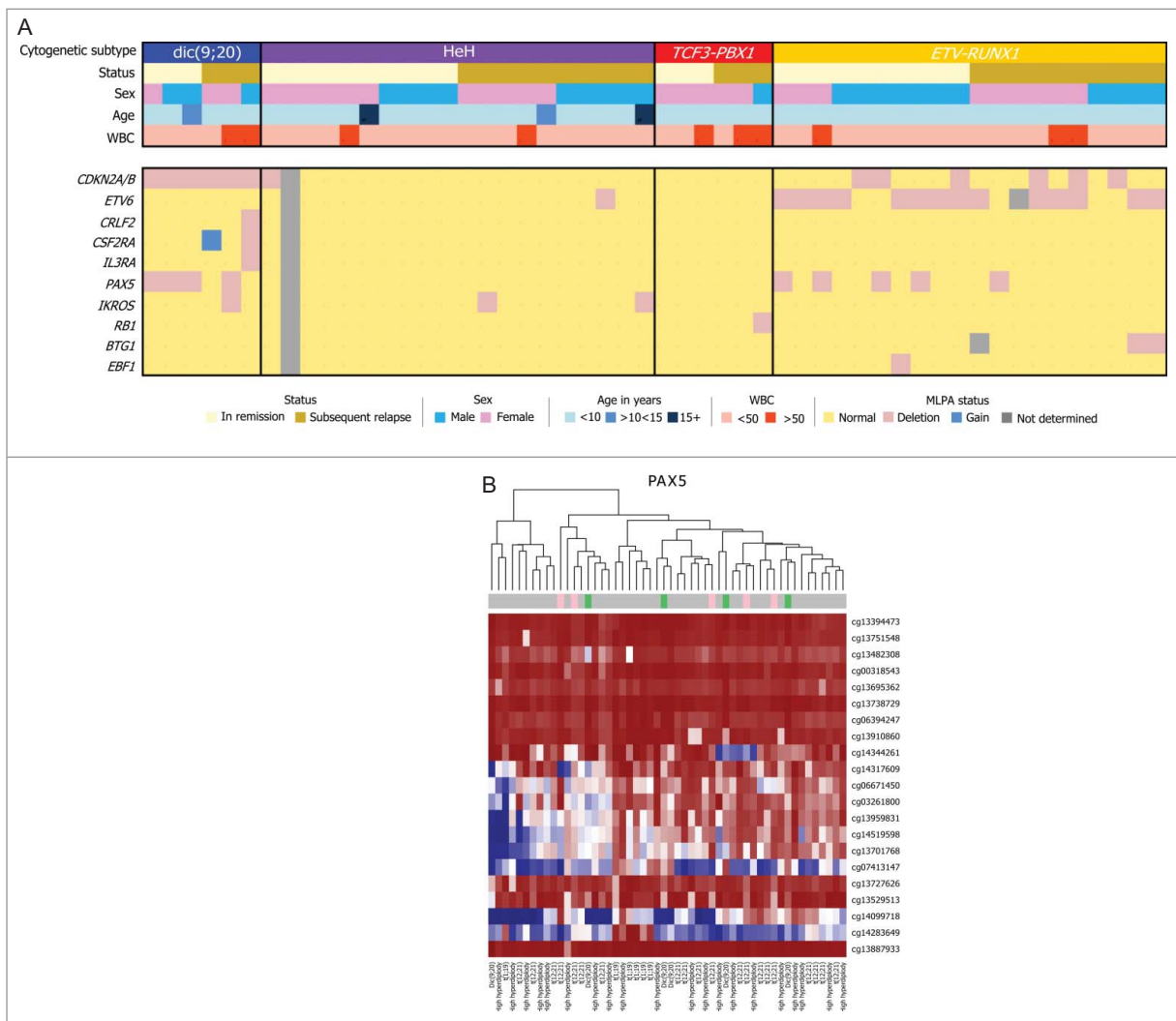


Figure 2. (A) Demographic and clinical features of 52 diagnostic bone marrow samples. Abbreviations: WBC, white blood cell; NCI risk, national cancer institute risk; SR, standard risk; HR, hazard ratio; HeH, High hyperdiploidy. Some of the gene aberrations listed are linked to the primary genetic aberrations (i.e., *CDKN2A/B* and *PAX5* in *dic(9;20)*), rather than true focal aberrations. Similarly, gene aberrations resulting from whole chromosome gain (HeH) or loss have not been shown. **(B)** Deletion of *PAX5* was observed in 5 patients via MLPA shown in pink blocks above. We clustered the data by looking at methylation probes 5 kb upstream and downstream of *PAX5*. The result shows that deletion status of *PAX5* does not seem to correlate with methylation values and seems independent of copy number.

proximity and number of differentially methylated probes, rather than by fixed size. *P*-values were also adjusted to control the false discovery rate (FDR) using the Benjamini–Hochberg method. However, neither approach identified regions in which the levels of methylation were significantly different between samples from the 2 different outcome categories at 5% FDR and 10% family-wise error rate (FWER), when all samples were analyzed simultaneously (Table 1). This was also largely true when the HeH and *ETV6-RUNX1* groups were analyzed separately, although a weak association was found at the *EXT1* loci specifically in the HeH subgroup (Table 1). Genetic subtype specific analysis was again restricted to *ETV6-RUNX1* and HeH subgroups. Single gene analysis utilizing pyrosequencing validated the array-generated data at the *EXT1* locus; however, expansion of the analysis to additional

diagnostic samples did not support an association between *EXT1* methylation and subsequent relapse (Fig. S4).

A recent study by Nordlund et al.²⁰ reported genome-wide methylation patterns for multiple childhood ALL cytogenetic subgroups, including the 4 included in this study. Similar to the results reported above, they also identified large-scale differences in DNA methylation between different cytogenetic subgroups. To assess the reproducibility of the genetic subgroup-specific methylation profiles identified by Nordlund et al., we determined whether the CpG marker sets identified as specific for individual cytogenetic subgroups in that study would also identify individual cytogenetic subgroups in our data set. As shown in Figure 3, all 4 cytogenetic subgroups were identified using their markers sets [with 54.0% (1141 / 2114), 37.7% (1136 / 3014), 28.3% (314 / 1110), 14.9% (353 / 2370) subtype specific CpG

Table 1. Differentially methylated regions identified by Bump Hunter analysis

All samples								
Chr	Start	End	No. of CpG sites	P-value	FWER ¹	Width (bps)	Nearest Gene	Distance to TSS ²
chr10	134765033	134765099	3	0.00029	0.354	67	<i>TTC40</i>	-8944
chr8	119124051	119124311	4	0.00049	0.443	261	<i>EXT1</i>	0
chr21	38468606	38468606	1	0.00042	0.750	1	<i>TTC3</i>	10516
chr2	77235218	77235218	1	0.00066	0.883	1	<i>LRRTM4</i>	514284
chr10	675888	675937	3	0.00242	0.887	50	<i>DIP2C</i>	59671
chr17	68164468	68164468	1	0.00093	0.934	1	<i>KCNJ2-AS1</i>	1075
chr9	124022933	124022933	1	0.00101	0.942	1	<i>GSN</i>	59172
chr3	168308798	168308798	1	0.00118	0.962	1	<i>EGFEM1P</i>	59276
chr6	32774788	32774788	1	0.00121	0.963	1	<i>HLA-DOB</i>	10037
chr1	248020692	248021091	4	0.00483	0.971	400	<i>TRIM58</i>	191
chr10	134765033	134765099	3	0.00029	0.354	67	<i>TTC40</i>	-8944
<i>ETV6-RUNX1</i>								
Chr	Start	End	No. of CpG sites	P-value	FWER	Width (bps)	Nearest Gene	Distance to TSS
chr5	158086454	158086454	1	1.64E-05	0.063	1	<i>EBF1</i>	437160
chr15	69744390	69744684	4	7.88E-05	0.236	295	<i>RPLP1</i>	-475
chr8	1651128	1651128	1	7.30E-05	0.242	1	<i>DLGAP2</i>	201596
chr4	134069593	134070441	10	0.00015	0.404	849	<i>PCDH10</i>	-29
chr5	178986620	178986906	5	0.00037	0.670	287	<i>RUFY1</i>	0
chr11	100760935	100760935	1	0.00031	0.693	1	<i>ARHGAP42</i>	202528
chr6	160023581	160024144	6	0.0009	0.898	564	<i>SOD2</i>	90209
chr21	46077454	46077731	6	0.0011	0.924	278	<i>TSPEAR</i>	53764
chr14	76015669	76015669	1	0.00079	0.946	1	<i>BATF</i>	26885
chr10	124639132	124639260	7	0.00135	0.952	129	<i>FAM24B</i>	0
HeH								
Chr	Start	End	No. of CpG sites	P-value	FWER	Width (bps)	Nearest Gene	Distance to TSS
chr8	119124051	119124462	5	1.58E-06	0.006	412	<i>EXT1</i>	0
chr10	675888	675937	3	0.00019	0.380	50	<i>DIP2C</i>	59671
chr16	66458043	66458043	1	0.00021	0.560	1	<i>BEAN1</i>	-3157
chr4	99850801	99851281	9	0.00047	0.625	481	<i>EIF4E</i>	505
chr22	25201958	25202163	6	0.00053	0.664	206	<i>SGSM1</i>	0
chr17	76875678	76876239	4	0.00055	0.669	562	<i>TIMP2</i>	42244
chr10	113944114	113944114	1	0.00031	0.695	1	<i>GPAM</i>	-577
chr5	150284600	150284796	2	0.00065	0.786	197	<i>ZNF300</i>	-55
chr1	248020377	248021091	8	0.00083	0.802	715	<i>TRIM58</i>	0
chr8	19459672	19460243	4	0.00089	0.807	572	<i>CSGALNACT1</i>	0
chr8	119124051	119124462	5	1.58E-06	0.006	412	<i>EXT1</i>	0

¹P-Value corrected for family-wise error rate, with B = 1000 permutations.²Transcriptional start site.

markers concordant with our analysis for *ETV6-RUNX1*, HeH, *TCF3-PBX1*, and *dic(9;20)* respectively, using the same criteria for defining subtype specific methylated loci].²⁰

Nordlund et al. also identified a set of 90 DMCs associated with relapse, specifically in *ETV6-RUNX1* positive cases. To validate this marker set for the identification of subsequent relapse, we applied these 90 relapse predicting DMCs to our datasets. However, as shown in Figure 3E, unsupervised clustering using the same 90 CpG sites, did not appear to differentiate between relapse and non-relapse samples (specifically in *ETV6-RUNX1* positive cases). To determine whether single CpG sites from within this group of 90 sites were associated with relapse, each locus was assessed individually in our *ETV6-RUNX1* positive sample set. However, only one of the 90 loci exhibited a statistically significant association with subsequent relapse in our sample set (cg17033047 within the *KCNA3* locus, $P = 0.01$, uncorrected P -value, higher methylation levels in relapse samples). Expanding the analysis for this site to an additional 57 *ETV6-*

RUNX1 positive cases (of which 3 relapsed) failed to confirm the differential methylation seen in the 20 samples used for the array analysis (Fig. S4D). As the size of the *ETV6-RUNX1* positive sample set was small ($n = 20$) with only 10 relapsed *ETV6-RUNX1* cases, the possibility cannot be ruled out that weak correlations may be detectable in larger sample sets. However, it should also be noted that our data set had more relapsed *ETV6-RUNX1* cases than Nordlund et al.²⁰ Taken together, these results suggest that DNA methylation at these loci is unlikely to be of significant clinical utility for the prediction of relapse in *ETV6-RUNX1* positive childhood ALL.

Discussion

Alterations in DNA methylation are highly prevalent in childhood ALL, suggesting that they may have a major impact on the biology and clinical behavior of the disease, as well as raising the

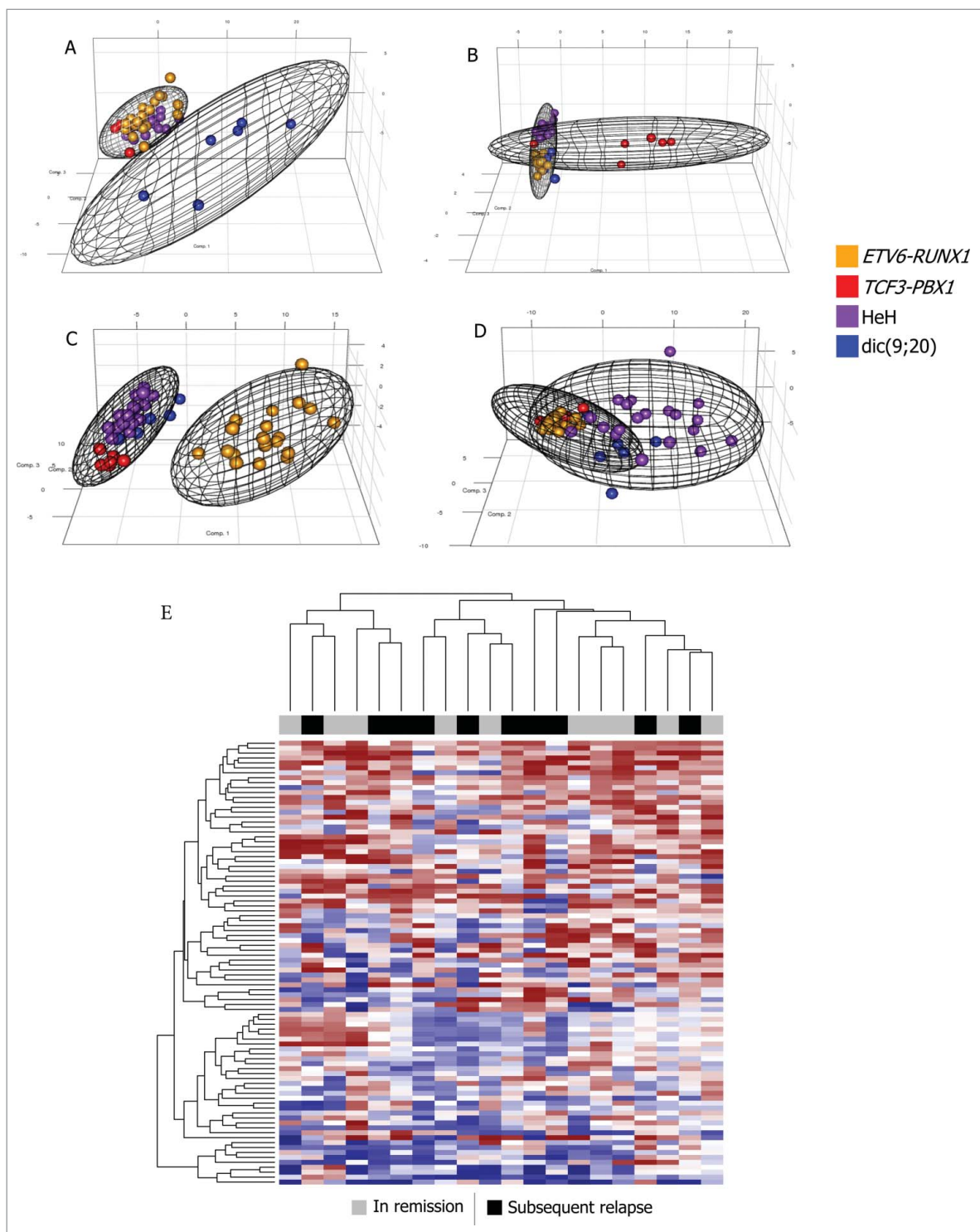


Figure 3. Principal component analysis of subtype-specific DMCs identified by Nordlund et al. recapitulates genetic subgroup separation and validates them as biomarkers for these subgroups. For each plot, 2 covariance spheroids have been plotted along the 95% confidence intervals for (A) dic(9;20) and others; (B) TCF3-PBX1 and others; (C) ETV6-RUNX1 and others; (D) HeH and others. Individual samples are colored by their cytogenetic status; TCF3-PBX1 cases are shown red, ETV6-RUNX1 in orange, dic(9;20) in blue and HeH in purple. (E) Previously reported relapse-associated DMCs in ETV6-RUNX1 positive cases are not recapitulated in our data set. The color bar at the top of the heatmap indicates sample type; continuous remission and subsequently relapsing patients are shown gray and black respectively. The heatmap shows the methylation status for 90 relapse-associated probes identified by Nordlund et al.²⁰ Samples (columns) and probes (rows) were clustered using complete linkage and Euclidean distance. Fully methylated probes are shown dark red, unmethylated probes shown dark blue and hemi-methylated probes are shown in white.

possibility that differences in DNA methylation patterns at diagnosis may be useful biomarkers for prediction of clinical outcome. In this study, genome-wide methylation analysis was carried out on 52 diagnostic ALL samples from 4 cytogenetic subgroups, in which 50% of patients subsequently relapsed and 50% remained in long-term remission. NMF consensus clustering identified multiple sub-groups within the methylation data; however, these were related to the underlying genetic differences and did not differentiate between samples with different relapse status. Different bioinformatics approaches were undertaken in an attempt to identify a single or a small number of methylation variable sites that could identify at diagnosis those patients most likely to relapse. However, none of these approaches identified any loci whose methylation status was significantly associated with subsequent relapse. Some limited evidence for an association with relapse was found at sites within the *EXT1* and *KCNA3* loci in the array data; however, neither were confirmed by pyrosequencing analysis of additional samples. The results presented here, in combination with previously published data,²⁰ suggest that genome-wide methylation profiles, identified using the Infinium HumanMethylation450 BeadChip array platform, may be unlikely to yield clinically useful biomarkers for prediction of relapse in childhood ALL over and above the prognostic information already provided by cytogenetic subgroups.

However, there was a clear correlation between genome-wide patterns of DNA methylation and the different cytogenetic subgroups, consistent with previous studies.^{15,20-23} In addition, we were able to use the marker sets recently identified by Nordlund et al.,²⁰ whose analysis utilized the same array platform, on our data set and validate their cytogenetic specific markers. Thus, while DNA methylation profiles did not appear to augment the prognostic information provided by standard cytogenetics, the consistency of the methylation changes in relation to cytogenetic subgroups suggests that these altered patterns of DNA methylation may be an important determinant of their different clinical behavior. Furthermore, the results suggest that genomic DNA methylation could be used as a surrogate for cytogenetic analysis in cases where genomic DNA, but no intact cells, was available.

While methylation profiles associated with cytogenetic subgroups could be validated, the marker set suggested to predict outcome in *ETV6-RUNX1* positive cases did not validate in our data set. A potential cause would be that the patients were treated on different protocols, although the treatment protocols used were highly similar.^{28,29} In addition, the patient populations were derived from different geographical locations, so differences in genetic background may have also played a role. However, the results do indicate that the identified methylation profile is not readily portable to other patient populations.

Although the cytogenetic subtypes showed a clear correlation with methylation profiles, the data presented here and previously published^{20,23} showed that this clustering was not absolute. For example, in the data set presented here, one dic(9;20) and one HeH case clustered with the *TCF3-PBX1* samples. In such cases, it is not clear whether the risk of relapse in the patients would

reflect their cytogenetic subgroup or be equivalent to the subgroup defined by the methylation profile. However, the comparative rarity of these cases (only 2/52 samples in this study) meant that a much larger study would be required to have sufficient power to address the potential prognostic significance of such "outlier" samples.

While analysis of DNA methylation profiles has identified subtype specific methylation changes, it is important to note that many of the alterations identified in this and other studies are shared across cytogenetic subgroups. This implies that a large set of epigenetic changes are either a prerequisite for, or an inevitable consequence of, the development of ALL. In general, the consistency of alterations seen in ALL and indeed other tumors has led to the hypothesis that cancer, in many instances, may be initiated specifically from a set of cells that have already undergone extensive epigenetic changes.³⁰ Thus, analysis of the targets for altered DNA methylation that are conserved across all ALL subtypes may be able to identify key drivers of the disease that could be targets for the development of novel therapeutic approaches.

The cytogenetic subtype specificity of many of the methylation changes suggests that these differential methylation patterns may be important in the different biological/clinical behavior of the different cytogenetic groups. Here we used gene enrichment pathway analysis to investigate whether subtype specific DMCs might be preferentially targeting specific biological pathways. Our observation of gene enrichment pathways in *ETV6-RUNX1* involving immunity and infection pathways (8/10 pathways with a significant association are related to immune function) is a potentially exciting avenue for future analysis, as it may relate to previous reports that have suggested abnormal immune response as a major factor shaping the trajectory of leukemogenesis.^{31,32} In the *TCF3-PBX1* subtype, we identified a remarkable enrichment for genes in the neuroactive ligand-receptor interaction pathway. Interestingly, a large fraction of these methylation changes map to gene promoter regions, suggesting that they are likely to be associated with functional changes in gene expression. Previous reports have shown that neuroactive ligand-receptor interaction pathways are associated with acute leukemias as well as several other diseases.^{33,34} Further investigation of the potential role of this pathway in *TCF3-PBX1* driven leukemia would be warranted. Integrating the methylation data with gene expression data would help clarify whether the subtype specific patterns of methylation correlate with clear differences in subtype specific gene expression and thus potentially with downstream gene function.

This study focused specifically on DNA methylation. To more clearly understand the epigenome of pediatric ALL it may be necessary to undertake integrative studies assessing other epigenetic mechanisms, such as nucleosome remodeling and histone modifications, as well as associations with microRNA and gene expression in the same sample sets. In addition, while the Infinium HumanMethylation450 BeadChip array platform used in this and other studies²⁰ has coverage of much of the genome, including 99% of RefSeq genes, it only contains probes for about 2% of the total number of CpG sites in the human genome. Some studies have used whole-genome bisulfite sequencing in an

attempt to address this limitation. However, at this time, it is challenging to apply this approach to more than just a small number of samples due to cost and increased DNA requirement.³⁵⁻³⁷ Thus, further technical developments in genome-wide bisulfite sequencing and integration with other epigenetic mechanisms will be required to allow the identification of a complete picture of the epigenetic changes in childhood ALL and how this relates to changes in gene expression profiles. Such integrative studies may reveal biologically relevant epigenetic changes.

Materials and Methods

Patients and sorted cells

Bone marrow samples from 52 pediatric patients with *ETV6-RUNX1* ($n = 20$), high hyperdiploidy (51–65 chromosomes) (HeH) ($n = 20$), *TCF3-PBX1* ($n = 6$) and *dic(9;20)(p11-13;q11)* ($n = 6$) who consented to be enrolled on the ethically-approved UK ALL treatment trial, ALL97/99. All samples used had a high blast count (average 93%). An additional 123 diagnostic bone marrow samples from MRC ALL97/99 (HeH, $n = 66$ and *ETV6-RUNX1*, $n = 57$) were used in confirmatory pyrosequencing analysis.

MLPA

Genomic DNA from patient bone marrow aspirates was extracted using standard procedures. Genomic DNA from healthy donors was obtained for use as control samples. DNA was analyzed using the SALSA MLPA Kit P335 (MRC Holland, Amsterdam, The Netherlands), as described previously.²⁵ This kit includes probes for *IKZF1* (8 probes), *CDKN2A/B* (3 probes), *PAX5* (6 probes), *EBF1* (4 probes), *ETV6* (6 probes), *BTG1* (4 probes), *RBI* (5 probes), and the *PAR1* region: *CRLF2*, *CSF2RA*, and *IL3RA* (one probe each). Data were analyzed using GeneMarker V1.85 analysis software (SoftGenetics). All loci were found to be deleted in at least one patient and the majority of patients (40/52) had deletion of one or more of the genes assessed.

Bisulfite conversion and 450K array hybridization

Bisulfite conversion was performed using the Zymo EZ-96 DNA methylation kit and the bisulfite converted DNA was hybridized to the HumanMethylation450 Analysis BeadChip (Illumina) and processed following the 450K methylation array procedure, according to manufacturer's instructions. Hybridization fluorescent signals were read by the Illumina BeadStation GX scanner. This procedure was performed at Wellcome Trust Clinical Research Facility, Edinburgh, UK.

Bioinformatics analyses

The arrays report DNA methylation status (β value) at $> 485,000$ CpG loci. The β value can range from 0 to 1, representing fully unmethylated and methylated values. Array processing, normalization and quality control checks, as well as derivation of the β values from the raw intensity values

(.idat files), were implemented using the R package 'minfi'.³⁸ We employed conservative quality control measures to filter out poorly performing and potentially confounding loci. Briefly, a filtering process removed unannotated probes (i.e., not mapped to the genome), probes located on chromosomes X/Y, probes that aligned to more than one place in the genome, allowing for 1 mismatch, and probes that had a SNP with a minor allele frequency of 5% or greater within 50 bp of the interrogated site. Probes that failed in $>5\%$ of samples were also removed.³⁹⁻⁴² The remaining probe β values (429,750) were converted to *M*-scores,⁴³ and the top 10,000 most variable probes by standard deviation were selected for subgroup identification. The 10,000 most variable probes were used for the clustering as this is equivalent to the inflection point in the curve (when mapping variability versus probe number), such that probes that were excluded were largely non-variable and would have added little to the clustering. A non-negative matrix factorization (NMF),⁴⁴ based consensus clustering approach was performed using the R package NMF.⁴⁵ An optimal factor number (i.e., subgroup number) was selected by maximizing cluster number while maintaining cophenetic correlation coefficient. Cluster stability measures (silhouette scores) were used to assess the quality of the identified subgroups. Pathway enrichment analysis was performed using 'Goseq' bioconductor package using the KEGG database. *P*-values were calculated by using both resampling and the Wallenius approximation based methods available in Goseq.⁴⁶ *P*-values were also adjusted to control the false discovery rate (FDR) using the Benjamini–Hochberg method. The list of genes associated with probes was derived from the annotation provided by Illumina. The DNA methylation dataset is available at the Gene Expression Omnibus (GEO) with accession number GSE69229.

Quantitative DNA methylation analysis using pyrosequencing

Genomic DNA (200 ng) was modified with sodium bisulfite using the MethylampTM One-Step DNA Modification Kit (Epigentek, Brooklyn, NY, USA) as per the manufacturer's instructions. All samples were re-suspended in 15 μ l of TE, and 1 μ l of this was used for subsequent PCR reactions. DNA samples were amplified in 25 μ l volumes containing 1X manufacturer's buffer, 1 unit of FastStart Taq polymerase (Roche, Welwyn Garden City, UK), 1–4 mM MgCl₂, 10 mM dNTPs, and 75 ng of each primer. PCR was performed for 40 cycles with an annealing temperature of 53–63°C, depending on the primer set being used. For each set of primers (listed in Supplementary Table 3) one of the forward or reverse primers included a 5' biotin label to allow for subsequent analysis by pyrosequencing. Following PCR amplification, sequencing was performed using a PSQ 96MA pyrosequencer (Qiagen, Hilden, Germany), as per manufacturer's protocol. For all loci, assays were performed in duplicate and values averaged between the duplicates. Only samples that were passed by the pyrosequencer were included and to further ensure a high degree of accuracy only runs in which single peak heights were in excess of 200 were included. For CpG island loci, between 3 and 6 consecutive CpG sites were measured and

the methylation value for each locus was taken as the mean of all CpG sites measured at that locus. For non CpG island associated CpG sites, pyrosequencing assays were designed to include that single specific CpG site. Primer design was performed using the manufacturer's provided PyroMark software and all pyrosequencing runs included *in vitro* methylated DNA (Millipore, Watford, UK) and normal peripheral blood derived DNA as control.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank member laboratories of the UK Cancer Cytogenetic Group for providing cytogenetic data and material as well as past and present members of the Leukemia Research Cytogenetics Group for their contribution to establishing this data set. Primary childhood leukemia samples used in this study

were provided by the Leukemia and Lymphoma Research Childhood Leukemia Cell Bank working with the laboratory teams in the Bristol Genetics Laboratory, Southmead Hospital, Bristol; Molecular Biology Laboratory, Royal Hospital for Sick Children, Glasgow; Molecular Haematology Laboratory, Royal London Hospital, London; and Molecular Genetics Service and Sheffield Children's Hospital, Sheffield, UK.

Funding

This work was supported by a PhD fellowship awarded to FML by the Iraqi Ministry of Higher Education and Scientific Research (MOHESR) as well as by a specialist program grant awarded by Leukemia & Lymphoma Research to CJH and AVM.

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website.

References

- Shah A, Coleman MP. Increasing incidence of childhood leukaemia: a controversy re-examined. *Br J Cancer* 2007; 97:1009-12; PMID:17712312
- Vora A, Goulden N, Wade R, Mitchell C, Hancock J, Hough R, Rowntree C, Richards S. Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial. *Lancet Oncol* 2013; 14:199-209; PMID:23395119; [http://dx.doi.org/10.1016/S1470-2045\(12\)70600-9](http://dx.doi.org/10.1016/S1470-2045(12)70600-9)
- Vora A, Goulden N, Mitchell C, Hancock J, Hough R, Rowntree C, Moorman AV, Wade R. Augmented post-remission therapy for a minimal residual disease-defined high-risk subgroup of children and young people with clinical standard-risk and intermediate-risk acute lymphoblastic leukaemia (UKALL 2003): a randomised controlled trial. *Lancet Oncol* 2014; 15:809-18; PMID:24924991; [http://dx.doi.org/10.1016/S1470-2045\(14\)70243-8](http://dx.doi.org/10.1016/S1470-2045(14)70243-8)
- Izraeli S. Application of genomics for risk stratification of childhood acute lymphoblastic leukaemia: from bench to bedside? *Br J Haematol* 2010; 151:119-31; PMID:20678159; <http://dx.doi.org/10.1111/j.1365-2141.2010.08312.x>
- Moorman AV, Enshaie A, Schwab C, Wade R, Chilton L, Elliott A, Richardson S, Hancock J, Kinsey SE, Mitchell CD, et al. A novel integrated cytogenetic and genomic classification refines risk stratification in pediatric acute lymphoblastic leukemia. *Blood* 2014; 124:1434-44; PMID:24957142; <http://dx.doi.org/10.1182/blood-2014-03-562918>
- Mullighan CG. Genomic characterization of childhood acute lymphoblastic leukemia. *Semin Hematol* 2013; 50:314-24; PMID:24246699; <http://dx.doi.org/10.1053/j.seminhematol.2013.10.001>
- Bird AP. The relationship of DNA methylation to cancer. *Cancer Surv* 1996; 28:87-101; PMID:8977030
- Bird AP, Wolffe AP. Methylation-induced repression—belts, braces, and chromatin. *Cell* 1999; 99:451-4; PMID:10589672; [http://dx.doi.org/10.1016/S0092-8674\(00\)81532-9](http://dx.doi.org/10.1016/S0092-8674(00)81532-9)
- Kulis M, Esteller M. DNA methylation and cancer. *Adv Genet* 2010; 70:27-56; PMID:20920744
- Jones PA, Laird PW. Cancer epigenetics comes of age. *Nat Genet* 1999; 21:163-7; PMID:9988266; <http://dx.doi.org/10.1038/5947>
- Tost J. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Mol Biotechnol* 2010; 44:71-81; PMID:19842073; <http://dx.doi.org/10.1007/s12033-009-9216-2>
- Weller M, Stupp R, Reifenberger G, Brandes AA, van den Bent MJ, Wick W, Hegi ME. MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nat Rev* 2010; 6:39-51; PMID:19997073
- Irving L, Mainou-Fowler T, Parker A, Ibbotson RE, Oscier DG, Strathdee G. Methylation markers identify high risk patients in IGHV mutated chronic lymphocytic leukemia. *Epigenetics* 2011; 6; PMID:21051931; <http://dx.doi.org/10.4161/epi.6.3.14038>
- Kraszevska MD, Dawidowska M, Larmonie NS, Kosmalka M, Sedek L, Szczepaniak M, Grzeszczak W, Langerak AW, Szczepanski T, Witt M. DNA methylation pattern is altered in childhood T-cell acute lymphoblastic leukemia patients as compared with normal thymic subsets: insights into CpG island methylator phenotype in T-ALL. *Leukemia* 2012; 26:367-71; <http://dx.doi.org/10.1038/leu.2011.208>
- Milani L, Lundmark A, Kiialainen A, Nordlund J, Flaegstad T, Forestier E, Heyman M, Jonmundsson G, Kanerva J, Schmiegelow K, et al. DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood* 2010; 115:1214-25; PMID:19965625; <http://dx.doi.org/10.1182/blood-2009-04-214668>
- Paulsson K, An Q, Moorman AV, Parker H, Molloy G, Davies T, Griffiths M, Ross FM, Irving J, Harrison CJ, et al. Methylation of tumour suppressor gene promoters in the presence and absence of transcriptional silencing in high hyperdiploid acute lymphoblastic leukaemia. *Br J Haematol* 2009; 144:838-47; PMID:19120349; <http://dx.doi.org/10.1111/j.1365-2141.2008.07523.x>
- Stumpel DJ, Schneider P, van Roon EH, Boer JM, de Lorenzo P, Valsecchi MG, de Menezes RX, Pieters R, Stam RW. Specific promoter methylation identifies different subgroups of MLL-rearranged infant acute lymphoblastic leukemia, influences clinical outcome, and provides therapeutic options. *Blood* 2009; 114:5490-8; PMID:19855078; <http://dx.doi.org/10.1182/blood-2009-06-227660>
- Borssen M, Palmqvist L, Karrman K, Abrahamsson J, Behrendtz M, Heldrup J, Forestier E, Roos G, Degerman S. Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. *PLoS One* 2013; 8:e65373; PMID:23762353; <http://dx.doi.org/10.1371/journal.pone.0065373>
- Sandoval J, Heyn H, Mendez-Gonzalez J, Gomez A, Moran S, Baiget M, Melo M, Badell I, Nomdedeu JF, Esteller M. Genome-wide DNA methylation profiling predicts relapse in childhood B-cell acute lymphoblastic leukaemia. *Br J Haematol* 2013; 160:406-9; PMID:23110451; <http://dx.doi.org/10.1111/bjh.12113>
- Nordlund J, Backlin CL, Wahlberg P, Busche S, Berglund EC, Eloranta ML, Flaegstad T, Forestier E, Frost BM, Harila-Saari A, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol* 2013; 14:r105; PMID:24063430; <http://dx.doi.org/10.1186/gb-2013-14-9-r105>
- Davidsson J, Liljebjorn H, Andersson A, Veerla S, Heldrup J, Behrendtz M, Fioretos T, Johansson B. The DNA methylome of pediatric acute lymphoblastic leukemia. *Hum Mol Genet* 2009; 18:4054-65; PMID:19679565; <http://dx.doi.org/10.1093/hmg/ddp354>
- Busche S, Ge B, Vidal R, Spinella JF, Saillour V, Richer C, Healy J, Chen SH, Droit A, Sinnenet D, et al. Integration of high-resolution methylome and transcriptome analyses to dissect epigenomic changes in childhood acute lymphoblastic leukemia. *Cancer Res* 2013; 73:4323-36; PMID:23722552; <http://dx.doi.org/10.1158/0008-5472.CAN-12-4367>
- Figuerola ME, Chen SC, Andersson AK, Phillips LA, Li Y, Sotzen J, Kundu M, Downing JR, Melnick A, Mullighan CG. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *J Clin Invest* 2013; 123:3099-111; PMID:23921123; <http://dx.doi.org/10.1172/JCI66203>
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al. High density DNA methylation array with single CpG site resolution. *Genomics* 2011; 98:288-95; PMID:21839163; <http://dx.doi.org/10.1016/j.ygeno.2011.07.007>
- Schwab CJ, Jones LR, Morrison H, Ryan SL, Yigitop H, Schouten JP, Harrison CJ. Evaluation of multiplex ligation-dependent probe amplification as a method for the detection of copy number abnormalities in B-cell precursor acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 2010; 49:1104-13; <http://dx.doi.org/10.1002/gcc.20818>
- Kwee I, Rinaldi A, Rancoita P, Rossi D, Capello D, Forconi F, Giuliani N, Piva R, Inghirami G, Gaidano

- G, et al. Integrated DNA copy number and methylation profiling of lymphoid neoplasms using a single array. *Br J Haematol* 2012; 156:354-7; PMID:22118580; <http://dx.doi.org/10.1111/j.1365-2141.2011.08946.x>
27. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012; 41:200-9; PMID:22422453; <http://dx.doi.org/10.1093/ije/dyr238>
28. Mitchell C, Payne J, Wade R, Vora A, Kinsey S, Richards S, Eden T. The impact of risk stratification by early bone-marrow response in childhood lymphoblastic leukaemia: results from the United Kingdom Medical Research Council trial ALL97 and ALL97/99. *Br J Haematol* 2009; 146:424-36; PMID:19549269; <http://dx.doi.org/10.1111/j.1365-2141.2009.07769.x>
29. Schmiegelow K, Forestier E, Hellebostad M, Heyman M, Kristinsson J, Soderhall S, Taskinen M, Nordic Society of Paediatric H, Oncology. Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia* 2010; 24:345-54; PMID:20010622; <http://dx.doi.org/10.1038/leu.2009.251>
30. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* 2006; 7:21-33; PMID:16369569; <http://dx.doi.org/10.1038/nrg1748>
31. Kinlen LJ. Epidemiological evidence for an infective basis in childhood leukaemia. *Br J Cancer* 1995; 71:1-5; PMID:7819022; <http://dx.doi.org/10.1038/bjc.1995.1>
32. O'Connor SM, Boneva RS. Infectious etiologies of childhood leukemia: plausibility and challenges to proof. *Environ Health Perspect* 2007; 115:146-50; PMID:17366835; <http://dx.doi.org/10.1289/ehp.9024>
33. Loudin MG, Wang J, Leung HC, Gurusiddappa S, Meyer J, Condos G, Morrison D, Tsimelzon A, Devadas M, Heerema NA, et al. Genomic profiling in Down syndrome acute lymphoblastic leukemia identifies histone gene deletions associated with altered methylation profiles. *Leukemia* 2011; 25:1555-63; <http://dx.doi.org/10.1038/leu.2011.128>
34. Ding Y, Chen M, Liu Z, Ding D, Ye Y, Zhang M, Kelly R, Guo L, Su Z, Harris SC, et al. atBioNet—an integrated network analysis tool for genomics and biomarker discovery. *BMC Genomics* 2012; 13:325; PMID:22817640; <http://dx.doi.org/10.1186/1471-2164-13-325>
35. Kreck B, Richter J, Ammerpohl O, Barann M, Esser D, Petersen BS, Vater I, Murga Penas EM, Bormann Chung CA, Seisenberger S, et al. Base-pair resolution DNA methylome of the EBV-positive Endemic Burkitt lymphoma cell line DAUDI determined by SOLiD bisulfite-sequencing. *Leukemia* 2013; 27:1751-3; PMID:23307032; <http://dx.doi.org/10.1038/leu.2013.4>
36. Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, Han X, Chen Q, Zhang X, Wang J. Whole genome DNA methylation analysis based on high throughput sequencing technology. *Methods* 2010; 52(3):203-12
37. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009; 462:315-22; PMID:19829295; <http://dx.doi.org/10.1038/nature08514>
38. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014; 30:1363-9; PMID:24478339; <http://dx.doi.org/10.1093/bioinformatics/btu049>
39. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 2013; 8:203-9; PMID:23314698; <http://dx.doi.org/10.4161/epi.23470>
40. Naeem H, Wong NC, Chatterton Z, Hong MK, Pedersen JS, Corcoran NM, Hovens CM, Macintyre G. Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. *BMC Genomics* 2014; 15:51; PMID:24447442; <http://dx.doi.org/10.1186/1471-2164-15-51>
41. Pajtler KW, Witt H, Sill M, Jones DT, Hovestadt V, Kratochwil F, Wani K, Tatevossian R, Punchihewa C, Johann P, et al. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. *Cancer Cell* 2015; 27:728-43; PMID:25965575; <http://dx.doi.org/10.1016/j.ccr.2015.04.002>
42. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, Konermann C, Pfaff E, Tonjes M, Sill M, Bender S, et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* 2012; 22:425-37; PMID:23079654; <http://dx.doi.org/10.1016/j.ccr.2012.08.024>
43. Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 2006; 7:55-65; PMID:16369572; <http://dx.doi.org/10.1038/nrg1749>
44. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* 2004; 101:4164-9; PMID:15016911; <http://dx.doi.org/10.1073/pnas.0308531101>
45. Schwalbe EC, Hayden JT, Rogers HA, Miller S, Lindsey JC, Hill RM, Nicholson SL, Kilday JP, Adamowicz-Brice M, Storer L, et al. Histologically defined central nervous system primitive neuro-ectodermal tumours (CNS-PNETs) display heterogeneous DNA methylation profiles and show relationships to other paediatric brain tumour types. *Acta Neuropathol* 2013; 126:943-6; PMID:24212602; <http://dx.doi.org/10.1007/s00401-013-1206-6>
46. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 2010; 11:R14; PMID:20132535; <http://dx.doi.org/10.1186/gb-2010-11-2-r14>



Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study



Edward C Schwalbe, Janet C Lindsey, Sirintra Nakjang, Stephen Crosier, Amanda J Smith, Debbie Hicks, Gholamreza Rafiee, Rebecca M Hill, Alice Iliasova, Thomas Stone, Barry Pizer, Antony Michalski, Abhijit Joshi, Stephen B Wharton, Thomas S Jacques, Simon Bailey, Daniel Williamson, Steven C Clifford

Summary

Background International consensus recognises four medulloblastoma molecular subgroups: WNT (MB_{WNT}), SHH (MB_{SHH}), group 3 (MB_{Gp3}), and group 4 (MB_{Gp4}), each defined by their characteristic genome-wide transcriptomic and DNA methylomic profiles. These subgroups have distinct clinicopathological and molecular features, and underpin current disease subclassification and initial subgroup-directed therapies that are underway in clinical trials. However, substantial biological heterogeneity and differences in survival are apparent within each subgroup, which remain to be resolved. We aimed to investigate whether additional molecular subgroups exist within childhood medulloblastoma and whether these could be used to improve disease subclassification and prognosis predictions.

Methods In this retrospective cohort study, we assessed 428 primary medulloblastoma samples collected from UK Children's Cancer and Leukaemia Group (CCLG) treatment centres (UK), collaborating European institutions, and the UKCCSG-SIOP-PNET3 European clinical trial. An independent validation cohort ($n=276$) of archival tumour samples was also analysed. We analysed samples from patients with childhood medulloblastoma who were aged 0–16 years at diagnosis, and had central review of pathology and comprehensive clinical data. We did comprehensive molecular profiling, including DNA methylation microarray analysis, and did unsupervised class discovery of test and validation cohorts to identify consensus primary molecular subgroups and characterise their clinical and biological significance. We modelled survival of patients aged 3–16 years in patients ($n=215$) who had craniospinal irradiation and had been treated with a curative intent.

Findings Seven robust and reproducible primary molecular subgroups of childhood medulloblastoma were identified. MB_{WNT} remained unchanged and each remaining consensus subgroup was split in two. MB_{SHH} was split into age-dependent subgroups corresponding to infant (<4.3 years; $MB_{SHH-Infant}$; $n=65$) and childhood patients (≥ 4.3 years; $MB_{SHH-Child}$; $n=38$). MB_{Gp3} and MB_{Gp4} were each split into high-risk (MB_{Gp3-HR} [$n=65$] and MB_{Gp4-HR} [$n=85$]) and low-risk (MB_{Gp3-LR} [$n=50$] and MB_{Gp4-LR} [$n=73$]) subgroups. These biological subgroups were validated in the independent cohort. We identified features of the seven subgroups that were predictive of outcome. Cross-validated subgroup-dependent survival models, incorporating these novel subgroups along with secondary clinicopathological and molecular features and established disease risk-factors, outperformed existing disease risk-stratification schemes. These subgroup-dependent models stratified patients into four clinical risk groups for 5-year progression-free survival: favourable risk (54 [25%] of 215 patients; 91% survival [95% CI 82–100]); standard risk (50 [23%] patients; 81% survival [70–94]); high-risk (82 [38%] patients; 42% survival [31–56]); and very high-risk (29 [13%] patients; 28% survival [14–56]).

Interpretation The discovery of seven novel, clinically significant subgroups improves disease risk-stratification and could inform treatment decisions. These data provide a new foundation for future research and clinical investigations.

Funding Cancer Research UK, The Tom Grahame Trust, Star for Harris, Action Medical Research, SPARKS, The JGW Patterson Foundation, The INSTINCT network (co-funded by The Brain Tumour Charity, Great Ormond Street Children's Charity, and Children with Cancer UK).

Copyright © The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

Introduction

The discovery of molecular disease subgroups represents the most fundamental advance in our understanding of medulloblastoma, the most common malignant brain tumour of childhood. Current international consensus recognises four subgroups of medulloblastoma: WNT (MB_{WNT}), SHH (MB_{SHH}), group 3 (MB_{Gp3}) and group 4

(MB_{Gp4}).¹ Each subgroup is defined empirically by genome-wide transcriptomic^{2–6} and DNA methylation patterns^{7,8} and characterised by distinct clinicopathological and molecular features.^{9–12} MB_{WNT} and MB_{SHH} are synonymous with WNT and SHH activating mutations.¹² By contrast, MB_{Gp3} and MB_{Gp4} have few mutations, but have multiple DNA copy number alterations.^{9–12}

Lancet Oncol 2017; 18: 958–71

Published Online

May 22, 2017

[http://dx.doi.org/10.1016/S1470-2045\(17\)30243-7](http://dx.doi.org/10.1016/S1470-2045(17)30243-7)

See [Comment](#) page 847

Wolfson Childhood Cancer Research Centre, Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne, UK (E C Schwalbe PhD, J C Lindsey PhD, S Nakjang PhD, S Crosier PhD, A J Smith MSc, D Hicks PhD, G Rafiee PhD, R M Hill PhD, A Iliasova MSc, Prof S Bailey PhD, D Williamson PhD, Prof S C Clifford PhD); Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, UK (E C Schwalbe, A Iliasova);

Department of Haematology and Oncology Department (A Michalski PhD), Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK; Neural Development Unit, UCL Institute of Child Health, London, UK (T Stone MRes, T S Jacques PhD); Institute of Translational Research, University of Liverpool, Liverpool, UK (Prof B Pizer PhD); Department of Neuropathology, Royal Victoria Infirmary, Newcastle University Teaching Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK (A Joshi MBBS); and Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK (Prof S B Wharton PhD)

Correspondence to:

Prof Steven C Clifford, Wolfson Childhood Cancer Research Centre, Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne NE1 7RU, UK steve.clifford@ncl.ac.uk

Research in context

Evidence before this study

The international consensus definition of medulloblastoma, published in 2012, recognises four primary molecular subgroups with distinct clinicopathological features: WNT (MB_{WNT}), SHH (MB_{SHH}), group 3 (MB_{Grp3}), and group 4 (MB_{Grp4}). Several studies established characteristic genome-wide transcriptomic and DNA methylomic profiles, using unsupervised class discovery techniques, which defined the consensus subgroups. These subgroups, described in the 2016 WHO classification of brain tumours, underpin current disease subclassification, research studies, and clinical trials. Profiling and class discovery studies published to date in medulloblastoma are based on cohorts typically with sample sizes less than 200 patients and, even within the consensus subgroups, significant heterogeneity of clinical and molecular features remains and many relationships to disease outcome are unresolved. Evidence from the component studies and reviews undertaken in the international consensus definition and the 2016 WHO classification, alongside our own reviews of the current literature, formed the foundation for the present study; no systematic reviews were carried out.

Added value of this study

We defined and characterised seven robust, reproducible, clinically significant, primary molecular subgroups within childhood medulloblastoma (in children aged up to 16 years at diagnosis), each with distinct clinicomolecular features. We propose a cross-validated, subgroup-dependent survival model that incorporates these novel subgroups, alongside established disease features and risk-factors and outperforms the disease risk-stratification schemes in current clinical use. Redistribution of disease risk using this scheme identifies substantial proportions of favourable-risk non-infant patients (>90% 5-year survival in 11% of patients) outside the MB_{WNT} subgroup (equivalent to approximately 70 patients per year in the European Union [EU]) who would be suitable for consideration of reduced intensity of therapy, and very-high-risk non-infant patients (<40% survival, 13% of patients, about 80 EU patients per year) for whom new treatment strategies should be prioritised.

Implications of all the available evidence

These data provide a step-change in our understanding and characterisation of molecular subgroups within medulloblastoma, with potential application to future disease subclassification, risk-stratification, and subgroup-dependent translational research.

Subgrouping is integral to the 2016 WHO medulloblastoma classification,¹³ and is used to direct treatment strategies aimed at improving cure rates (5-year survival across all four subgroups is 65–70%), and reducing long-term intellectual and neuroendocrine impairments associated with existing multimodality therapies. Patients with childhood MB_{WNT} consistently show a favourable prognosis (>90% survival^{14,15}) and reduced intensity risk-adapted therapies are being studied in these patients in international clinical trials,¹⁶ whereas SHH pathway inhibitors show promise in MB_{SHH} disease in early-phase trials,¹⁷ although treatment of infants (younger than 3 years at diagnosis) and young children with these inhibitors should be approached with caution, because of the risk of premature fusion of growth-plates.¹⁸

Substantial biological heterogeneity is evident within each non- MB_{WNT} subgroup; for instance, *TP53* mutations are associated with a poor outcome in MB_{SHH} .^{13,19} High-risk clinical factors (metastatic disease [M+]; large-cell, anaplastic [LCA] pathology; incomplete surgical resection [R+]; and *MYC/MYCN* amplification), which are currently used to stratify risk in medulloblastoma in children aged 3 years or older, were derived from cohort-wide investigations before discovery of the consensus subgroups, and thus did not consider their effect.^{15,16,20,21}

Studies that defined the four-subgroup consensus used modestly sized cohorts (typically fewer than 200 patients).^{2–6} In this Article, we describe comprehensive molecular profiling of clinically annotated discovery and validation cohorts totalling more than 700 tumours.

We report the discovery and characterisation of seven stable and reproducible primary subgroups of childhood medulloblastoma (in patients younger than 16 years at diagnosis), which subdivide each of the classic consensus non- MB_{WNT} subgroups (MB_{SHH} , MB_{Grp3} , and MB_{Grp4}) into two clinically significant groups with distinct clinicomolecular features and survival outcomes.

Methods

Study design and participants

In this retrospective cohort study, we assessed 428 centrally reviewed, clinically annotated primary medulloblastomas from patients aged 0–16 years at diagnosis, collected from UK Children's Cancer and Leukaemia Group (CCLG) treatment centres (UK; 366 [86%]), collaborating European institutions in Budapest (Hungary; 20 [5%]) and Warsaw (Poland; 15 [4%]), and samples from the European UKCCSG-SIOP-PNET3 clinical trial (27 [6%]). As is typical for medulloblastoma, we regarded patients younger than 3 years at diagnosis as infants. 108 (26%) of 408 patient samples used were collected in 2010–14, 192 (47%) in 2000–10, 85 (21%) in 1990–2000, and the remaining 23 (6%) were collected before 1990 (18 were from the 1980s, four from the 1970s, and one was from 1968). Year of diagnosis was unavailable for 20 samples.

Tumour samples were provided by the UK CCLG as part of CCLG-approved biological study BS-2007–04; informed, written consent was obtained from parents of all patients because all assessed patients were younger than 16 years. Tumour investigations were done with

approval from Newcastle North Tyneside Research Ethics Committee (study reference 07/Q0905/71); all tumour material was collected in accordance with this approval. We used 276 medulloblastomas (GSE54880) from a published tumour archive,⁸ comprising patients aged from 0–18 years at diagnosis, as an independent validation cohort. 18 post-mortem cerebellar samples were collected from the Newcastle Brain Tissue Resource and used as controls in some analyses; all samples were collected with written, informed consent.

Procedures

We tested medulloblastoma samples with the Illumina HumanMethylation450K DNA methylation array (Illumina, San Diego, CA, USA). The Gene Expression Omnibus accession number for 450K DNA methylation array profiles we used for the determination of human medulloblastoma molecular subgroup status is GSE93646.

To identify methylation-dependent subgroups, we did unsupervised class discovery by NMF-metagene and k-means clustering, testing all combinations of 3–10 metagenes and clusters for reproducibility using bootstrapped resampling methods (250 iterations) as described previously.⁷ This analysis identified metagenes (a single score that reflects the methylation status of several CpG loci) representing the main biological effects present in the genome-wide dataset. We assessed cluster stability using the cophenetic index, a shorthand measure of the robustness of sample clustering as determined by consensus non-negative matrix factorisation (appendix p 3). We visualised clusters with t-SNE.²² We assigned samples classified with less than 80% confidence (by resampling procedures) as non-classifiable (NC; appendix pp 2–3).

We projected metagenes derived from our discovery cohort onto the validation cohort. Additionally, we combined the discovery and validation cohorts to do equivalent consensus clustering.

We assessed established medulloblastoma clinical, pathological, and molecular features as described previously.⁷ Briefly, we defined histopathological variants according to the WHO 2016 guidelines.¹³ We assigned metastatic status (M+) based on Chang's criteria (appendix p 3). Tumours were designated as R+ if their residuum after surgical excision exceeded 1.5 cm². Pathology was centrally reviewed by three experienced neuropathologists for 380 (89%) of 428 samples, and clinical data were collated from contributing centres and reviewed centrally (appendix p 3). We assessed *MYC* and *MYCN* status by fluorescence in situ hybridisation or copy-number estimates from methylation array. We assessed *TP53*, *CTNNB1*, and *TERT* mutation status by Sanger sequencing. We identified subgroup-specific differentially methylated CpG loci or methylated regions (DMRs) using limma or DMRcate^{23,24} (appendix p 3). RNA-seq expression data were generated for discovery cohort samples for which mRNA of sufficient quantity

and quality was available. We identified subgroup-specific differentially expressed genes using DESeq2,²⁵ and these genes were included in ontology enrichment analyses (appendix p 4). We identified *GFI1* mutations from RNA-seq data (appendix p 4).

MB_{SHH} mutation data were obtained from a previous study.²⁶ Although 450K methylation data for MB_{SHH} subgroup assignment were not available for this sample cohort, the tightly defined age cutoff that we defined between the molecularly determined MB_{SHH-Infant} and MB_{SHH-Child} subgroups enabled us to infer subgroups for this sequencing cohort (appendix p 4).²⁶ We tested recurrent MB_{SHH} mutations (*TP53*, *SUFU*, *PTCH1*, *SMO*, and *TERT*) and gene amplifications (*MYCN* and *GLI2*) identified by whole genome sequencing, for association with the age-defined MB_{SHH-Child} or MB_{SHH-Infant} subgroups using Fisher's exact test (appendix p 4).

Statistical analysis

We did survival analyses (overall survival and progression-free survival) on samples from patients aged 3–16 years within our discovery cohort, who received maximal surgical resection and craniospinal irradiation with curative intent. Overall survival was defined as the time from date of surgery to death or date of last follow-up and progression-free survival as the time from date of surgery to first event (progression or relapse) or date of last follow-up. Patients with follow-up time that exceeded 10 years were right-censored at 10 years.

The tightly defined age cutoff between the molecularly determined MB_{SHH-Infant} and MB_{SHH-Child} subgroups enabled us to assess an expanded survival cohort of MB_{SHH-Child} disease (n=55), including additional samples with insufficient DNA for methylation array analysis, classified as MB_{SHH-Child} on the basis of their age (appendix p 4). In this group, we assessed the prognostic potential of currently used clinical and molecular risk markers (M+ disease, R+ disease, LCA pathology, sex, *MYCN* amplification, *TERT* mutation, and *TP53* mutation [appendix pp 4–5]). Patients in the MB_{SHH-Infant} subgroup were typically younger than 3 years at diagnosis and were, therefore, treated on infant protocols. Treatment in this group of patients is heterogeneous, and is focused on omitting or delaying radiotherapy to reduce treatment-associated morbidities as far as possible. As a consequence, we report only overall survival in this group.

We created univariate and cross-validated multivariate Cox models based on subgroups, established risk factors, and cytogenetic changes. Prognostic markers in the multivariate analysis were identified by performing 100 rounds of 10-fold cross-validation, evaluating the performance of markers by measuring area under the curve (AUC) at 5 years for progression-free survival in the left out fold, and calculating the overall mean AUC over all rounds (appendix p 5). We added variables conferring an increase in AUC, as measured by time-dependent receiver operating characteristic curves at 5 years, to the model.

See Online for appendix

Because MB_{GTP3} and MB_{GTP4} shared a metagene (V1), which defined a low-risk outcome and implied shared biology, we considered MB_{GTP3/4} as a single entity, and MB_{GTP3} and MB_{GTP4} separately for creation of survival models. In addition to currently understood clinical and molecular risk markers in these groups (M+ disease, R+ disease, LCA pathology, gender, *MYC/MYCN* amplification, and i17q [isochromosome 17q]), we additionally tested for recurrent cytogenetic changes, MB_{GTP3} membership, and membership of the high-risk methylomic group composed of members from both MB_{GTP3} and MB_{GTP4}, defined by metagene V1 (appendix pp 5–6). We categorised identified independent prognostic markers into risk-stratification schemes (favourable-risk, >90% survival; standard-risk, >75–90% survival; high-risk, 40–75% survival; very high-risk, <40% survival) and survival-dependent ROC analysis of progression-free survival at 5 years, to assess performance²⁷ by comparison with previously reported classification schemes (appendix pp 5–6).^{16,28}

We constructed Kaplan-Meier curves and compared patient groups with log-rank tests. For Kaplan-Meier comparison of two groups, we calculated hazard ratios (HRs) for the 0–5 year survival interval and 95% CIs from the Wald statistic. We tested the proportionality assumption for Cox modelling using scaled Schoenfeld residuals. Missing data were assumed to be missing completely at random and affected samples were removed from multivariate analyses. We implemented array processing, normalisation, quality-control checks, and copy-number estimation, relative to a panel of 18 normal cerebella with the R packages *minfi*²⁹ and *cnmeme* (appendix p 2).

The significance threshold was set at $p < 0.05$ for all statistical tests in this study, unless otherwise stated. Significance of association was assessed using Fisher's exact and chi-squared tests with Yates' continuity correction. We identified subgroup-specific age-differences between the non-MB_{WNT} or non-MB_{SHH} medulloblastoma subgroups using ANOVA (appendix p 4). Statistical or bioinformatics analyses were done using R (version 3.2.3).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all of the data and had the final responsibility to submit for publication.

Results

Clinicopathological and molecular diagnostic characteristics of 428 patients younger than 16 years who had primary childhood medulloblastoma (discovery cohort) are shown in table 1. Consensus analysis identified two equally robust cluster solutions (cophenetic index 0.998 [four metagenes] and 0.997 [six metagenes]; appendix p 10). The first cluster solution (four metagenes, four clusters) recapitulated the established four-subgroup

consensus,¹ whereas the second (six metagenes, seven clusters) revealed further clusters within the established subgroups (figure 1A, appendix pp 10–11).

	Discovery cohort (n=428)	Validation cohort (n=276)	MB _{SHH-Child} survival cohort (n=55)	MB _{GTP3/4} survival cohort (n=175)
Sex				
Male	278 (65%)	174 (63%)	32 (58%)	124 (71%)
Female	150 (35%)	102 (37%)	23 (42%)	51 (29%)
Male:female ratio	1.9:1	1.7:1	1.4:1	2.4:1
Age at diagnosis (years)				
Median (range)	6.34 (0.24–15.97)	7.50 (0.0–18.0)	10.86 (3.5–15.54)	7.33 (3.4–15.97)
<3	101 (24%)	30 (11%)	0	0
≥3	327 (76%)	244 (89%)	55 (100%)	175 (100%)
Pathology variant				
Classic	276 (70%)	NA	23 (44%)	131 (83%)
DN/MBEN	58 (15%)	NA	15 (29%)	7 (5%)
LCA	60 (15%)	NA	14 (27%)	19 (12%)
MB-NOS	34	NA	3	18
Metastatic stage				
M–	285 (73%)	NA	47 (85%)	124 (73%)
M+	104 (27%)	NA	8 (15%)	47 (27%)
Resection				
Sub-total resection (R+)	98 (26%)	NA	9 (16%)	51 (29%)
Gross total resection (R–)	285 (74%)	NA	46 (84%)	123 (71%)
Treatment				
RTX alone	28 (8%)	NA	16 (33%)	16 (9%)
RTX and CTX	314 (92%)	NA	32 (67%)	157 (91%)
Follow-up time (years)	4.91 (0.2–25.7)	NA	6.52 (0.5–16.8)	4.58 (0.4–25.7)
CTNNB1 mutation				
Mutant	24 (7%)	NA	NA	0
Wild-type	297 (93%)	NA	NA	144 (100%)
Chromosome 6				
Loss	30 (8%)	28 (10%)	NA	0
Normal	361 (92%)	248 (90%)	NA	158 (100%)
Chromosome 17				
i17q	111 (28%)	87 (32%)	NA	72 (46%)
No i17q	280 (72%)	189 (68%)	NA	86 (54%)
MYC amplification				
Positive	22 (5%)	12 (4%)	NA	8 (5%)
Negative	404 (95%)	264 (96%)	NA	165 (95%)
MYCN amplification				
Positive	29 (7%)	17 (6%)	NA	11 (6%)
Negative	397 (93%)	259 (94%)	NA	162 (94%)
TP53 mutation				
Positive	18 (7%)	NA	13 (27%)	1 (1%)
Negative	238 (93%)	NA	35 (73%)	89 (99%)
TERT mutation				
Positive	16 (4%)	NA	18 (35%)	1 (1%)
Negative	357 (96%)	NA	34 (65%)	150 (99%)

(Table 1 continues on next page)

	Discovery cohort (n=428)	Validation cohort (n=276)	MB _{SHH-Child} survival cohort (n=55)	MB _{Grp3/4} survival cohort (n=175)
(Continued from previous page)				
450K 4 subgroup assignment				
MB _{WNT}	33 (8%)	33 (12%)	NA	0
MB _{SHH}	109 (26%)	60 (22%)	24 (100%)	1 (1%)
MB _{Grp3}	130 (31%)	72 (26%)	NA	63 (36%)
MB _{Grp4}	153 (36%)	111 (40%)	NA	109 (63%)
Non-classifiable	3	NA	NA	2
450K 7 subgroup assignment				
MB _{WNT}	33 (8%)	33 (12%)	NA	NA
MB _{SHH-Child}	38 (9%)	32 (12%)	24 (100%)	NA
MB _{SHH-Infant}	65 (16%)	28 (10%)	NA	NA
MB _{Grp3-HighRisk}	65 (16%)	51 (18%)	NA	44 (25%)
MB _{Grp3-LowRisk}	50 (12%)	20 (7%)	NA	13 (7%)
MB _{Grp4-HighRisk}	85 (21%)	54 (20%)	NA	63 (36%)
MB _{Grp4-LowRisk}	73 (18%)	58 (21%)	NA	55 (31%)
Non-classifiable	19	0	NA	NA

Data are n (%) or median (range), unless otherwise specified. MB=medulloblastoma. SHH=sonic hedgehog. Grp3=group 3. Grp4=group 4. Grp3/4=combined groups 3 and 4. NA=data not available. DN=desmoplastic or nodular. MBEN=medulloblastoma with extensive nodularity. LCA=large-cell anaplastic. MB-NOS=medulloblastoma not otherwise specified. M=non-metastatic disease. M+=metastatic disease. RTX=radiotherapy. CTX=chemotherapy. WNT=wnt/wingless.

Table 1: Demographics and clinicopathological characteristics of all cohorts

MB_{WNT} tumours formed a single subgroup (n=33) characterised by *CTNNB1* mutations, loss of chromosome 6, and an expected favourable prognosis (5-year overall survival: 93% [95% CI 82–100]; figure 1B). Our newly detected metagenes split each remaining consensus subgroup (MB_{SHH}, MB_{Grp3}, and MB_{Grp4}) in two. MB_{SHH} was split into age-dependent subgroups corresponding to infant (<4·3 years; MB_{SHH-Infant}; n=65) and childhood patients (≥4·3 years; MB_{SHH-Child}; n=38) by the respective absence or presence of metagene V4. Both have intermediate prognoses (5-year overall survival MB_{SHH-Child}: 58% [95% CI 41–82]; MB_{SHH-Infant}: 62% [50–77]; figure 1B). MB_{Grp3} and MB_{Grp4} are each split into high-risk (MB_{Grp3-HR} [n=65] and MB_{Grp4-HR} [n=85]) and low-risk (MB_{Grp3-LR} [n=50] and MB_{Grp4-LR} [n=73]) subgroups by common metagene V1 (figure 1A). 5-year overall survival was 37% [95% CI 25–53] in the MB_{Grp3-HR} subgroup, 69% [55–87] in the MB_{Grp3-LR} subgroup, 69% [58–83] in the MB_{Grp4-HR} subgroup, and 80% [70–92] in the MB_{Grp4-LR} subgroup (figure 1B). The subdivision of MB_{Grp3} and MB_{Grp4} distinguishes patients with a superior stratification (5-year overall survival AUC 0·649 [MB_{Grp3/4} combined with low-risk or high-risk subdivision]) compared with the current consensus MB_{Grp3} and MB_{Grp4} subgroups (AUC 0·610). Moreover, in the patients aged 3–16 years at diagnosis and receiving craniospinal irradiation, the high-risk or low-risk subdivision of MB_{Grp3/4} stratifies this group into standard (MB_{Grp3-LR} 81% [95% CI 60–100%]; MB_{Grp4-LR} 81% [71–93%]) and high-risk (MB_{Grp3-HR} 35% [23–55%]; MB_{Grp4-HR} 47% [34–66%]) 5-year progression-free survival outcomes, by contrast with the current

consensus MB_{Grp3/4} designations, which show intermediate outcomes (figure 1C, 1D).

Clinicopathological and biological features were non-randomly distributed in all seven subgroups (figure 1A, appendix pp 12–15). Patients in the MB_{SHH-Infant} subgroup had significantly enriched desmoplastic or nodular pathology compared with all other subgroups (p<0·0001), and *TP53* mutation (p<0·0001) and *MYCN* amplifications (p<0·0001) were significantly more frequent in MB_{SHH-Child} than in all other subgroups. Patients in the MB_{Grp3-HR} subgroup significantly more frequently had LCA pathology (p<0·0001) and *MYC* amplification (p<0·0001), than all other subgroups. Although patients in the MB_{Grp3-HR} and MB_{Grp4-HR} subgroups had similar 10-year overall survival (22% [95% CI 10–46] vs 36% [22–59]; figure 1B), patients in the MB_{Grp4-HR} subgroup died later of their disease (ten [36%] of 28 deaths in the MB_{Grp4-HR} subgroup occurred more than 5 years after diagnosis) than did those in the MB_{Grp3-HR} subgroup (33 [92%] of 36 deaths occurred within 5 years of diagnosis; appendix p 26).

Validation by projection of six metagenes onto an independent cohort⁸ of 276 patients (table 1) confirmed their existence (appendix pp 10–11). Moreover, reapplying consensus clustering to the combined cohort of 704 patients confirmed a seven subgroup model as optimal, giving 100% concordance to the classifications derived separately from our discovery cohort (appendix pp 10–11).

Age distributions differed between the two MB_{SHH} subgroups; age distributions are log-normally distributed and intersect at 4·3 years (figure 2A). The two peak incidences of age at diagnosis in infants and in older children for MB_{SHH} disease,²⁶ when observed as a whole, are resolved by their classification into distinct MB_{SHH-Infant} and MB_{SHH-Child} subgroups (appendix pp 12–13). Each MB_{SHH} subgroup possesses characteristic molecular or clinicopathological features (appendix pp 12–13). LCA pathology (p=0·00050), *MYCN* amplification (p<0·0001), and mutations of *TP53* (p<0·0001) and *TERT* (p=0·0015) were all significantly enriched in the MB_{SHH-Child} subgroup compared with the MB_{SHH-Infant} subgroup; whereas gender, M+ disease status, and R+ disease status were not significantly different between groups (figure 2B; appendix pp 12–13). *TERT* promoter mutation and *MYCN* amplification or LCA pathology were mutually exclusive (figure 2B; appendix pp 12–13). Mutational data from an independent MB_{SHH} cohort²⁶ showed that *SUFU* mutation was significantly associated with MB_{SHH-Infant}, whereas *PTCH1* mutations were observed in both MB_{SHH} subgroups (figure 2C). *GLI2* amplification, *MYCN* amplification, and *TP53* mutations (both somatic and germline) were significantly associated with the MB_{SHH-Child} subgroup (figure 2C).

Compared with normal cerebella and patients in the MB_{SHH-Infant} subgroup, patients in the MB_{SHH-Child} subgroup had subgroup-specific DNA methylation changes

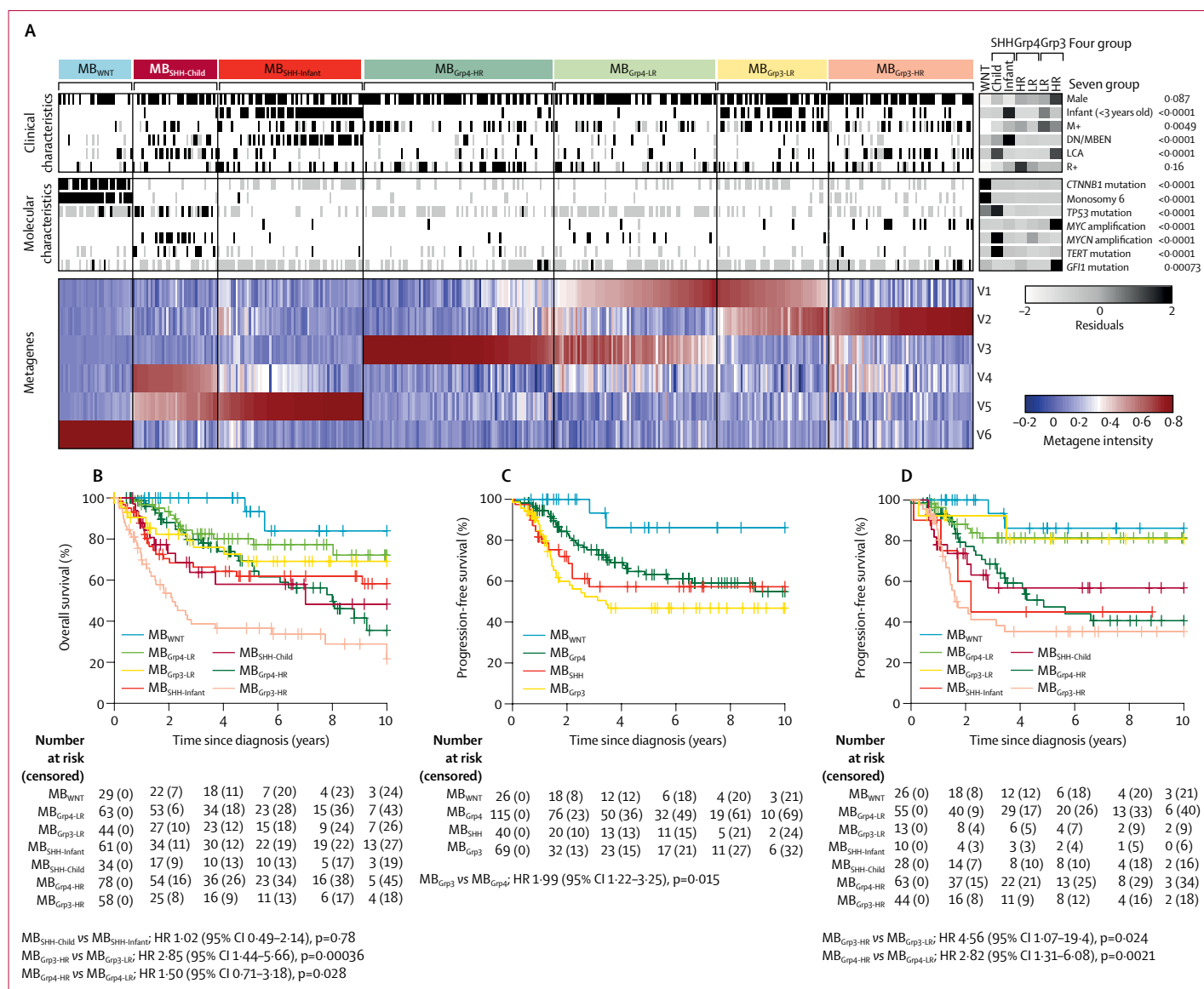


Figure 1: Novel clinically significant subgroups within the established medulloblastoma subgroups

(A) Non-negative matrix factorisation consensus clustering of methylome data from 428 primary medulloblastomas. Each column represents one patient. Missing data are shown in grey. Residuals from χ^2 tests indicate where subgroup-enrichment has occurred (darker shades of grey indicate stronger relationships), p values are from χ^2 tests of enrichment; scale bar for residuals (-2 to 2) is shown. Methylation-derived metagene levels (V1-V6), which define subgroup membership, are also shown (red indicates high metagene levels, blue indicates low levels). (B) Overall survival of patients in the seven identified subgroups. All discovery cohort patients with available overall survival information are shown (n=367). (C) Progression-free survival of patients in the consensus four subgroups of medulloblastoma in discovery cohort patients receiving craniospinal irradiation and aged 3-16 years at diagnosis (n=250). (D) Progression-free survival of patients in the seven identified subgroups of medulloblastoma in patients receiving craniospinal irradiation and aged 3-16 years at diagnosis (n=239). Discrepancy in the numbers of patients in (C) and (D) is due to consensus clustering; certain samples could not be confidently classified for the seven subgroup model or the four subgroup model, and were omitted from the figures. DN/MBEN=desmoplastic or nodular medulloblastoma with extensive nodularity. HR=hazard ratio. LCA=large-cell anaplastic. M+=metastatic disease. R+=residual disease.

(predominantly hypermethylation), at both individual CpG loci and at the gene level (figure 2B; appendix pp 12-13), frequently involving developmental genes (79 [14%] of 584 genes with gene ontology term embryonic morphogenesis had aberrant hypermethylation). DNA methylation changes were validated in an independent cohort⁸ (appendix pp 12-13). When discovery cohort MB_{SHH} RNA-seq expression data were available

(190 [44%] of 428 samples), significant differential expression was observed between the subgroups (1593 genes, fold change >1.5; adjusted p<0.01; appendix pp 12-13). Although there were few recurrent cytogenetic changes, many tumours in patients in the MB_{SHH-Child} subgroup (18 [51%] of 35 tumours) had loss of chromosome 9q, often associated with gain of 9p (appendix pp 12-13).

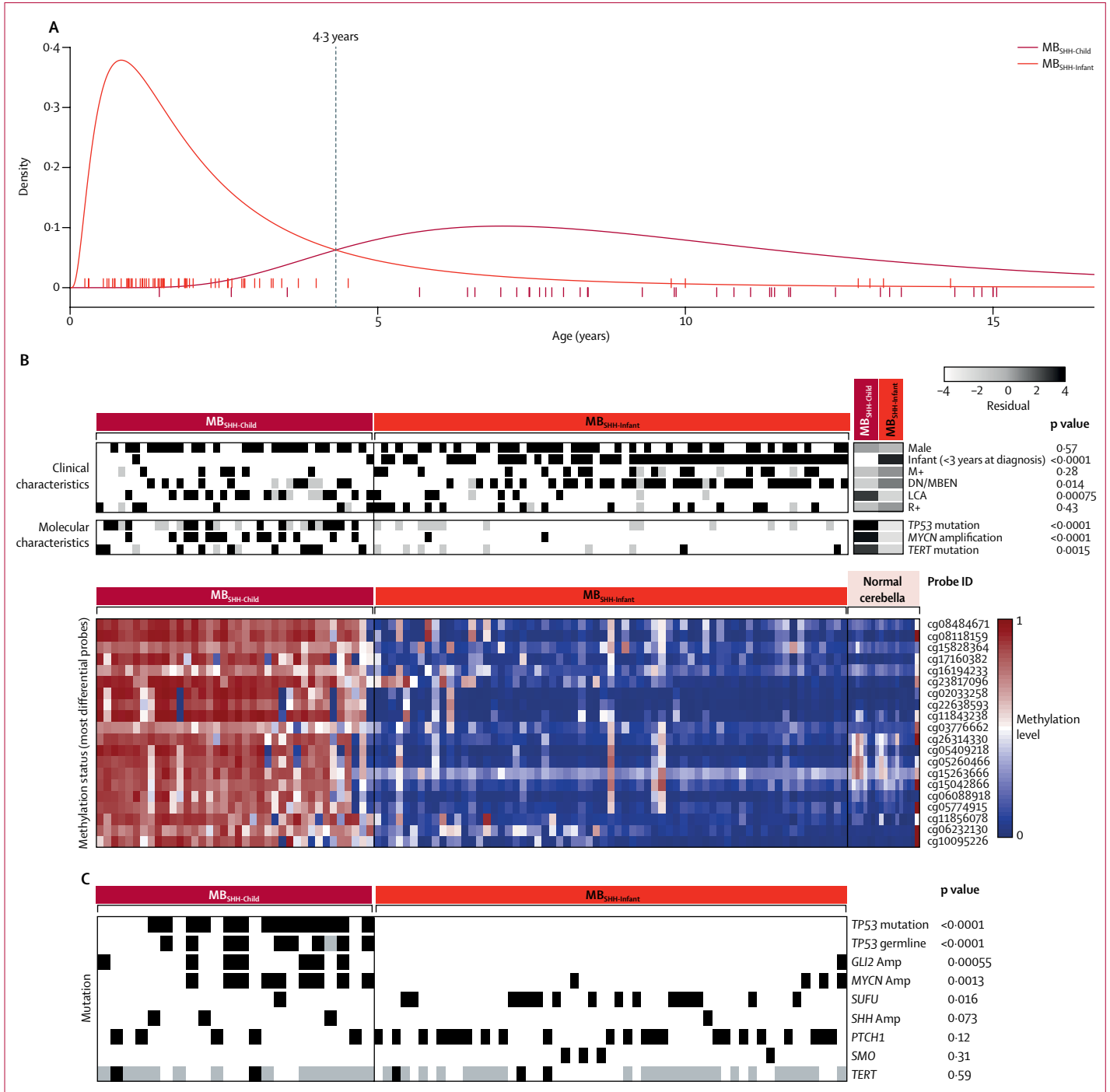


Figure 2: MB_{SHH} disease comprises two age-dependent molecular subgroups

(A) Log-normal age distributions of MB_{SHH-Infant} (red) and MB_{SHH-Child} disease (dark red). Patient ages at diagnosis are shown as ticks along the x-axis and are coloured by subgroup. (B) Clinicopathological and molecular disease features of MB_{SHH-Infant} and MB_{SHH-Child} subgroups. Residuals from χ^2 tests indicate where subgroup-enrichment has occurred (darker shades of grey indicate stronger relationships); scale bar for residuals (-4 to 4) is shown. p values from χ^2 tests are shown. Differentially methylated probes: Illumina probe identifiers for the top 20 most differentially methylated probes, alongside methylation status of 18 normal cerebella (pink). Each column represents one patient. (C) SHH genome-sequencing data²⁶ was classified into methylation subgroups on the basis of age. Each column represents one patient. Amp=amplification. DN/MBEN=desmoplastic or nodular medulloblastoma with extensive nodularity. LCA=large-cell anaplastic. M+=metastatic disease. R+=residual disease.

The age distributions of patients in the four MB_{Grp3} and MB_{Grp4} subgroups differed significantly ($p < 0.0001$). Patients in the MB_{Grp3-LR} and MB_{Grp3-HR} subgroups were younger at diagnosis than those in the MB_{Grp4-LR} and MB_{Grp4-HR} subgroups (appendix pp 14–15). Infants in the MB_{Grp3-HR} subgroup frequently had amplified MYC (seven

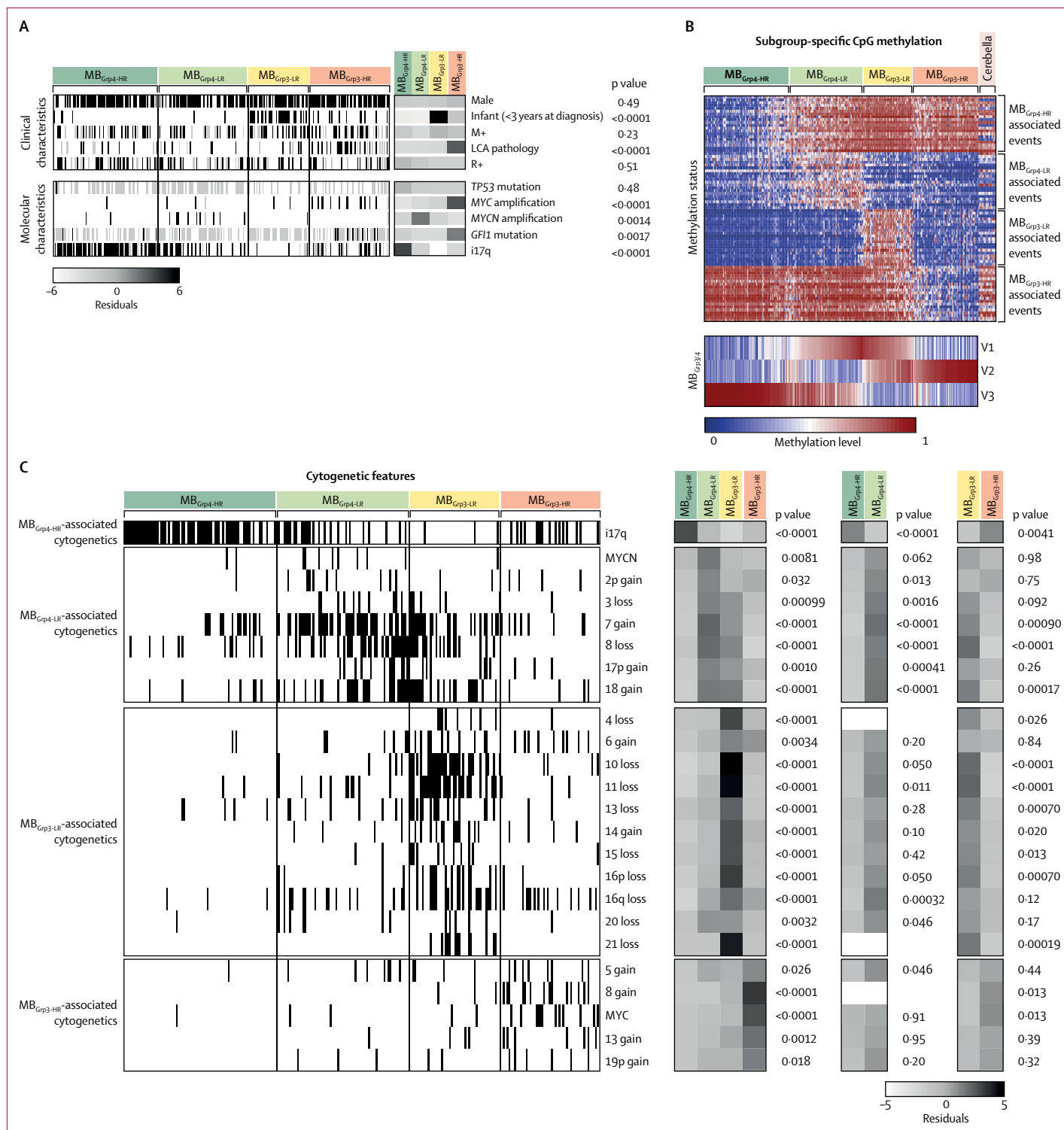


Figure 3: Characterisation of MB_{Cp3} and MB_{Cp4} subgroups

(A) Clinicopathological and molecular disease features. Residuals from χ^2 tests indicate where subgroup-enrichment has occurred (darker shades of grey indicate stronger relationships); scale bar for residuals (-6 to 6) is shown. p values from χ^2 tests are shown. (B) Heat map shows the top 20 differentially methylated probes for these subgroups. Methylation data of 18 normal cerebella are shown alongside and magnitude of MB_{Cp3} and MB_{Cp4} metagenes is shown below. (C) Identification of MB_{Cp3} and MB_{Cp4} medulloblastoma cytogenetic determinants. Markers with $p < 0.05$ and present in at least 10% of one subgroup are ordered by their subgroup association and then by chromosomal order. Residuals from χ^2 tests indicate where subgroup enrichment has occurred (darker shades of grey indicate stronger relationships), across all subgroups and within MB_{Cp3} and MB_{Cp4} individually. p values from χ^2 tests are shown. i17q=iso chromosome 17q. LCA=large-cell anaplastic. M+=metastatic disease. R+=residual disease.

[64%] of 11 infants). MB_{Grp3-HR} tumours were strongly associated with LCA pathology (20 [35%] of 57) and *GFI1* mutations (nine [29%] of 31; figure 3A, appendix pp 14–15). i17q is the sole significantly enriched feature of MB_{Grp4-HR} (60 [76%] of 79 [figures 3A, 3C]. Clinicopathological and molecular disease features of the MB_{Grp3} and MB_{Grp4} subgroups are shown in figure 3A.

Several hundred differentially methylated CpG probes or regions defined the four subgroups. MB_{Grp3-HR} was characterised by the greatest number of significantly differentially methylated CpGs compared with other subgroups, commonly hypomethylated CpG loci (figure 3B; appendix pp 14–15). Notably, the low-risk subgroups were defined primarily by hypermethylation with respect to normal cerebellum, whereas the high-risk

subgroups were defined by hypomethylation (figure 3B; appendix pp 14–15). Cytogenetic changes distinguished each subgroup as unique from the others (figure 3C). These distinguishing cytogenetic features were validated in an independent cohort (appendix pp 14–15).

We did survival analyses in an MB_{SHH-Child} cohort that included 31 additional SHH cases unsuitable for 450k array analysis and classified as MB_{SHH-Child} on the basis of age (appendix pp 4–5). In this cohort, one out of three assessable *TP53* mutations were germline (appendix pp 16–17). *TP53* mutation was significantly associated with *MYCN* amplification (p=0.022) and LCA pathology (p=0.0033), *MYCN* amplification was associated with LCA pathology (p<0.0001), and LCA pathology and *MYCN* amplification were never observed with *TERT* mutations (p=0.00079 for LCA and p=0.0090 for *MYCN* amplification). There was no significant association between metastatic (M+) disease and *TP53* mutation (p=1), *MYCN* amplification (p=0.15), or LCA pathology (p=0.67), or an association between subtotally resected (R+) disease and *TP53* mutation (p=1), *MYCN* amplification (p=1) or LCA pathology (p=0.41). Univariate survival analysis of clinicobiological features (including risk features established in disease-wide studies¹⁶) in this cohort showed significantly shorter progression-free survival associated with *MYCN* amplification *TP53* mutation, LCA pathology, M+ disease, and R+ disease, but no associations with *TERT* mutation status or sex (table 2; appendix pp 18–19). Multivariate Cox modelling, showed that *MYCN* amplification, *TP53* mutation, and M+ disease are independent risk factors for progression-free survival (table 2). Only the 42 samples with complete clinical information for the considered variables were included. The disease-wide risk-stratification scheme currently in use for the HIT-SIOP-PNET5-MB clinical trial,¹⁶ which deems *MYCN* amplification, LCA pathology, M+ disease, and R+ disease as high-risk factors, outperformed the MB_{SHH-Child} subgroup stratification in AUC analysis (appendix pp 16–17). We used this HIT-SIOP-PNET5-MB stratification scheme as the basis of a combined risk-stratification model for MB_{SHH-Child} (appendix pp 16–17), classifying patients with any one of these risk factors as very high risk. 50 patients had sufficient clinical data for classification using the scheme. This model discriminates favourable (24 [48%] of 50 patients, 5-year progression-free survival: 96% [95% CI 88–100]) and very high-risk (26 (52%), 5-year progression-free survival: 29% [14–58]) groups of patients within the MB_{SHH-Child} subgroup (p<0.0001; appendix pp 16–17).

Combining all craniospinally irradiated patients in the MB_{Grp3/4} subgroup aged 3–16 years who had outcome data (n=175), allocation to the MB_{Grp3-HR} and MB_{Grp4-HR} subgroups was a significant high-risk factor for shorter progression-free survival in univariate analysis (table 3). Additionally, in multivariate analysis, *MYC* amplification was identified as an independently prognostic high-risk factor, and chromosome 13 loss was associated with an improved outcome (table 3).

	n	Univariate (n=55)		Cross-validated multivariate (n=42)	
		HR (95% CI)	p value	HR (95% CI)	p value
<i>MYCN</i> amplification vs no amplification	52	4.47 (1.65–12.1)	0.0032	2.83 (0.87–9.22)	0.084
M+ vs M– disease	55	5.69 (2.01–16.0)	0.0011	4.59 (1.28–16.4)	0.019
<i>TP53</i> mutation vs no mutation	48	3.47 (1.29–9.30)	0.014	3.44 (1.15–10.2)	0.027
LCA pathology vs non-LCA pathology	52	2.88 (1.15–7.24)	0.025
<i>TERT</i> wild-type vs <i>TERT</i> mutation	52	2.21 (0.78–6.25)	0.13
R+ vs R– disease	55	3.45 (1.30–9.19)	0.013
Male vs female	55	1.13 (0.45–2.82)	0.79

p values are from Cox proportional hazards analyses. The prognostic significance of covariates selected in cross-validated multivariate models are also shown. HR=hazard ratio. M+=metastatic disease. M–=non-metastatic disease. LCA=large-cell anaplastic. R+=residual disease (subtotal surgical resection). R–=no residual disease (gross total resection).

Table 2: Identification of prognostic survival markers in MB_{SHH-Child} cohort

	n	Univariate (n=175)		Cross-validated multivariate (n=133)	
		HR (95% CI)	p value	HR (95% CI)	p value
High-risk methylation group vs low-risk methylation group	175	3.73 (1.94–7.18)	<0.0001	3.21 (1.59–6.51)	0.0012
<i>MYC</i> amplification vs no amplification	173	2.94 (1.06–8.13)	0.038	18.4 (5.01–67.7)	<0.0001
Loss of chromosome 13 vs no loss	158	0.10 (0.01–0.74)	0.024	0.06 (0.01–0.49)	0.0090
MB _{Grp3} vs MB _{Grp4}	175	2.04 (1.23–3.40)	0.006
M+ vs M– disease	171	1.77 (1.03–3.05)	0.039
i17q vs no i17q	158	1.71 (0.99–2.95)	0.056
Male vs female	175	1.56 (0.86–2.84)	0.144
<i>MYCN</i> amplification vs no amplification	173	0.72 (0.23–2.29)	0.576
LCA pathology vs non-LCA pathology	157	1.08 (0.49–2.39)	0.848
R+ vs R– disease	171	1.22 (0.72–2.09)	0.464

Identification of prognostic survival markers in combined childhood non-MB_{SHH} and non-MB_{WNT} survival cohort (aged 3.0–16.0 years, receiving craniospinal irradiation, with survival information). p values from Cox proportional hazards analyses are shown. The characteristics of covariates selected in cross-validated multivariate models are also shown. The high-risk methylation group comprised samples from both MB_{Grp3} and MB_{Grp4}, defined by the shared MB_{Grp3/4} metagene V1. HR=hazard ratio. MB=medulloblastoma. Grp3=group 3. Grp4=group 4. M+=metastatic disease. M–=non-metastatic disease. i17q=isochromosome 17q. LCA=large-cell anaplastic. R+=residual disease (subtotal surgical resection). R–=no residual disease (gross total resection).

Table 3: Identification of prognostic survival markers in MB_{Grp3} and MB_{Grp4} cohorts

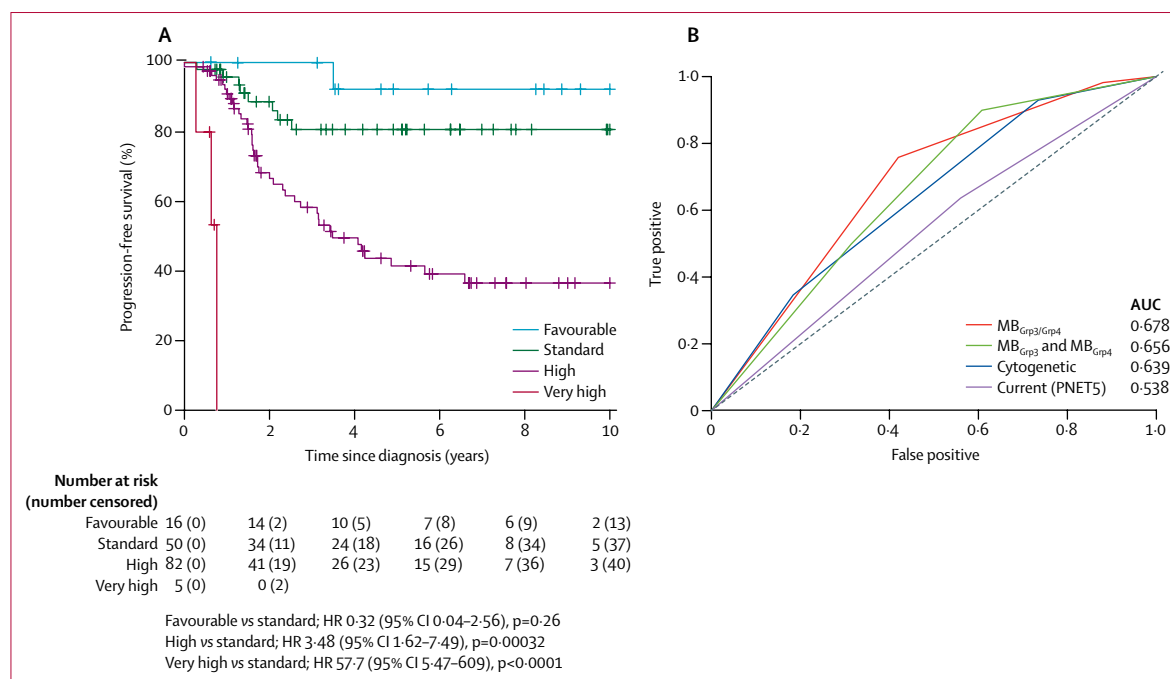


Figure 4: Novel risk stratification scheme for MB_{Grp3} and MB_{Grp4} medulloblastoma

(A) Progression-free survival plots for identified risk subgroups (n=156) defined in table 3 and the appendix (p 20). (B) Time-dependent ROC curves at 5 years are shown for this novel risk stratification alongside a published cytogenetic stratification scheme²⁸ (MB_{Grp4} with chromosome 11 loss or chromosome 17 gain, low risk; MB_{Grp4} with M– disease, standard risk; MB_{Grp4} with M+ disease, high risk; MB_{Grp3} with MYC amplification, i17q, or M+ disease, high risk; MB_{Grp3} without MYC amplification, i17q, or M+ disease, standard risk), and the PNET5 risk stratification (patients positive for one or more of LCA pathology, M+ disease, R+ disease, MYC(N) amplification are high risk; patients absent for all high-risk features, standard risk), as well as the stratification derived from considering MB_{Grp3} and MB_{Grp4} as separate entities (appendix p 22). AUC=area under curve. LCA=large-cell anaplastic. M+=metastatic disease. M–=non-metastatic disease. ROC=receiver operating characteristic.

A stratification model was developed that divided MB_{Grp3/4} into different risk groups for 5-year progression-free survival: favourable risk (chromosome 13 loss and no MYC amplification; 16 [10%] of 153 patients; 92% [95% CI 79–100]); standard risk (MB_{Grp4-LR} or MB_{Grp3-LR} with no MYC amplification; 50 [33%] patients; 81% [70–94]); high risk (MB_{Grp4-HR} or MB_{Grp3-HR} with no MYC amplification; 82 [54%] patients; 42% [31–56]); and very high risk (MB_{Grp3} with MYC amplification; five [3%] patients; 0%; figure 4A; appendix pp 20–21). 156 patients had information for chromosome 13 loss and MYC amplification, of which three were classed as unassignable because they were MB_{Grp4} with MYC amplification (appendix pp 20–21). This stratification scheme outperformed current risk-stratification models (figure 4B).

For comparison, we developed equivalent separate survival stratification schemes for MB_{Grp3} and MB_{Grp4} (appendix pp 22–23). Risk factors identified were broadly consistent with the factors identified in the combined scheme, although the combined scheme was a better predictor of progression-free survival than when MB_{Grp3} and MB_{Grp4} were considered separately (figure 4B). Taking MB_{Grp4} patients in isolation, in univariate analysis, a designation of MB_{Grp4-HR}, chromosome 7q status, M+ disease, and male sex were associated with poor progression-free survival, whereas MYCN amplification, R+ disease, and LCA pathology were not (appendix

pp 22–23). Chromosome 7q gain and M+ disease were retained as independent prognostic factors in multivariate analysis (appendix pp 22–23). A 5-year progression-free survival model incorporating chromosome 7q gain and M+ disease defined standard-risk (35 [32%] of 110 patients; 87% [95% CI 76–100]) and high-risk groups (75 [68]; 49% [37–66]), and outperformed other published models by AUC analysis (appendix pp 22–23).

Taking patients with MB_{Grp3} in isolation, MYC amplification was the only risk factor significantly associated with progression-free survival in multivariate analysis, and outcomes were poor for these very high-risk patients (appendix pp 22–23). Patients in the MB_{Grp3} with non-MYC amplified tumours were at high risk, with progression-free survival similar to that for the MB_{Grp4-HR} subgroup (51 [91%] of 56 patients; 46% [95% CI 33–64] for MB_{Grp3} with non-MYC amplified tumours vs 41% [28–60] for MB_{Grp4-HR}). MB_{Grp3-HR} shows a worse outcome than MB_{Grp3-LR} ($p=0.040$; appendix p 22). LCA pathology (11 [21%] of 53 patients), M+ disease (17 [29%] of 58 patients), and R+ disease (13 [22%] of 59 patients),¹ were frequent in patients in MB_{Grp3} but none were associated with prognosis, and no stratification scheme based on MB_{Grp3} alone markedly improved outcome prediction compared with standard stratification schemes (appendix pp 22–23).

	WNT	MB _{SHH-Child}	MB _{SHH-Infant}	MB _{Grp4-HR}	MB _{Grp4-LR}	MB _{Grp3-LR}	MB _{Grp3-HR}	
Demographics	Infant disease % (<3 years)	0	5	78	5	3	54	17
	Male %	48	63	55	67	66	68	77
	n	33	38	65	85	73	50	65
Clinical features	Histology (%) CLAS:DN:LCA	86:3:10	32:26:41	35:55:10	86:5:9	85:6:9	90:2:8	61:4:35
	Metastasis (%)	3	16	28	30	23	41	33
	Sub-total resection (%)	10	17	26	35	28	24	25
	10 year overall survival (95% CI)	72% (66–100)	48% (29–80)	58% (46–75)	36% (22–59)	72% (59–88)	69% (55–87)	22% (10–46)
Molecular features	Mutation	CTNNB1, TP53	TP53, TP53 GL, TERT, SUFU, PTCH1	SUFU, PTCH1				GFI1
	Cytogenetics		 MYCN, GLI2 amplification			 MYCN amplification		 MYC amplification
	Gene expression*		↑RUNX3, HCAR1, HCAR2, FOXG1	↑TRABD2A, TTC9, SLFN11, CHRM2	↑ESYT2, WDR60, DAPK2, PRDM6	↑BMP5, SPTLC3, COL9A3, ZIC5	↑FGD6, BRMS1L, FAM122B, REV3L	↑PVT1, TRAP1, NMRAL1, CNTLN Ribosome biogenesis genes
	DNA methylation	Global: ↓ vs CB Probe level*: PI3K-Akt, Ras signalling pathways Gene level*: ↓ vs MB _{SHH-Infant} , CB DLX6-AS1, ACTA1, GCM2, FEZF2	Global: ↓ vs CB ↑ vs MB _{SHH-Infant} Probe level*: Ras signalling pathway Gene level*: ↑ vs MB _{SHH-Infant} , CB DLX6-AS1, ACTA1, GCM2, FEZF2	Global: ↓ vs CB ↓ vs MB _{SHH-Child} Probe level*: Hippo signalling pathway Gene level*: ↓ vs MB _{SHH-Child} , CB DLX6-AS1, ACTA1, GCM2, FEZF2	Global: ↓ vs CB ↓ vs MB _{Grp4-LR} Probe level*: PI3K-Akt signalling pathway Gene level*: ↓ vs MB _{Grp4-LR} , CB HLA-DRB5, NXK2-5, ABLIM1, HOXC6	Global: ↓ vs CB ↑ vs MB _{Grp4-HR} Probe level*: PI3K-Akt signalling pathway Gene level*: ↑ vs MB _{Grp4-HR} , CB HLA-DRB5, NXK2-5, ABLIM1, HOXC6	Global: ↓ vs CB ↑ vs MB _{Grp3-HR} Probe level*: PI3K-Akt signalling pathway Gene level*: ↑ vs MB _{Grp3-HR} , CB PRKCZ, MCF2L, MIR662	Global: ↓ vs CB ↓ vs MB _{Grp3-LR} Probe level*: PI3K-Akt signalling pathway Gene level*: ↑ vs MB _{Grp3-LR} , CB GALNT9, MIR662

Figure 5: Summary of the seven primary childhood medulloblastoma subgroups

Demographic, clinicopathological, and molecular features are summarised. *Comparisons of cytogenetic, gene expression, and DNA methylation changes are made with respect to their counterpart subgroup, except for MB_{WNT} cases, which were compared with normal cerebella if data were available. For probe-level comparisons, Kyoto Encyclopedia of Genes and Genomes pathway enrichment of demethylated loci was investigated, after correcting for multiple probes mapping to the same gene (data summarised in appendix pp 27–31). CB=normal cerebella. CLAS=classic histological subtype. DN=desmoplastic nodular. LCA=large-cell anaplastic.

The clinicopathological and molecular features of the new seven clinically significant subgroups are summarised in figure 5. The combination of subgroup-specific survival models creates an overarching risk stratification for all childhood medulloblastoma (figure 6A). Patients are stratified into four clinical risk groups for 5-year progression-free survival: favourable risk (comprising MB_{WNT}, MB_{SHH-Child} with no high-risk features, and non-MYC amplified MB_{Grp3/Grp4} with chromosome 13 loss; 54 [25%] of 215 patients; 91% [95% CI 82–100]); standard risk (comprising non-MYC amplified MB_{Grp3-LR/Grp4-LR} subgroups; 50 [23%] patients; 81% [70–94]); high-risk (comprising non-MYC amplified MB_{Grp3-HR/Grp4-HR} subgroups; 82 [38%] patients; 42% [31–56]); and very high-risk (comprising MB_{SHH-Child} with high-risk features and MYC-amplified MB_{Grp3}; 29 [13%] patients; 28% [14–56]; figure 6B). 215 patients aged 3–16 years at diagnosis had data available for these factors. The AUC from our proposed stratification of childhood medulloblastoma outperforms current and proposed cytogenetic risk stratifications (figure 6C).²⁸ We note that

M+ disease status is a strong risk factor for poor progression-free survival in MB_{Grp4}. Incorporation of M+ disease status into MB_{Grp4-LR} and non-MYC amplified MB_{Grp3-LR} survival modelling does not affect model performance, but potentially allows redistribution of standard-risk patients to create larger favourable (90 [41%] of 218 patients) and high-risk groups (99 [45%] of 218 patients; figure 6A, C; appendix pp 24–25), which could be considered as an alternative stratification scheme. The proposed refinement to the stratification enables additional cases classified as MB_{Grp3-LR} and MB_{Grp4-LR} that do not have copy number information (other than MYC amplification status) and are non-metastatic to be classified as favourable.

Discussion

The discovery and validation of seven robust and reproducible primary molecular subgroups of childhood medulloblastoma in this retrospective cohort study represents, to our knowledge, the first clinically significant elaboration of the four-subgroup consensus

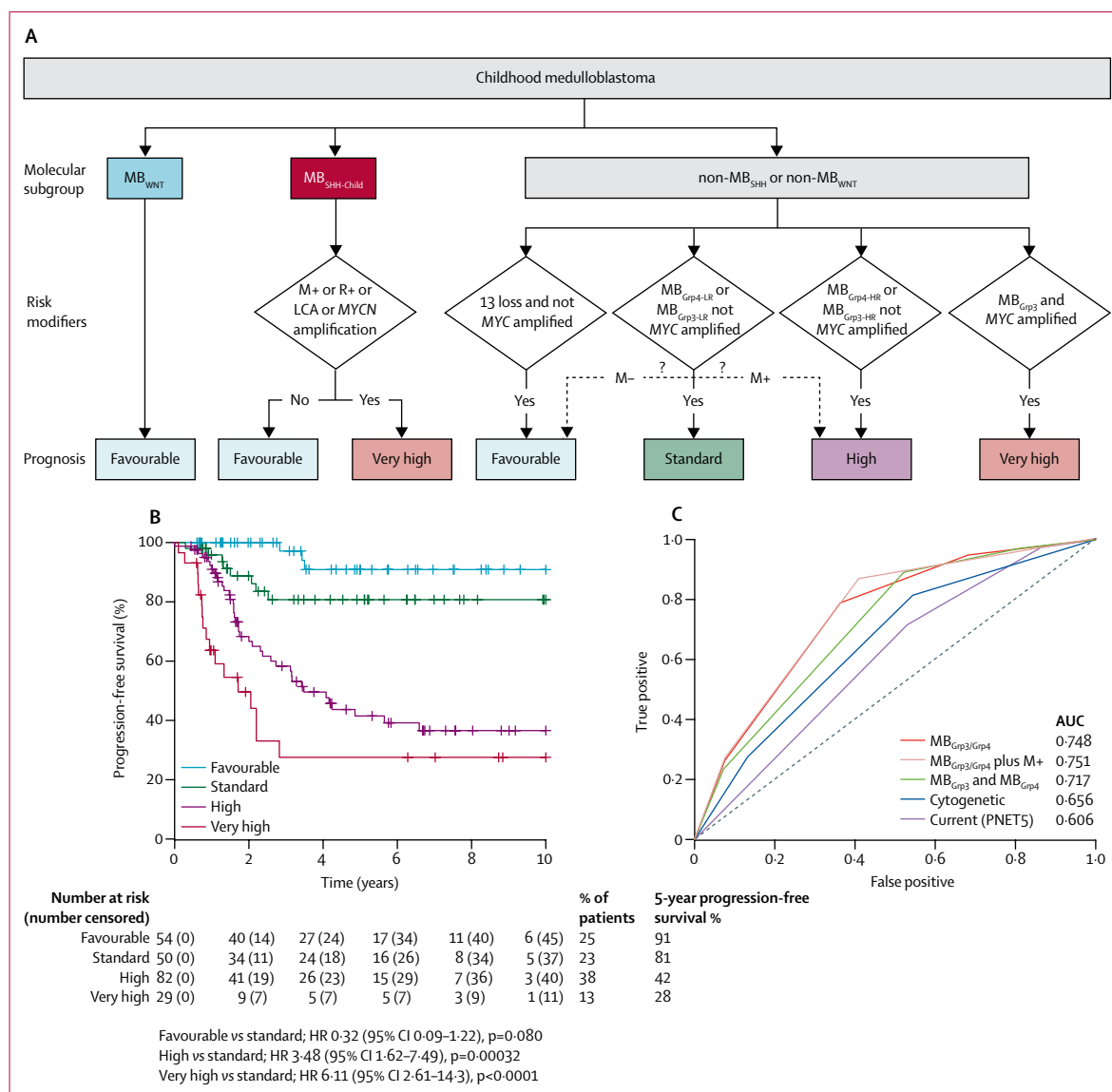


Figure 6: Summary of survival modelling of novel medulloblastoma subgroups

(A) Summary of a novel risk-stratification scheme for childhood medulloblastoma in a cohort of patients aged 3–16 years receiving craniospinal irradiation (n=215). The potential to further stratify MB_{Grp4-LR} patients into favourable and high-risk groups by their metastatic stage is shown (dashed arrows). (B) Kaplan-Meier plot of childhood medulloblastoma risk stratification. (C) Performance of novel stratification scheme in comparison with time-dependent ROC curves of existing schemes of progression-free survival at 5 years. MB_{Grp3/4}; MB_{Grp3} and MB_{Grp4} considered as a single entity; MB_{Grp3/4} plus M+: MB_{Grp3} and MB_{Grp4} considered as a single entity with MB_{Grp4-LR} and non-MYC amplified MB_{Grp3-LR} further stratified by M+ disease status; MB_{Grp3} and MB_{Grp4}; MB_{Grp3} and MB_{Grp4} stratified separately; cytogenetic: cytogenetically defined scheme;²⁶ PNET5: scheme employed by HIT-SIOP-PNET5-MB clinical trial. LCA=large-cell anaplastic. M+=metastatic disease. M-=non-metastatic disease. R+=residual disease.

established in 2012.¹ While our work supports the stability of the four established groups, it also reveals significant substructures within each group with distinct clinicopathological and molecular features. Importantly, these primary subgroups emerge from unsupervised analysis, and are supported by distinguishing DNA methylation, gene expression, and copy-number profiles, consistent in discovery and validation cohorts. Notably, these subgroups were not identifiable in a previously published dataset, which

included fewer samples and, specifically, fewer infant patients.⁸ Our seven subgroups reveal a biological overlap between MB_{Grp3} and MB_{Grp4}. They share a biological signature, defined by a common metagene, indicating a clinicobiological overlap, which might suggest a common origin.

These primary subgroups may be further subdivided by the presence or absence of secondary molecular characteristics, many of which, in turn, have subgroup-specific clinical and prognostic significance (eg, *MYC*

amplification in MB_{Grp3} or *TP53* mutation, *MYCN* amplification, LCA pathology, M+ disease, and R+ disease in MB_{SHH-Child}). Some of these secondary features have been described and assigned clinical significance in previous studies; in this Article, their association with specific novel subgroups (eg, chromosome 11 loss and chromosome 17 gain in MB_{Grp4-LR}²⁸) has revealed the underlying biological basis of these subgroup-specific biomarkers. Moreover, re-evaluation of currently used high-risk factors derived from cohort-wide studies that did not consider subgroup shows that their importance is either low (eg, LCA pathology, M+ disease, or R+ disease in MB_{Grp3}; *MYCN* in MB_{Grp4}) or high (*MYCN* amplification, LCA pathology, *TP53* mutation, and M+ disease in MB_{SHH-Child}; *MYC* in MB_{Grp3}; M+ in MB_{Grp4}) when considered in the context of our new subgroups. Finally, the biological definition of MB_{SHH-Infant} (<4·3 years) is at odds with current clinical definitions of infant disease (<3 years) and this should prompt consideration in the future as to whether infant treatment protocols are appropriate for MB_{SHH-Infant} patients older than 3 years.¹⁶ Survival modelling in children younger than 3 years is qualitatively different from analysis in those over 3 years of age, because of the heterogeneity of treatment of infant disease. As such, we regarded further risk modelling in this patient group to be outside the scope of this study, to be addressed in future investigations. To our knowledge, no previous study has directly assessed survival of the molecularly-defined MB_{SHH-Infant} subgroup. The overall survival at 5 years that we observed in MB_{SHH-Infant} disease (62%, 95% CI 50–77) is lower than previously reported in an international meta-analysis of the MB_{SHH} subgroup in age-defined infants (<4 years at diagnosis; 77%),³⁰ but these patients were not molecularly defined and, as such, are not directly comparable.

Our survival analysis focused on the 3–16-year-old clinical group who received current conventional therapies: surgical resection followed by adjuvant radiotherapy with or without chemotherapy at diagnosis with curative intent. Combined risk-modelling across all patients in the non-MB_{WNT} or non-MB_{SHH} subgroups identified *MYC* amplification, high-risk methylation subgroup membership, and loss of chromosome 13 as independent risk factors. Survival models incorporating these factors outperformed the clinical risk-stratification used in current clinical trials (HIT-SIOP-PNET5-MB¹⁶) and subgroup-dependent cytogenetic stratification schemes.²⁸

We have defined a risk-stratification of childhood medulloblastoma that allows patients to be assigned into four overarching risk groups. Favourable-risk patients, including both MB_{WNT} and novel non-MB_{WNT} groups, should be urgently considered for therapy-reducing strategies. Very high-risk patients, typically refractory to conventional therapies (eg, amplified *MYCN*, mutated *TP53*, LCA pathology, and M+ disease in MB_{SHH} and

amplified *MYC* in MB_{Grp3}) should be prioritised for alternative upfront treatment strategies. The priority for high-risk patients, comprising the novel MB_{Grp4-HR} and patients with non-amplified *MYC* in the MB_{Grp3-HR} subgroup, and a standard-risk group, comprising all other patients, should be optimisation of current therapies and the application of novel, biologically targeted agents.

We note the limitations of developing survival models in retrospective patient cohorts, who received heterogeneous treatments. Notwithstanding that models were developed using patients aged 3–16 years, who all received maximal surgical resection and craniospinal irradiation with curative intent, caution should be applied to their clinical implementation. We also note the statistical limitations of stratifications identifying small numbers of patients (eg, very high-risk, 13% of patients). Moreover, some of the identified biomarkers (notably loss of chromosome 13) have not previously been reported as prognostic. We therefore emphasise that validation in additional cohorts, and ideally in prospective, uniformly treated patients in clinical trials, is essential. A small number of samples (<5 samples) from this study were used to assist with the creation of the four-subgroup classification consensus.⁵ Similarly, our own publication that described four methylation-dependent subgroups of medulloblastoma⁷ contained 87 samples that overlapped with this study, although the previously published study contained fewer samples (discovery cohort size of 100 and validation cohort size of 130 patients) and DNA methylation profiling was at much lower resolution (1505 vs >400 000 CpG loci).

The existence of novel primary medulloblastoma subgroups, coupled with the characterisation of secondary prognostic features within each group, represents a significant advance in our understanding of medulloblastoma biology and its application in clinical management and future trials design. We provide clear evidence of the shared biology between MB_{Grp3} and MB_{Grp4} which affects clinical behaviour and has significant implications for understanding disease biology, developmental origins, and experimental modelling. These investigations constitute a blueprint for a new consensus in medulloblastoma molecular sub-classification with important implications for future molecular diagnostics and clinical management.

Contributors

ECS, DW, SB, and SCC designed the study and wrote the manuscript. JCL, SC, AJS, and RMH did laboratory experimentation and analysis. ECS, DW, TS, and SN did bioinformatics analysis. SN, DH, GR, and AI prepared figures. TSJ, AJ, and SBW provided central pathology review. BP, AM, AJ, SBW, TSJ, and SB gathered samples and patient data and provided clinical interpretation. All authors contributed to and approved the final manuscript.

Declaration of interests

We declare no competing interests.

Acknowledgments

This study was funded by Cancer Research UK (C8464/A13457), The Tom Grahame Trust, Star for Harris, Action Medical Research, SPARKS, The JGW Patterson Foundation, and The INSTINCT network (co-funded by The Brain Tumour Charity, Great Ormond Street Children's Charity, and Children with Cancer UK). TSJ is supported by the National Institute for Health Research and a Great Ormond Street Hospital UCL Biomedical Research Centre award.

References

- Taylor MD, Northcott PA, Korshunov A, et al. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol* 2012; **123**: 465–72.
- Northcott PA, Korshunov A, Witt H, et al. Medulloblastoma comprises four distinct molecular variants. *J Clin Oncol* 2011; **29**: 1408–14.
- Fattet S, Haberler C, Lecoix P, et al. Beta-catenin status in paediatric medulloblastomas: correlation of immunohistochemical expression with mutational status, genetic profiles, and clinical characteristics. *J Pathol* 2009; **218**: 86–94.
- Cho Y-J, Tsherniak A, Tamayo P, et al. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J Clin Oncol* 2011; **29**: 1424–30.
- Kool M, Koster J, Bunt J, et al. Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS One* 2008; **3**: e3088.
- Thompson MC, Fuller C, Hogg TL, et al. Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *J Clin Oncol* 2006; **24**: 1924–31.
- Schwalbe EC, Williamson D, Lindsey JC, et al. DNA methylation profiling of medulloblastoma allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. *Acta Neuropathol* 2013; **125**: 359–71.
- Hovestadt V, Remke M, Kool M, et al. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol* 2013; **125**: 913–16.
- Pugh TJ, Weeraratne SD, Archer TC, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* 2012; **488**: 106–10.
- Jones DTW, Jäger N, Kool M, et al. Dissecting the genomic complexity underlying medulloblastoma. *Nature* 2012; **488**: 100–05.
- Robinson G, Parker M, Kranenburg TA, et al. Novel mutations target distinct subgroups of medulloblastoma. *Nature* 2012; **488**: 43–48.
- Northcott PA, Jones DTW, Kool M, et al. Medulloblastomics: the end of the beginning. *Nat Rev Cancer* 2012; **12**: 818–34.
- Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization Classification of Tumours of the Central Nervous System: a summary. *Acta Neuropathol* 2016; **131**: 803–20.
- Ellison DW, Onilude OE, Lindsey JC, et al. beta-Catenin status predicts a favourable outcome in childhood medulloblastoma: the United Kingdom Children's Cancer Study Group Brain Tumour Committee. *J Clin Oncol* 2005; **23**: 7951–57.
- Ellison DW, Kocak M, Dalton J, et al. Definition of disease-risk stratification groups in childhood medulloblastoma using combined clinical, pathologic, and molecular variables. *J Clin Oncol* 2011; **29**: 1400–07.
- Pizer BL, Clifford SC. The potential impact of tumour biology on improved clinical practice for medulloblastoma: progress towards biologically driven clinical trials. *Br J Neurosurg* 2009; **23**: 364–75.
- Robinson GW, Orr BA, Wu G, et al. Vismodegib exerts targeted efficacy against recurrent sonic hedgehog-subgroup medulloblastoma: results from phase II pediatric brain tumor consortium studies PBTC-025B and PBTC-032. *J Clin Oncol* 2015; **33**: 2646–54.
- Kimura H, Ng JMY, Curran T. Transient inhibition of the Hedgehog pathway in young mice causes permanent defects in bone structure. *Cancer Cell* 2008; **13**: 249–60.
- Zhukova N, Ramaswamy V, Remke M, et al. Subgroup-specific prognostic implications of TP53 mutation in medulloblastoma. *J Clin Oncol* 2013; **31**: 2927–35.
- Ramaswamy V, Remke M, Bouffet E, et al. Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. *Acta Neuropathol* 2016; **131**: 821–31.
- von Bueren AO, Kortmann R-D, von Hoff K, et al. Treatment of children and adolescents with metastatic medulloblastoma and prognostic relevance of clinical and biologic parameters. *J Clin Oncol* 2016; **34**: 4151–60.
- Maaten LVD, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008; **9**: 2579–605.
- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**: e47.
- Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* 2015; **8**: 6.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550.
- Kool M, Jones DTW, Jäger N, et al. Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothed inhibition. *Cancer Cell* 2014; **25**: 393–405.
- Heagerty PJ, Saha-Chaudhuri P, Saha-Chaudhuri MP. Package 'survivalROC'. Jan 13, 2013. <https://cran.r-project.org/web/packages/survivalROC/survivalROC.pdf> (accessed March 13, 2017).
- Shih DJH, Northcott PA, Remke M, et al. Cytogenetic prognostication within medulloblastoma subgroups. *J Clin Oncol* 2014; **32**: 886–96.
- Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014; **30**: 1363–69.
- Kool M, Korshunov A, Remke M, et al. Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, group 3, and group 4 medulloblastomas. *Acta Neuropathol* 2012; **123**: 473–84.

References

- Adamski, J., Ramaswamy, V., Huang, A., Bouffet, E., 2014. Advances in managing medulloblastoma and intracranial primitive neuro-ectodermal tumors. *F1000Prime Reports* 6, 56.
- Adan, L., Trivin, C., Sainte-Rose, C., Zucker, J.M., Hartmann, O., Brauner, R., 2001. GH deficiency caused by cranial irradiation during childhood: Factors and markers in young adults. *Journal of Clinical Endocrinology and Metabolism* 86, 5245–5251.
- Archer, T.C., Mahoney, E.L., Pomeroy, S.L., 2017. Medulloblastoma: Molecular Classification-Based Personal Therapeutics. *Neurotherapeutics* 14, 265–273.
- Archer, T.C., Weeraratne, S.D., Pomeroy, S.L., 2012. Hedgehog-GLI pathway in medulloblastoma. *Journal of Clinical Oncology* 30, 2154–2156.
- Ashley, D.M., Merchant, T.E., Strother, D., Zhou, T., Duffner, P., Burger, P.C., Miller, D.C., Lyon, N., Bonner, M.J., Msall, M., Buxton, A., Geyer, R., Kun, L.E., Coleman, L., Pollack, I.F., 2012. Induction chemotherapy and conformal radiation therapy for very young children with nonmetastatic medulloblastoma: Children’s Oncology Group study P9934. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 30, 3181–6.
- Bachmann, I.M., Halvorsen, O.J., Collett, K., Stefansson, I.M., Straume, O., Haukaas, S.A., Salvesen, H.B., Otte, A.P., Akslén, L.A., 2006. EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. *Journal of Clinical Oncology* 24, 268–273.
- Baer, C., Claus, R., Plass, C., 2013. Genome-wide epigenetic regulation of miRNAs in cancer. *Cancer Research* 73, 473–477.
- Bahadur, G., 2000. Age definitions, childhood cancers in relationship to reproductive issues. *Human Reproduction* vol.15 no.1 pp.227–230, 2000 15, 227–230.
- Bailey, C.C., Gnekow, A., Wellek, S., Jones, M., Round, C., Brown, J., Phillips, A., Neidhardt, M.K., 1995. Prospective randomised trial of chemotherapy given before radiotherapy in childhood medulloblastoma. *International society of paediatric oncology (SIOP) and the (German) society of paediatric oncology (GPO): SIOP II. Medical and Pediatric Oncology*

- 25, 166–178.
- Bailey, P., Cushing, H., 1925. Medulloblastoma cerebelli: A common type of midcerebellar glioma of childhood. *Archives of Neurology And Psychiatry* 14, 192–224.
- Baker, S.J., Markowitz, S., Fearon, E.R., Willson, J.K., Vogelstein, B., 1990. Suppression of human colorectal carcinoma cell growth by wild-type p53. *Science (New York, N.Y.)* 249, 912–915.
- Banister, C.E., 2012. Review of Epigenetics: A Reference Manual: A book edited by Jeffrey M. Craig and Nicholas C. Wong. *Epigenetics* 7, 963–964.
- Barber, G.N., 2015. STING: infection, inflammation and cancer. *Nature Reviews Immunology* 2015 15:12 15, 760–770.
- Batora, N.V., Sturm, D., Jones, D.T., Kool, M., Pfister, S.M., Northcott, P.A., 2014. Transitioning from genotypes to epigenotypes: Why the time has come for medulloblastoma epigenomics.
- Bautista, F., Fioravanti, V., de Rojas, T., Carceller, F., Madero, L., Lassaletta, A., Moreno, L., 2017. Medulloblastoma in children and adolescents: a systematic review of contemporary phase I and II clinical trials and biology update. *Cancer Medicine* 6, 2606–2624.
- Baylin, S.B., Jones, P.A., 2011. A decade of exploring the cancer epigenome — biological and translational implications .
- Beauchamp, E.M., Ringer, L., Bulut, G., Sajwan, K.P., Hall, M.D., Lee, Y.C., Peaceman, D., Ozdemirli, M., Rodriguez, O., Macdonald, T.J., Albanese, C., Toretsky, J.A., Uren, A., 2011. Arsenic trioxide inhibits human cancer cell growth and tumor development in mice by blocking Hedgehog/GLI pathway. *The Journal of clinical investigation* 121, 148–60.
- Benayoun, B.A., Pollina, E.A., Brunet, A., 2015. Epigenetic regulation of ageing: Linking environmental inputs to genomic stability. [arXiv:15334406](https://arxiv.org/abs/15334406).
- Bernstein, B.E., Meissner, A., Lander, E.S., 2007. The Mammalian Epigenome. [arXiv:NIHMS150003](https://arxiv.org/abs/NIHMS150003).
- Beroukhi, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S., Du, J., Kau, T., Thomas, R.K., Shah, K., Soto, H., Perner, S., Prensner, J., Debiasi, R.M., Demichelis, F., Hatton, C., Rubin, M.A., Garraway, L.A., Nelson, S.F., Liao, L., Mischel, P.S., Cloughesy, T.F., Meyerson, M., Golub, T.A., Lander, E.S., Mellinghoff, I.K., Sellers, W.R., 2007. Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proceedings of the National Academy of Sciences of the United States of America* 104, 20007.

- Beroukhi, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., McHenry, K.T., Pinchback, R.M., Ligon, A.H., Cho, Y.J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M.S., Weir, B.A., Tanaka, K.E., Chiang, D.Y., Bass, A.J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F.J., Sasaki, H., Tepper, J.E., Fletcher, J.A., Taberner, J., Baselga, J., Tsao, M.S., Demicheli, F., Rubin, M.A., Janne, P.A., Daly, M.J., Nucera, C., Levine, R.L., Ebert, B.L., Gabriele, S., Rustgi, A.K., Antonescu, C.R., Ladanyi, M., Letai, A., Garraway, L.A., Loda, M., Beer, D.G., True, L.D., Okamoto, A., Pomeroy, S.L., Singer, S., Golub, T.R., Lander, E.S., Getz, G., Sellers, W.R., Meyerson, M., 2010. The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905. [arXiv:15334406](#).
- Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., Fan, J.B., Shen, R., 2011. High density DNA methylation array with single CpG site resolution. *Genomics* 98, 288–295.
- Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., Gunderson, K.L., 2009. Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics* .
- Biswas, S., Rao, C.M., 2017. Epigenetics in cancer: Fundamentals and Beyond. *Pharmacology and Therapeutics* 173, 118–134.
- Bland, J.M., Altman, D.G., 2004. The logrank test. *BMJ* 328, 1073.
- Bouffet, E., 2019. Management of high-risk medulloblastoma. *Neurochirurgie* .
- Bovelstad, H., Bøvelstad, H.M., Nygård, S., Størvold, H.L., Aldrin, M., Borgan, Frigessi, A., Lingjærde, O.C., 2007. Predicting survival from microarray data - A comparative study. *Bioinformatics* .
- Boveri, T., 1914. Zur Frage der Entstehung maligner Tumoren. *Die Naturwissenschaften* 2, 801–802.
- Bradburn, M.J., Clark, T.G., Love, S.B., Altman, D.G., 2003. Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer* 89, 431–436.
- Brena, R.M., Costello, J.F., 2007. Genome-epigenome interactions in cancer. *Human molecular genetics* 16 Spec No, R96–105.
- Brentnall, A.R., Cuzick, J., 2018. Use of the concordance index for predictors of censored survival data. *Statistical Methods in Medical Research* 27, 2359–2373.

- Bruce, J.P., Hui, A.B.Y., Shi, W., Perez-Ordóñez, B., Weinreb, I., Xu, W., Haibe-Kains, B., Waggott, D.M., Boutros, P.C., O'Sullivan, B., Waldron, J., Huang, S.H., Chen, E.X., Gilbert, R., Liu, F.F., 2015. Identification of a microRNA signature associated with risk of distant metastasis in nasopharyngeal carcinoma. *Oncotarget* 6, 4537.
- Calin, G.A., Croce, C.M., 2006. MicroRNA signatures in human cancers. *Nat Rev Cancer* 6, 857–866.
- Cambruzzi, E., 2018. Medulloblastoma , WNT-activated / SHH-activated : clinical impact of molecular analysis and histogenetic evaluation , 809–815.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempest, P., Jones, R.S., Zhang, Y., 2002. Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* 298, 1039–1043.
- Capper, D., Jones, D.T., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., Kratz, A., Wefers, A.K., Huang, K., Pajtler, K.W., Schweizer, L., Stichel, D., Olar, A., Engel, N.W., Lindenberg, K., Harter, P.N., Braczynski, A.K., Plate, K.H., Dohmen, H., Garvalov, B.K., Coras, R., Hölsken, A., Hewer, E., Bewerunge-Hudler, M., Schick, M., Fischer, R., Beschorner, R., Schittenhelm, J., Staszewski, O., Wani, K., Varlet, P., Pages, M., Temming, P., Lohmann, D., Selt, F., Witt, H., Milde, T., Witt, O., Aronica, E., Giangaspero, F., Rushing, E., Scheurlen, W., Geisenberger, C., Rodriguez, F.J., Becker, A., Preusser, M., Haberler, C., Bjerkvig, R., Cryan, J., Farrell, M., Deckert, M., Hench, J., Frank, S., Serrano, J., Kannan, K., Tsirigos, A., Brück, W., Hofer, S., Brehmer, S., Seiz-Rosenhagen, M., Hänggi, D., Hans, V., Rozsnoki, S., Hansford, J.R., Kohlhof, P., Kristensen, B.W., Lechner, M., Lopes, B., Mawrin, C., Ketter, R., Kulozik, A., Khatib, Z., Heppner, F., Koch, A., Jouvett, A., Keohane, C., Mühleisen, H., Mueller, W., Pohl, U., Prinz, M., Benner, A., Zapatka, M., Gottardo, N.G., Driever, P.H., Kramm, C.M., Müller, H.L., Rutkowski, S., Von Hoff, K., Frühwald, M.C., Gnekow, A., Fleischhack, G., Tippelt, S., Calaminus, G., Monoranu, C.M., Perry, A., Jones, C., Jacques, T.S., Radlwimmer, B., Gessi, M., Pietsch, T., Schramm, J., Schackert, G., Westphal, M., Reifenberger, G., Wesseling, P., Weller, M., Collins, V.P., Blümcke, I., Bendszus, M., Debus, J., Huang, A., Jabado, N., Northcott, P.A., Paulus, W., Gajjar, A., Robinson, G.W., Taylor, M.D., Jaunmuktane, Z., Ryzhova, M., Platten, M., Unterberg, A., Wick, W., Karajannis, M.A., Mittelbronn, M., Acker, T., Hartmann, C., Aldape, K., Schüller, U., Buslei, R., Lichter, P., Kool, M., Herold-Mende, C., Ellison, D.W., Hasselblatt, M., Snuderl, M., Brandner, S., Korshunov, A., Von Deimling, A., Pfister, S.M., 2018. DNA methylation-based classification of central nervous system tumours. *Nature* 555, 469–474.

- Carlotti, C.G., Smith, C., Rutka, J.T., 2008. The molecular genetics of medulloblastoma: An assessment of new therapeutic targets. *Neurosurgical Review* 31, 359–368.
- Castellone, M.D., Teramoto, H., Williams, B.O., Druey, K.M., Gutkind, J.S., 2005. Prostaglandin E2 promotes colon cancer cell growth through a Gs-axin-beta-catenin signaling axis. *Science (New York, N.Y.)* 310, 1504–10.
- Cavalli, F.M., Remke, M., Rampasek, L., Peacock, J., Shih, D.J., Luu, B., Garzia, L., Torchia, J., Nor, C., Morrissy, A.S., Agnihotri, S., Thompson, Y.Y., Kuzan-Fischer, C.M., Farooq, H., Isaev, K., Daniels, C., Cho, B.K., Kim, S.K., Wang, K.C., Lee, J.Y., Grajkowska, W.A., Perek-Polnik, M., Vasiljevic, A., Faure-Contier, C., Jouvett, A., Giannini, C., Nageswara Rao, A.A., Li, K.K.W., Ng, H.K., Eberhart, C.G., Pollack, I.F., Hamilton, R.L., Gillespie, G.Y., Olson, J.M., Leary, S., Weiss, W.A., Lach, B., Chambless, L.B., Thompson, R.C., Cooper, M.K., Vibhakkar, R., Hauser, P., van Veelen, M.L.C., Kros, J.M., French, P.J., Ra, Y.S., Kumabe, T., López-Aguilar, E., Zitterbart, K., Sterba, J., Finocchiaro, G., Massimino, M., Van Meir, E.G., Osuka, S., Shofuda, T., Klekner, A., Zollo, M., Leonard, J.R., Rubin, J.B., Jabado, N., Albrecht, S., Mora, J., Van Meter, T.E., Jung, S., Moore, A.S., Hallahan, A.R., Chan, J.A., Tirapelli, D.P., Carlotti, C.G., Fouladi, M., Pimentel, J., Faria, C.C., Saad, A.G., Massimi, L., Liau, L.M., Wheeler, H., Nakamura, H., Elbabaa, S.K., Perezpeña-Diazconti, M., Chico Ponce de León, F., Robinson, S., Zapotocky, M., Lassaletta, A., Huang, A., Hawkins, C.E., Tabori, U., Bouffet, E., Bartels, U., Dirks, P.B., Rutka, J.T., Bader, G.D., Reimand, J., Goldenberg, A., Ramaswamy, V., Taylor, M.D., 2017. Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* 31, 737–754.e6.
- Chang, C.H., Housepian, E.M., Herbert, C., 1969. An Operative Staging System and a Megavoltage Radiotherapeutic Technic for Cerebellar Medulloblastomas. *Radiology* 93, 1351–1359.
- Cheang, M.C.U., Chia, S.K., Voduc, D., Gao, D., Leung, S., Snider, J., Watson, M., Davies, S., Bernard, P.S., Parker, J.S., Perou, C.M., Ellis, M.J., Nielsen, T.O., 2009. Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer. *JNCI: Journal of the National Cancer Institute* 101, 736–750.
- Chen, Y.X., Fang, J.Y., Lu, R., Qiu, D.K., 2007. Expression of p21WAF1 is related to acetylation of histone H3 in total chromatin in human colorectal cancer. *World Journal of Gastroenterology* 13, 2209–2213.
- Chen, Z., Yin, X., Li, K., Chen, S., Li, H., Li, Y., Zhang, Q., Wang, H., Qiu, Y., 2018. Serum levels of TRIM72 are lower among patients with colon cancer: Identification of a potential

- diagnostic marker. *Tohoku Journal of Experimental Medicine* 245, 61–68.
- Chervona, Y., Costa, M., 2012. Histone modifications and cancer: biomarkers of prognosis? *American journal of cancer research* 2, 589–97.
- Chial, B.H., Write, P.D., Right, S., Education, N., 2008. Proto-oncogenes to Oncogenes to Cancer.
- Chiang, D.Y., Getz, G., Jaffe, D.B., O’Kelly, M.J.T., Zhao, X., Carter, S.L., Russ, C., Nusbaum, C., Meyerson, M., Lander, E.S., 2009. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nature Methods* 6, 99–103.
- Chin, D., Sklar, C., Donahue, B., Uli, N., Geneiser, N., Allen, J., Nirenberg, A., David, R., Kohn, B., Oberfield, S.E., 1997. Thyroid dysfunction as a late effect in survivors of pediatric medulloblastoma/primitive neuroectodermal tumors: A comparison of hyperfractionated versus conventional radiotherapy. *Cancer* 80, 798–804.
- Cho, Y.J., Tsherniak, A., Tamayo, P., Santagata, S., Ligon, A., Greulich, H., Berhoukim, R., Amani, V., Goumnerova, L., Eberhart, C.G., Lau, C.C., Olson, J.M., Gilbertson, R.J., Gajjar, A., Delattre, O., Kool, M., Ligon, K., Meyerson, M., Mesirov, J.P., Pomeroy, S.L., 2011. Integrative Genomic Analysis of Medulloblastoma Identifies a Molecular Subgroup That Drives Poor Clinical Outcome. *Journal of Clinical Oncology* 29, 1424–1430.
- Cho, Y.W., Hong, T., Hong, S.H., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G.R., Copeland, T.D., Kalkum, M., Ge, K., 2007. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *Journal of Biological Chemistry* 282, 20395–20406.
- Clifford, S.C., Lannering, B., Schwalbe, E.C., Hicks, D., O’ Toole, K., Nicholson, S.L., Goschzik, T., zur Mühlen, A., Figarella-Branger, D., Doz, F., Rutkowski, S., Gustafsson, G., Pietsch, T., 2015. Biomarker-driven stratification of disease-risk in non-metastatic medulloblastoma: Results from the multi-center HIT-SIOP-PNET4 clinical trial. *Oncotarget* 6, 38827–38839.
- Clifford, S.C., Lusher, M.E., Lindsey, J.C., Langdon, J.A., Gilbertson, R.J., Straughton, D., Ellison, D.W., 2006. Wnt/Wingless Pathway Activation and Chromosome 6 Loss Characterise a Distinct Molecular Sub-Group of Medulloblastomas Associated with a Favourable Prognosis. *Cell Cycle* 5, 2666–2670.
- Colditz, G., 2014. *Encyclopedia of Cancer and Society*. Encyclopedia of Cancer and Society

- Cortessis, V.K., Thomas, D.C., Joan Levine, A., Breton, C.V., Mack, T.M., Siegmund, K.D., Haile, R.W., Laird, P.W., 2012. Environmental epigenetics: Prospects for studying epigenetic mediation of exposure-response relationships.
- Cox, D.R., 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 187–202.
- Crawford, J.R., MacDonald, T.J., Packer, R.J., 2007. Medulloblastoma in childhood: new biological advances. *The Lancet Neurology* 6, 1073–1085.
- Croce, C.M., 1995. Oncogenes and cancer. *Science (New York, N.Y.)* 267, 1408–1409.
- CRUK, 2018. Cancer Research UK-Available at: <https://www.cancerresearchuk.org/about-cancer>. Accessed 23-02-2018.
- De Luca, A., Parmigiani, E., Buffo, A., Cerrato, V., Fuca, E., Leto, K., 2016. Sonic hedgehog patterning during cerebellar development , 291–303.
- De Visser, K.E., Eichten, A., Coussens, L.M., 2006. Paradoxical roles of the immune system during cancer development. *Nature Reviews Cancer* 2006 6:1 6, 24–37.
- DeBerardinis, R.J., Lum, J.J., Hatzivassiliou, G., Thompson, C.B., 2008. The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metabolism* 7, 11–20.
- DeNardo, D.G., Andreu, P., Coussens, L.M., 2010. Interactions between lymphocytes and myeloid cells regulate pro-versus anti-tumor immunity. *Cancer and Metastasis Reviews* 29, 309–316.
- Diede, S.J., Guenthoer, J., Geng, L.N., Mahoney, S.E., Marotta, M., Olson, J.M., Tanaka, H., Tapscott, S.J., 2010. DNA methylation of developmental genes in pediatric medulloblastomas identified by denaturation analysis of methylation differences. *Proceedings of the National Academy of Sciences* 107, 234–239.
- Dijkgraaf, G.J., Aliche, B., Weinmann, L., Januario, T., West, K., Modrusan, Z., Burdick, D., Goldsmith, R., Robarge, K., Sutherlin, D., Scales, S.J., Gould, S.E., Yauch, R.L., De Sauvage, F.J., 2011. Small molecule inhibition of GDC-0449 refractory smoothed mutants and downstream mechanisms of drug resistance. *Cancer Research* 71, 435–444.
- Du, P., Zhang, X., Huang, C.C., Jafari, N., Kibbe, W.A., Hou, L., Lin, S.M., 2010. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11.
- Dubuc, A.M., Remke, M., Korshunov, A., Northcott, P.A., Zhan, S.H., Mendez-Lago, M., Kool, M., Jones, D.T., Unterberger, A., Morrissy, A.S., Shih, D., Peacock, J., Ramaswamy,

- V., Rolider, A., Wang, X., Witt, H., Hielscher, T., Hawkins, C., Vibhakar, R., Croul, S., Rutka, J.T., Weiss, W.A., Jones, S.J., Eberhart, C.G., Marra, M.A., Pfister, S.M., Taylor, M.D., 2013. Aberrant patterns of H3K4 and H3K27 histone lysine methylation occur across subgroups in medulloblastoma. *Acta Neuropathologica* 125, 373–384.
- Dvorak, H.F., Flier, J.S., Underhill, L.H., 1986. Tumors: Wounds That Do Not Heal. *New England Journal of Medicine* 315, 1650–1659.
- van Dyk, E., Hoogstraat, M., ten Hoeve, J., Reinders, M.J.T., Wessels, L.F.A., 2016. RUBIC identifies driver genes by detecting recurrent DNA copy number breaks. *Nature Communications* 2016 7:1 7, 1–10.
- Eberhart, C.G., Cohen, K.J., Tihan, T., Goldthwaite, P.T., Burger, P.C., 2003. Medulloblastomas with systemic metastases: Evaluation of tumor histopathology and clinical behavior in 23 patients. *Journal of Pediatric Hematology/Oncology* 25, 198–203.
- Edelstein, K., Spiegler, B.J., Fung, S., Panzarella, T., Mabbott, D.J., Jewitt, N., D'Agostino, N.M., Mason, W.P., Bouffet, E., Tabori, U., Laperriere, N., Hodgson, D.C., 2011. Early aging in adult survivors of childhood medulloblastoma: long-term neurocognitive, functional, and physical outcomes. *Neuro-oncology* 13, 536–45.
- Ellison, D., 2002. Classifying the medulloblastoma: Insights from morphology and molecular genetics.
- Ellison, D.W., Kocak, M., Dalton, J., Megahed, H., Lusher, M.E., Ryan, S.L., Zhao, W., Nicholson, S.L., Taylor, R.E., Bailey, S., Clifford, S.C., 2011. Definition of disease-risk stratification groups in childhood medulloblastoma using combined clinical, pathologic, and molecular variables. *Journal of Clinical Oncology* 29, 1400–1407.
- Ellison, D.W., Onilude, O.E., Lindsey, J.C., Lusher, M.E., Weston, C.L., Taylor, R.E., Pearson, A.D., Clifford, S.C., 2005. Beta-catenin status predicts a favorable outcome in childhood medulloblastoma: The United Kingdom Children's Cancer Study Group Brain Tumour Committee. *Journal of Clinical Oncology* 23, 7951–7957.
- Ensor, J.E., 2014. Biomarker Validation: Common Data Analysis Concerns. *The Oncologist* 19, 886–891.
- Esquela-Kerscher, A., Slack, F.J., 2006. Oncomirs - MicroRNAs with a role in cancer.
- Esteller, M., 2007. Epigenetic gene silencing in cancer: The DNA hypermethylation.
- Fan, S., Chi, W., 2016. Methods for genome-wide DNA methylation analysis in human cancer. *Briefings in Functional Genomics* 15, 432–442.
- Fearon, E.R., 1997. Human cancer syndromes: Clues to the origin and nature of cancer.

- Fearon, E.R., Vogelstein, B., 1990. A genetic model for colorectal tumorigenesis. *Cell* 61, 759–767.
- Feber, A., Guilhamon, P., Lechner, M., Fenton, T., Wilson, G.A., Thirlwell, C., Morris, T.J., Flanagan, A.M., Teschendorff, A.E., Kelly, J.D., Beck, S., 2014. Using high-density DNA methylation arrays to profile copy number alterations. *Genome Biology* 15, R30—R30.
- Feinberg, A.P., Ohlsson, R., Henikoff, S., 2006. The epigenetic progenitor origin of human cancer.
- Ferguson-Smith, A.C., 2011. Genomic imprinting: The emergence of an epigenetic paradigm.
- Fernández-Aceñero, M.J., Cruz, M., Sastre-Varela, J., Casal, J.I., Nieto, M.A., del Puerto-Nevado, L., García-Foncillas, J., Cebrián, A., 2019. TRIM72 Immunohistochemical Expression Can Predict Relapse in Colorectal Carcinoma. *Pathology and Oncology Research* .
- Fleiss, J., 1981. The measurement of interrater agreement. In *Statistical Methods for Rates and Proportions*, 2nd Edition, John Wiley, NewYork, 212-236.
- Fortin, J.P., Hansen, K.D., 2014. Minfi tutorial BioC2014. Technical Report.
- Fossati, P., Ricardi, U., Orecchia, R., 2009. Pediatric medulloblastoma: Toxicity of current treatment and potential role of protontherapy.
- Frank, B., Bermejo, J.L., Hemminki, K., Sutter, C., Wappenschmidt, B., Meindl, A., Kiechle-Bahat, M., Bugert, P., Schmutzler, R.K., Bartram, C.R., Burwinkel, B., 2007. Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. *Carcinogenesis* 28, 1442–1445.
- Frühwald, M.C., Rutkowski, S., 2011. Tumors of the central nervous system in children and adolescents. *Deutsches Arzteblatt international* 108, 390–7.
- Fuccillo, M., Joyner, A.L., Fishell, G., 2006. Morphogen to mitogen: The multiple roles of hedgehog signalling in vertebrate neural development.
- Gabriel, A.S., Lafta, F.M., Schwalbe, E.C., Nakjang, S., Cockell, S.J., Iliasova, A., Enshaei, A., Schwab, C., Rand, V., Clifford, S.C., Kinsey, S.E., Mitchell, C.D., Vora, A., Harrison, C.J., Moorman, A.V., Strathdee, G., 2015. Epigenetic landscape correlates with genetic subtype but does not predict outcome in childhood acute lymphoblastic leukemia. *Epigenetics* .
- Gajjar, A., Chintagumpala, M., Ashley, D., Kellie, S., Kun, L.E., Merchant, T.E., Woo, S., Wheeler, G., Ahern, V., Krasin, M.J., Fouladi, M., Broniscer, A., Krance, R., Hale, G.A., Stewart, C.F., Dauser, R., Sanford, R.A., Fuller, C., Lau, C., Boyett, J.M., Wallace,

- D., Gilbertson, R.J., 2006. Risk-adapted craniospinal radiotherapy followed by high-dose chemotherapy and stem-cell rescue in children with newly diagnosed medulloblastoma (St Jude Medulloblastoma-96): long-term results from a prospective, multicentre trial. *The Lancet. Oncology* 7, 813–20.
- Gajjar, S., Mazloom, A., Chintagumpala, M., Mahajan, A., Paulino, A.C., 2014. Secondary glioblastoma multiform in a patient with CHARGE syndrome and prior radiation therapy for medulloblastoma.
- Gareth James, Trevor Hastie, Robert Tibshirani, D.W., 2014. *An Introduction to Statistical Learning: With Applications in R*. *Journal of Agricultural, Biological, and Environmental Statistics* .
- Gautreau, A., Louvard, D., Arpin, M., 2002. ERM proteins and NF2 tumor suppressor: the Yin and Yang of cortical actin organization and cell growth signaling. *Current opinion in cell biology* 14, 104–109.
- Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N.Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C.R., Nohadani, M., Eklund, A.C., Spencer-Dene, B., Clark, G., Pickering, L., Stamp, G., Gore, M., Szallasi, Z., Downward, J., Futreal, P.A., Swanton, C., 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine* .
- Gibson, F., Walsh, J., Mburu, P., Varela, A., Brown, K.A., Antonio, M., Beisel, K.W., Steel, K.P., Brown, S.D., 1995. A type VII myosin encoded by the mouse deafness gene shaker-1. *Nature* 1995 374:6517 374, 62–64.
- Gibson, P., Tong, Y., Robinson, G., Thompson, M.C., Currele, D.S., Eden, C., Kranenburg, T.A., Hogg, T., Poppleton, H., Martin, J., Finkelstein, D., Pounds, S., Weiss, A., Patay, Z., Scoggins, M., Ogg, R., Pei, Y., Yang, Z.J., Brun, S., Lee, Y., Zindy, F., Lindsey, J.C., Taketo, M.M., Boop, F.A., Sanford, R.A., Gajjar, A., Clifford, S.C., Rouse, M.F., McKinnon, P.J., Gutmann, D.H., Ellison, D.W., Wechsler-Reya, R., Gilbertson, R.J., 2010. Subtypes of medulloblastoma have distinct developmental origins. *Nature* 468, 1095–1099.
- Gilbert, M.T.P., Haselkorn, T., Bunce, M., Sanchez, J.J., Lucas, S.B., Jewell, L.D., van Marck, E., Worobey, M., 2007. The Isolation of Nucleic Acids from Fixed, Paraffin-Embedded Tissues-Which Methods Are Useful When? *PLoS ONE* 2, 537.
- Gilbertson, R.J., Ellison, D.W., 2008. The Origins of Medulloblastoma Subtypes. *Annual Review of Pathology: Mechanisms of Disease* 3, 341–365.

- GoldenHelix, 2014. CNV Quality Assurance Tutorial. SNP Variation Suite User Guide , 2.
- Gorlin, R.J., Goltz, R.W., 1960. Multiple Nevoid Basal-Cell Epithelioma, Jaw Cysts and Bifid Rib. *New England Journal of Medicine* 262, 908–912.
- Goschzik, T., Schwalbe, E.C., Hicks, D., Smith, A., zur Muehlen, A., Figarella-Branger, D., Doz, F., Rutkowski, S., Lannering, B., Pietsch, T., Clifford, S.C., 2018. Prognostic effect of whole chromosomal aberration signatures in standard-risk, non-WNT/non-SHH medulloblastoma: a retrospective, molecular analysis of the HIT-SIOP PNET 4 trial. *The Lancet Oncology* 19.
- Greaves, M., Maley, C.C., 2012. Clonal evolution in cancer. *Nature* 481, 306–313.
- Grivennikov, S.I., Greten, F.R., Karin, M., 2010. Immunity, Inflammation, and Cancer. *Cell* 140, 883–899.
- Gschwind, A., Fischer, O.M., Ullrich, A., 2004. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nature reviews. Cancer* 4, 361–370.
- Guil, S., Esteller, M., 2009. DNA methylomes, histone codes and miRNAs: Tying it all together.
- Hanahan, D., Folkman, J., 1996. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* 86, 353–64.
- Hanahan, D., Weinberg, R.A., 2000. The Hallmarks of Cancer. *Cell* 100, 57–70. [arXiv:0208024](https://arxiv.org/abs/0208024).
- Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell* 144, 646–74.
- Hargreaves, D.C., Crabtree, G.R., 2011. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Research* 21:3 21, 396–420.
- Hegi, M.E., Diserens, A.C., Gorlia, T., Hamou, M.F., de Tribolet, N., Weller, M., Kros, J.M., Hainfellner, J.A., Mason, W., Mariani, L., Bromberg, J.E., Hau, P., Mirimanoff, R.O., Cairncross, J.G., Janzer, R.C., Stupp, R., 2005. MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *New England Journal of Medicine* 352, 997–1003.
- Heikens, J., Michiels, E.M.C., Behrendt, H., Endert, E., Bakker, P.J.M., Fliers, E., 1998. Long-term neuro-endocrine sequelae after treatment for childhood medulloblastoma. *European Journal of Cancer* 34, 1592–1597.
- Hilgenfeld, R., 1995. Regulatory GTPases. *Current Opinion in Structural Biology* 5, 810–817.

- Hill, R.M., Kuijper, S., Lindsey, J.C., Petrie, K., Schwalbe, E.C., Barker, K., Boulton, J.K., Williamson, D., Ahmad, Z., Hallsworth, A., Ryan, S.L., Poon, E., Robinson, S.P., Ruddle, R., Raynaud, F.I., Howell, L., Kwok, C., Joshi, A., Nicholson, S.L., Crosier, S., Ellison, D.W., Wharton, S.B., Robson, K., Michalski, A., Hargrave, D., Jacques, T.S., Pizer, B., Bailey, S., Swartling, F.J., Weiss, W.A., Chesler, L., Clifford, S.C., 2015. Combined MYC and P53 defects emerge at medulloblastoma relapse and define rapidly progressive, therapeutically targetable disease. *Cancer Cell* 27, 72–84.
- Hosmer, D.W., Lemeshow, S., 2005. *Applied Logistic Regression*.
- Hovestadt, V., Jones, D.T., Picelli, S., Wang, W., Kool, M., Northcott, P.A., Sultan, M., Stachurski, K., Ryzhova, M., Warnatz, H.J., Ralser, M., Brun, S., Bunt, J., Jäger, N., Kleinheinz, K., Erkek, S., Weber, U.D., Bartholomae, C.C., Von Kalle, C., Lawrenz, C., Eils, J., Koster, J., Versteeg, R., Milde, T., Witt, O., Schmidt, S., Wolf, S., Pietsch, T., Rutkowski, S., Scheurlen, W., Taylor, M.D., Brors, B., Felsberg, J., Reifenberger, G., Borkhardt, A., Lehrach, H., Wechsler-Reya, R.J., Eils, R., Yaspo, M.L., Landgraf, P., Korshunov, A., Zapatka, M., Radlwimmer, B., Pfister, S.M., Lichter, P., 2014. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* .
- Hovestadt, V., Zapatka, M., 2015. conumee: Enhanced copy-number variation analysis using Illumina 450k methylation arrays. R package version 0.99.
- Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., Eden, A., 2008. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* 27, 404–408.
- Howlader, N., Noone, A., Krapcho, M., Miller, D., Bishop, K., Kosary, C., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D., Chen, H., Feuer, E., Cronin, K., 2016. *Cancer Statistics Review, 1975-2014 - SEER Statistics*, National Cancer Institute.
- Huang, G.H., Xu, Q.F., Cui, Y.H., Li, N., Bian, X.W., Lv, S.Q., 2016. Medulloblastoma stem cells: Promising targets in medulloblastoma therapy. *Cancer Science* 107, 583–589.
- Huang, S.M.A., Mishina, Y.M., Liu, S., Cheung, A., Stegmeier, F., Michaud, G.A., Charlat, O., Wiellette, E., Zhang, Y., Wiessner, S., Hild, M., Shi, X., Wilson, C.J., Mickanin, C., Myer, V., Fazal, A., Tomlinson, R., Serluca, F., Shao, W., Cheng, H., Shultz, M., Rau, C., Schirle, M., Schlegl, J., Ghidelli, S., Fawell, S., Lu, C., Curtis, D., Kirschner, M.W., Lengauer, C., Finan, P.M., Tallarico, J.A., Bouwmeester, T., Porter, J.A., Bauer, A., Cong, F., 2009. Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* 461, 614–620.

- Huarte, M., 2015. The emerging role of lncRNAs in cancer.
- Hunter, D.J., Losina, E., Guermazi, A., Burstein, D., Lassere, M.N., Kraus, V., 2012. NIH Public Access 11, 536–545.
- Hupé, P., Stransky, N., Thiery, J.P., Radvanyi, F., Barillot, E., 2004. Analysis of array CGH data: From signal ratio to gain and loss of DNA regions. *Bioinformatics* 20, 3413–3422.
- Hur, Y., Lee, H., 2011. Wavelet-based identification of DNA focal genomic aberrations from single nucleotide polymorphism arrays. *BMC Bioinformatics* 12, 1–14.
- Illumina, 2020. Infinium™ Mouse Methylation BeadChip. www.illumina.com. Accessed 20-11-2021.
- Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., Weissman, J.S., 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols* 7, 1534–1550.
- Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddloh, J.A., Wen, B., Feinberg, A.P., 2008. Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Research* 18, 780–790.
- Jiang, H., Zeng, X., He, N., 2012. An Integrated View of Current Progress in Copy Number Variations Analysis of Genome. *Advanced Science Letters* 7, 1–8.
- Jones, D.T., Jäger, N., Kool, M., Zichner, T., Hutter, B., Sultan, M., Cho, Y.J., Pugh, T.J., Hovestadt, V., Stütz, A.M., Rausch, T., Warnatz, H.J., Ryzhova, M., Bender, S., Sturm, D., Pleier, S., Cin, H., Pfaff, E., Sieber, L., Wittmann, A., Remke, M., Witt, H., Hutter, S., Tzaridis, T., Weischenfeldt, J., Raeder, B., Avci, M., Amstislavskiy, V., Zapatka, M., Weber, U.D., Wang, Q., Lasitschka, B., Bartholomae, C.C., Schmidt, M., von Kalle, C., Ast, V., Lawerenz, C., Eils, J., Kabbe, R., Benes, V., van Sluis, P., Koster, J., Volckmann, R., Shih, D., Betts, M.J., Russell, R.B., Coco, S., Tonini, G.P., Schüller, U., Hans, V., Graf, N., Kim, Y.J., Monoranu, C., Roggendorf, W., Unterberg, A., Herold-Mende, C., Milde, T., Kulozik, A.E., von Deimling, A., Witt, O., Maass, E., Rössler, J., Ebinger, M., Schuhmann, M.U., Frühwald, M.C., Hasselblatt, M., Jabado, N., Rutkowski, S., von Bueren, A.O., Williamson, D., Clifford, S.C., Mc Cabe, M.G., Collins, V.P., Wolf, S., Wiemann, S., Lehrach, H., Brors, B., Scheurlen, W., Felsberg, J., Reifenberger, G., Northcott, P.A., Taylor, M.D., Meyerson, M., Pomeroy, S.L., Yaspo, M.L., Korbel, J.O., Korshunov, A., Eils, R., Pfister, S.M., Lichter, P., 2012. Dissecting the genomic complexity underlying medulloblastoma. *Nature* 488, 100–105.

- Jones, M.J.K., Jallepalli, P.V., 2012. Chromothripsis: chromosomes in crisis. *Developmental cell* 23, 908–17.
- Jones, P.A., Baylin, S.B., 2002. The fundamental role of epigenetic events in cancer.
- Jones, P.A., Baylin, S.B., 2007a. The epigenomics of cancer. *An Omics Perspective on Cancer Research* , 683–692 [arXiv:NIHMS150003](#).
- Jones, P.A., Baylin, S.B., 2007b. The Epigenomics of Cancer. *Cell* 128, 683–692.
- Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M., Pulice, J.L., Chory, E.J., Crabtree, G.R., 2017. Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nature Genetics* 49, 213–222. [arXiv:15334406](#).
- Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., Pinkel, D., 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258, 818–821.
- Kaplan, E.L., Meier, P., 1958. Nonparametric Estimation from Incomplete Observations **NONPARAMETRIC ESTIMATION FROM INCOMPLETE OBSERVATIONS***. Source: *Journal of the American Statistical Association* 53, 457–481.
- Kastan, M.B., 2008. DNA Damage Responses: Mechanisms and Roles in Human Disease. *Molecular Cancer Research* 6, 517–524.
- Kawauchi, D., Robinson, G., Uziel, T., Gibson, P., Rehg, J., Gao, C., Finkelstein, D., Qu, C., Pounds, S., Ellison, D.W., Gilbertson, R.J., Roussel, M.F., 2012. A Mouse Model of the Most Aggressive Subgroup of Human Medulloblastoma. *Cancer Cell* 21, 168–180.
- Kemp, C.J., Fero, M.L., Randel, E., Gurley, K.E., Roberts, J.M., 1998. The murine gene p27(Kip 1) is haplo-insufficient for tumour suppression. *Nature* 396, 177–180.
- Kiltie, A.E., Lashford, L.S., Gattamaneni, H.R., 1997. Survival and late effects in medulloblastoma patients treated with craniospinal irradiation under three years old. *Medical and Pediatric Oncology* 28, 348–354.
- Kim, J., Lee, J.J., Kim, J., Gardner, D., Beachy, P.A., 2010. Arsenic antagonizes the Hedgehog pathway by preventing ciliary accumulation and reducing stability of the Gli2 transcriptional effector. *Proceedings of the National Academy of Sciences* 107, 13432–13437.
- Kim, T.M., Xi, R., Luquette, L.J., Park, R.W., Johnson, M.D., Park, P.J., 2013. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome research* 23, 217–27.

- Kleihues, P., Louis, D.N., Scheithauer, B.W., Rorke, L.B., Reifenberger, G., Burger, P.C., Cavenee, W.K., 2002. The WHO classification of tumors of the nervous system. *Journal of Neuropathology and Experimental Neurology* 61, 215–225.
- Kling, T., Wenger, A., Beck, S., Carén, H., 2017. Validation of the MethylationEPIC Bead-Chip for fresh-frozen and formalin-fixed paraffin-embedded tumours. *Clinical Epigenetics* 9, 1–6.
- Knudson, A.G., 1971. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proceedings of the National Academy of Sciences* 68, 820–823.
- Kogerman, P., Grimm, T., Kogerman, L., Krause, D., Undén, A.B., Sandstedt, B., Toftgård, R., Zaphiropoulos, P.G., 1999. Mammalian Suppressor-of-Fused modulates nuclear–cytoplasmic shuttling of GLI-1. *Nature Cell Biology* 1, 312–319.
- Kool, M., Korshunov, A., Remke, M., Jones, D., Schlanstein, M., Northcott, P., Cho, Y.J., Koster, J., Schouten-van Meeteren, A., van Vuurden, D., Clifford, S., Pietsch, T., von Bueren, A., Rutkowski, S., McCabe, M., Collins, V., Bäcklund, M., Haberler, C., Bourdeaut, F., Delattre, O., Doz, F., Ellison, D., Gilbertson, R., Pomeroy, S., Taylor, M., Lichter, P., Pfister, S., 2012. Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. *Acta Neuropathologica* 123, 473–484.
- Lamont, J.M., McManamy, C.S., Pearson, A.D., Clifford, S.C., Ellison, D.W., 2004. Combined histopathological and molecular cytogenetic stratification of medulloblastoma patients. *Clinical Cancer Research* 10, 5482–5493.
- Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159.
- Laneve, P., Po, A., Favia, A., Legnini, I., Alfano, V., Rea, J., Carlo, V.D., Bevilacqua, V., Miele, E., Mastronuzzi, A., Carai, A., Locatelli, F., Bozzoni, I., Ferretti, E., Caffarelli, E., 2017. The long noncoding RNA linc-NeD125 controls the expression of medulloblastoma driver genes by microRNA sponge activity. *Oncotarget* 8, 31003–31015.
- Lannering, B., Rutkowski, S., Doz, F., Pizer, B., Gustafsson, G., Navajas, A., Massimino, M., Reddingius, R., Benesch, M., Carrie, C., Taylor, R., Gandola, L., Bjork-Eriksson, T., Giralt, J., Oldenburger, F., Pietsch, T., Figarella-Branger, D., Robson, K., Forni, M., Clifford, S.C., Warmuth-Metz, M., Von Hoff, K., Faldum, A., Mosseri, V., Kortmann, R., 2012. Hyperfractionated versus conventional radiotherapy followed by chemotherapy in standard-risk medulloblastoma: Results from the randomized multicenter HIT-SIOP PNET 4 trial. *Journal of Clinical Oncology* 30, 3187–3193.

- Lau, J.C., Hanel, M.L., Wevrick, R., 2004. Tissue-specific and imprinted epigenetic modifications of the human NDN gene. *Nucleic Acids Research* 32, 3376–3382.
- Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S., Li, C., Amamoto, R., Peters, D.T., Turczyk, B.M., Marblestone, A.H., Inverso, S.A., Bernard, A., Mali, P., Rios, X., Aach, J., Church, G.M., 2014. Highly multiplexed subcellular RNA sequencing in situ. *Science* 343, 1360–1363.
- Lee, J.W., Devanarayan, V., Barrett, Y.C., Weiner, R., Allinson, J., Fountain, S., Keller, S., Weinryb, I., Green, M., Duan, L., Rogers, J.A., Millham, R., O'Brien, P.J., Sailstad, J., Khan, M., Ray, C., Wagner, J.A., 2006. Fit-for-Purpose Method Development and Validation for Successful Biomarker Measurement. *Pharmaceutical Research* 23, 312–328.
- Lemckert, F.A., Bournazos, A., Eckert, D.M., Kenzler, M., Hawkes, J.M., Butler, T.L., Ceely, B., North, K.N., Winlaw, D.S., Egan, J.R., Cooper, S.T., 2016. Lack of MG53 in human heart precludes utility as a biomarker of myocardial injury or endogenous cardioprotective factor. *Cardiovascular Research* 110, 178–187.
- Leung, C., Lingbeek, M., Shakhova, O., Liu, J., Tanger, E., Saremaslani, P., Van Lohuizen, M., Marino, S., 2004. *Bmi1* is essential for cerebellar development and is overexpressed in human medulloblastomas. *Nature* 428, 337–341.
- Lev Maor, G., Yearim, A., Ast, G., 2015. The alternative role of DNA methylation in splicing regulation. *Trends in genetics : TIG* 31, 274–280.
- Lim, J., Hao, T., Shaw, C., Patel, A.J., Szabó, G., Rual, J.F., Fisk, C.J., Li, N., Smolyar, A., Hill, D.E., Barabási, A.L., Vidal, M., Zoghbi, H.Y., 2006. A Protein-Protein Interaction Network for Human Inherited Ataxias and Disorders of Purkinje Cell Degeneration. *Cell* 125, 801–814.
- Lin, C.Y., Erkek, S., Tong, Y., Yin, L., Federation, A.J., Zapatka, M., Haldipur, P., Kawauchi, D., Risch, T., Warnatz, H.J., Worst, B.C., Ju, B., Orr, B.A., Zeid, R., Polaski, D.R., Segura-Wang, M., Waszak, S.M., Jones, D.T., Kool, M., Hovestadt, V., Buchhalter, I., Sieber, L., Johann, P., Chavez, L., Gröschel, S., Ryzhova, M., Korshunov, A., Chen, W., Chizhikov, V.V., Millen, K.J., Amstislavskiy, V., Lehrach, H., Yaspo, M.L., Eils, R., Lichter, P., Korb, J.O., Pfister, S.M., Bradner, J.E., Northcott, P.A., 2016. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* 530, 57–62.
- Lin, T.L., Matsui, W., 2012. Hedgehog pathway as a drug target: Smoothed inhibitors in development.

- Lindsey, J.C., Lusher, M.E., Anderton, J.A., Bailey, S., Gilbertson, R.J., Pearson, A.D., Ellison, D.W., Clifford, S.C., 2004. Identification of tumour-specific epigenetic events in medulloblastoma development by hypermethylation profiling. *Carcinogenesis* 25, 661–668.
- Liu, J., Pan, S., Hsieh, M.H., Ng, N., Sun, F., Wang, T., Kasibhatla, S., Schuller, A.G., Li, A.G., Cheng, D., Li, J., Tompkins, C., Pferdekamper, A., Steffy, A., Cheng, J., Kowal, C., Phung, V., Guo, G., Wang, Y., Graham, M.P., Flynn, S., Brenner, J.C., Li, C., Villarroel, M.C., Schultz, P.G., Wu, X., McNamara, P., Sellers, W.R., Petruzzelli, L., Boral, A.L., Seidel, H.M., McLaughlin, M.E., Che, J., Carey, T.E., Vanasse, G., Harris, J.L., 2013. Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. *Proceedings of the National Academy of Sciences* 110, 20224–20229.
- Liu, Y., Wei, X., Guan, L., Xu, S., Yuan, Y., Lv, D., He, X., Zhan, J., Kong, Y., Guo, J., Zhang, H., 2018. Unconventional myosin VIIA promotes melanoma progression. *Journal of Cell Science* .
- Lodish, H.F., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., James, D., 2013. *Molecular Cell Biology*, 7th edition. volume 5. [arXiv:arXiv:1011.1669v3](#).
- Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., Kleihues, P., 2007. The 2007 WHO classification of tumours of the central nervous system. [arXiv:arXiv:1011.1669v3](#).
- Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W., 2016. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica* 131, 803–820. [arXiv:15334406](#).
- Lowe, S.W., Bodis, S., McClatchey, a., Remington, L., Ruley, H.E., Fisher, D.E., Housman, D.E., Jacks, T., 1994. P53 Status and the Efficacy of Cancer Therapy in Vivo.
- Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., Golub, T.R., 2005. MicroRNA expression profiles classify human cancers. *Nature* 435, 834–838. [arXiv:1512.00567](#).
- MacFarlane, L.A., R. Murphy, P., 2010. MicroRNA: Biogenesis, Function and Role in Cancer. *Current Genomics* 11, 537–561.
- Machold, R., Fishell, G., 2002. Hedgehog patterns midbrain ARChitecture.
- Maher, C.A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., Chinnaiyan, A.M., 2009. Transcriptome sequencing to detect

- gene fusions in cancer. *Nature* 458, 97–101.
- Malkin, D., 2011. Li-fraumeni syndrome. *Genes cancer* 2, 475–84.
- Mantel, N., 1966. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports*. Part 1 .
- Marcus, K.J., Haas-Kogan, D., 2009. Pediatric Radiation Oncology, in: *Oncology of Infancy and Childhood*, pp. 241–255.
- Martin, A.M., Raabe, E., Eberhart, C., Cohen, K.J., 2014. Management of Pediatric and Adult Patients with Medulloblastoma. *arXiv:NIHMS150003*.
- Martin, D.M., 2015. Epigenetic Developmental Disorders: CHARGE syndrome, a case study. *Current genetic medicine reports* 3, 1–7. *arXiv:NIHMS150003*.
- Martín-Subero, J.I., Esteller, M., 2017. Epigenetic Mechanisms in Cancer Development. *The Molecular Basis of Human Cancer* , 263–275.
- Marzouka, N.a.d., Nordlund, J., Bäcklin, C.L., Lönnerholm, G., Syvänen, A.C., Carlsson Almlöf, J., 2016. CopyNumber450kCancer: baseline correction for accurate copy number calling from the 450k methylation array. *Bioinformatics* 32, 1080–1082.
- Massimino, M., Biassoni, V., Gandola, L., Garrè, M.L., Gatta, G., Giangaspero, F., Poggi, G., Rutkowski, S., 2016. Childhood medulloblastoma. *Critical Reviews in Oncology/Hematology* 105, 35–51.
- Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., Dsouza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V.M., Rowitch, D.H., Xing, X., Fiore, C., Schillebeeckx, M., Jones, S.J., Haussler, D., Marra, M.A., Hirst, M., Wang, T., Costello, J.F., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 2010 466:7303 466, 253–257.
- Mburu, P., Liu, X.Z., Walsh, J., Saw, D., Cope, M.J., Gibson, F., Kendrick-Jones, J., Steel, K.P., Brown, S.D., 1997. Mutation analysis of the mouse myosin VIIA deafness gene. *Genes and function* 1, 191–203.
- McCarroll, S.A., Altshuler, D.M., 2007. Copy-number variation and association studies of human disease. *Nature Genetics* 39, S37–S42. *arXiv:NIHMS150003*.
- McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., De Bakker, P.I., Maller, J.B., Kirby, A., Elliott, A.L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P.J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K.W., Rava, R., Daly, M.J., Gabriel, S.B., Altshuler, D., 2008. Integrated

- detection and population-genetic analysis of SNPs and copy number variation. *Nature Genetics* 40, 1166–1174.
- McMahon, A.P., Bradley, A., 1990. The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell* 62, 1073–1085.
- McManamy, C.S., Lamont, J.M., Taylor, R.E., Cole, M., Pearson, A.D., Clifford, S.C., Ellison, D.W., 2003. Morphophenotypic variation predicts clinical behavior in childhood non-desmoplastic medulloblastomas. *Journal of Neuropathology and Experimental Neurology* 62, 627–632.
- McManamy, C.S., Pears, J., Weston, C.L., Hanzely, Z., Ironside, J.W., Taylor, R.E., Grundy, R.G., Clifford, S.C., Ellison, D.W., 2007. Nodule formation and desmoplasia in medulloblastomas - Defining the nodular/desmoplastic variant and its biological behavior. *Brain Pathology* 17, 151–164.
- Meirson, T., Gil-Henn, H., Samson, A.O., 2020. Invasion and metastasis: the elusive hallmark of cancer. *Oncogene* 39, 2024–2026.
- Menyhárt, O., Giangaspero, F., Gyorffy, B., 2019. Molecular markers and potential therapeutic targets in non-WNT/non-SHH (group 3 and group 4) medulloblastomas. *Journal of Hematology and Oncology* 12, 1–17.
- Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., Getz, G., 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biology* 12, R41.
- Miettinen, S., Laurikainen, E., Johansson, R., Minn, H., Laurell, G., Salmi, T.T., 1997. Radiotherapy enhanced ototoxicity of cisplatin in children. *Acta oto-laryngologica. Supplementum* 529, 90–94.
- Miyake, N., Koshimizu, E., Okamoto, N., Mizuno, S., Ogata, T., Nagai, T., Kosho, T., Ohashi, H., Kato, M., Sasaki, G., Mabe, H., Watanabe, Y., Yoshino, M., Matsuishi, T., Takanashi, J.I., Shotelersuk, V., Tekin, M., Ochi, N., Kubota, M., Ito, N., Ihara, K., Hara, T., Tonoki, H., Ohta, T., Saito, K., Matsuo, M., Urano, M., Enokizono, T., Sato, A., Tanaka, H., Ogawa, A., Fujita, T., Hiraki, Y., Kitanaka, S., Matsubara, Y., Makita, T., Taguri, M., Nakashima, M., Tsurusaki, Y., Saitsu, H., Yoshiura, K.I., Matsumoto, N., Niikawa, N., 2013. MLL2 and KDM6A mutations in patients with Kabuki syndrome. *American Journal of Medical Genetics, Part A* 161, 2234–2243.
- Moran, S., Arribas, C., Esteller, M., 2016. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* 8,

389–399.

- Morris, T.J., Beck, S., 2014. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods (San Diego, Calif.)* 72, 3–8.
- Moxon-Emre, I., Taylor, M.D., Bouffet, E., Hardy, K., Campen, C.J., Malkin, D., Hawkins, C., Laperriere, N., Ramaswamy, V., Bartels, U., Scantlebury, N., Janzen, L., Law, N., Walsh, K.S., Mabbott, D.J., 2016. Intellectual outcome in molecular subgroups of medulloblastoma. *Journal of Clinical Oncology* 34, 4161–4170.
- Mukherjee, S., 2010. *The emperor of all maladies : a biography of cancer*. Scribner, New York.
- Murphy, B.L., Obad, S., Bihannic, L., Ayrault, O., Zindy, F., Kauppinen, S., Roussel, M.F., 2013. Silencing of the miR-1792 Cluster Family Inhibits Medulloblastoma Progression. *Cancer Research* 73, 7068–7078.
- Nathan, P.C., Greenberg, M.L., Ness, K.K., Hudson, M.M., Mertens, A.C., Mahoney, M.C., Gurney, J.G., Donaldson, S.S., Leisenring, W.M., Robison, L.L., Oeffinger, K.C., 2008. Medical care in long-term survivors of childhood cancer: A report from the childhood cancer survivor study. *Journal of Clinical Oncology* 26, 4401–4409.
- NCI, 2018. National Cancer Institute.
- Negrini, S., Gorgoulis, V.G., Halazonetis, T.D., 2010. Genomic instability — an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology* 11:3 11, 220–228.
- Neumann, J.E., Swartling, F.J., Schüller, U., 2017. Medulloblastoma: experimental models and reality. *Acta Neuropathologica* 134, 679–689.
- Newton, H.B., 2001. Review of the molecular genetics and chemotherapeutic treatment of adult and paediatric medulloblastoma. *Expert Opinion on Investigational Drugs* 10, 2089–2104.
- Nord, H., Pfeifer, S., Nilsson, P., Sandgren, J., Popova, S., Strömberg, B., Alafuzoff, I., Nistér, M., De Stahl, T.D., 2012. Novel amplifications in pediatric medulloblastoma identified by genome-wide copy number profiling. *Journal of Neuro-Oncology* 107, 37–49.
- Northcott, P.A., Buchhalter, I., Morrissy, A.S., Hovestadt, V., Weischenfeldt, J., Ehrenberger, T., Gröbner, S., Segura-Wang, M., Zichner, T., Rudneva, V.A., Warnatz, H.J., Sidiropoulos, N., Phillips, A.H., Schumacher, S., Kleinheinz, K., Waszak, S.M., Erkek, S., Jones, D.T., Worst, B.C., Kool, M., Zapatka, M., Jäger, N., Chavez, L., Hutter, B., Bieg, M., Paramasivam, N., Heinold, M., Gu, Z., Ishaque, N., Jäger-Schmidt, C., Imbusch,

- C.D., Jugold, A., Hübschmann, D., Risch, T., Amstislavskiy, V., Gonzalez, F.G.R., Weber, U.D., Wolf, S., Robinson, G.W., Zhou, X., Wu, G., Finkelstein, D., Liu, Y., Cavalli, F.M., Luu, B., Ramaswamy, V., Wu, X., Koster, J., Ryzhova, M., Cho, Y.J., Pomeroy, S.L., Herold-Mende, C., Schuhmann, M., Ebinger, M., Liau, L.M., Mora, J., McLendon, R.E., Jabado, N., Kumabe, T., Chuah, E., Ma, Y., Moore, R.A., Mungall, A.J., Mungall, K.L., Thiessen, N., Tse, K., Wong, T., Jones, S.J., Witt, O., Milde, T., Von Deimling, A., Capper, D., Korshunov, A., Yaspo, M.L., Kriwacki, R., Gajjar, A., Zhang, J., Beroukhim, R., Fraenkel, E., Korbel, J.O., Brors, B., Schlesner, M., Eils, R., Marra, M.A., Pfister, S.M., Taylor, M.D., Lichter, P., 2017. The whole-genome landscape of medulloblastoma subtypes. *Nature* 547, 311–317.
- Northcott, P.A., Jones, D.T.W., Kool, M., Robinson, G.W., Gilbertson, R.J., Cho, Y.J., Pomeroy, S.L., Korshunov, A., Lichter, P., Taylor, M.D., Pfister, S.M., 2012a. Medulloblastomics: the end of the beginning. *Nat Rev Cancer* 12, 818–834.
- Northcott, P.a., Korshunov, A., Witt, H., Hielscher, T., Eberhart, C.G., Mack, S., Bouffet, E., Clifford, S.C., Hawkins, C.E., French, P., Rutka, J.T., Pfister, S., Taylor, M.D., 2011. Medulloblastoma comprises four distinct molecular variants. *Journal of Clinical Oncology* 29, 1408–1414.
- Northcott, P.A., Nakahara, Y., Wu, X., Feuk, L., Ellison, D.W., Croul, S., Mack, S., Kongkham, P.N., Peacock, J., Dubuc, A., Ra, Y.s.S., Zilberberg, K., Mcleod, J., Scherer, S.W., Sunil, J., Sunil Rao, J., Eberhart, C.G., Grajkowska, W., Gillespie, Y., Lach, B., Grundy, R., Pollack, I.F., Hamilton, R.L., Van Meter, T., Carlotti, C.G., Boop, F., Bigner, D., Gilbertson, R.J., Rutka, J.T., Taylor, M.D., 2009. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nature Genetics* 41, 465–472. [arXiv:15334406](https://arxiv.org/abs/15334406).
- Northcott, P.A., Shih, D.J., Peacock, J., Garzia, L., Sorana Morrissy, A., Zichner, T., Stütz, A.M., Korshunov, A., Reimand, J., Schumacher, S.E., Beroukhim, R., Ellison, D.W., Marshall, C.R., Lionel, A.C., MacK, S., Dubuc, A., Yao, Y., Ramaswamy, V., Luu, B., Rolider, A., Cavalli, F.M., Wang, X., Remke, M., Wu, X., Chiu, R.Y., Chu, A., Chuah, E., Corbett, R.D., Hoad, G.R., Jackman, S.D., Li, Y., Lo, A., Mungall, K.L., Ming Nip, K., Qian, J.Q., Raymond, A.G., Thiessen, N., Varhol, R.J., Birol, I., Moore, R.A., Mungall, A.J., Holt, R., Kawachi, D., Roussel, M.F., Kool, M., Jones, D.T., Witt, H., Fernandez-L, A., Kenney, A.M., Wechsler-Reya, R.J., Dirks, P., Aviv, T., Grajkowska, W.A., Perek-Polnik, M., Haberler, C.C., Delattre, O., Reynaud, S.S., Doz, F.F., Pernet-Fattet, S.S., Cho, B.K., Kim, S.K., Wang, K.C., Scheurlen, W., Eberhart, C.G., Fèvre-Montange, M.,

- Jouvet, A., Pollack, I.F., Fan, X., Muraszko, K.M., Yancey Gillespie, G., Di Rocco, C., Massimi, L., Michiels, E.M., Kloosterhof, N.K., French, P.J., Kros, J.M., Olson, J.M., Ellenbogen, R.G., Zitterbart, K., Kren, L., Thompson, R.C., Cooper, M.K., Lach, B., McLendon, R.E., Bigner, D.D., Fontebasso, A., Albrecht, S., Jabado, N., Lindsey, J.C., Bailey, S., Gupta, N., Weiss, W.A., Bognár, L., Klekner, A., Van Meter, T.E., Kumabe, T., Tominaga, T., Elbabaa, S.K., Leonard, J.R., Rubin, J.B., Liau, L.M., Van Meir, E.G., Fouladi, M., Nakamura, H., Cinalli, G., Garami, M., Hauser, P., Saad, A.G., Iolascon, A., Jung, S., Carlotti, C.G., Vibhakar, R., Shin Ra, Y., Robinson, S., Zollo, M., Faria, C.C., Chan, J.A., Levy, M.L., Sorensen, P.H., Meyerson, M., Pomeroy, S.L., Cho, Y.J., Bader, G.D., Tabori, U., Hawkins, C.E., Bouffet, E., Scherer, S.W., Rutka, J.T., Malkin, D., Clifford, S.C., Jones, S.J., Korbelt, J.O., Pfister, S.M., Marra, M.A., Taylor, M.D., 2012b. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* 487, 49–56.
- Nowell, P.C., 1976. The clonal evolution of tumor cell populations. *Science (New York, N.Y.)* 194, 23–8.
- Nusse, R., Nusse, R., 2005. Wnt signaling in disease and in development. *Cell research* 15, 28–32.
- Nusse, R., Varmus, H.E., 1992. Wnt genes.
- Nüsslein-volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in drosophila. *Nature* 287, 795–801. [arXiv:arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- O’Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., Jenuwein, T., 2001. The Polycomb-Group Gene *Ezh2* Is Required for Early Mouse Development. *Molecular and Cellular Biology* 21, 4330–4336.
- Odame, I., Duckworth, J., Talsma, D., Beaumont, L., Furlong, W., Webber, C., Barr, R., 2006. Osteopenia, physical activity and health-related quality of life in survivors of brain tumors treated in childhood. *Pediatric blood cancer* 46, 357–62.
- Ogilvy-Stuart, A.L., Clayton, P.E., Shalet, S.M., 1994. Cranial irradiation and early puberty. *Journal of Clinical Endocrinology and Metabolism* 78, 1282–1286.
- Olshen, A.B., Venkatraman, E.S., Lucito, R., Wigler, M., 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5, 557–572.
- Packer, R.J., Gajjar, A., Vezina, G., Rorke-Adams, L., Burger, P.C., Robertson, P.L., Bayer, L., LaFond, D., Donahue, B.R., Marymont, M.A.H., Muraszko, K., Langston, J., Sposto,

- R., 2006. Phase III study of craniospinal radiation therapy followed by adjuvant chemotherapy for newly diagnosed average-risk medulloblastoma. *Journal of Clinical Oncology* 24, 4202–4208.
- Packer, R.J., Goldwein, J., Nicholson, H.S., Vezina, L.G., Allen, J.C., Ris, M.D., Muraszko, K., Rorke, L.B., Wara, W.M., Cohen, B.H., Boyett, J.M., 1999. Treatment of children with medulloblastomas with reduced-dose craniospinal radiation therapy and adjuvant chemotherapy: A Children's Cancer Group Study. *J Clin Oncol* 17, 2127–2136.
- Packer, R.J., Macdonald, T., Vezina, G., Keating, R., Santi, M., 2012. Medulloblastoma and primitive neuroectodermal tumors. *Handb Clin Neurol* 105, 529–548.
- Packer, R.J., Vezina, G., 2008. Management of and prognosis with medulloblastoma: Therapy at a crossroads.
- Palmer, S.L., Hassall, T., Evankovich, K., Mabbott, D.J., Bonner, M., Deluca, C., Cohn, R., Fisher, M.J., Morris, E.B., Broniscer, A., Gajjar, A., 2010. Neurocognitive outcome 12 months following cerebellar mutism syndrome in pediatric patients with medulloblastoma. *Neuro-oncology* 12, 1311–7.
- Paraf, F., Jothy, S., Van Meir, E.G., 1997. Brain tumor-polyposis syndrome: Two genetic diseases?
- Parkes, J., Hendricks, M., Ssenyonga, P., Mugamba, J., Molyneux, E., Schouten-van Meeteren, A., Qaddoumi, I., Fieggen, G., Luna-Fineman, S., Howard, S., Mitra, D., Bouffet, E., Davidson, A., Bailey, S., 2015. SIOP PODC adapted treatment recommendations for standard-risk medulloblastoma in low and middle income settings. *Pediatric Blood and Cancer* 62, 553–564. [arXiv:NIHMS150003](https://arxiv.org/abs/1503.0003).
- Patel, J.N., Mandock, K., McLeod, H.L., 2014. Clinically relevant cancer biomarkers and pharmacogenetic assays. *Journal of Oncology Pharmacy Practice* 20, 65–72.
- Paulino, A.C., 2002. Hypothyroidism in children with medulloblastoma: A comparison of 3600 and 2340 cGy craniospinal radiotherapy. *International Journal of Radiation Oncology Biology Physics* 53, 543–547.
- Peate, I., 2011. Men and cancer: The gender dimension. *British Journal of Nursing* 20, 340–343.
- Pei, Y., Moore, C.E., Wang, J., Tewari, A.K., Eroshkin, A., Cho, Y.J., Witt, H., Korshunov, A., Read, T.A., Sun, J.L., Schmitt, E.M., Miller, C.R., Buckley, A.F., McLendon, R.E., Westbrook, T.F., Northcott, P.A., Taylor, M.D., Pfister, S.M., Febbo, P.G., Wechsler-Reya, R.J., 2012. An Animal Model of MYC-Driven Medulloblastoma. *Cancer Cell* 21,

155–167.

- Pfister, S., Schlaeger, C., Mendrzyk, F., Wittmann, A., Benner, A., Kulozik, A., Scheurlen, W., Radlwimmer, B., Lichter, P., 2007. Array-based profiling of reference-independent methylation status (aPRIMES) identifies frequent promoter methylation and consecutive downregulation of *ZIC2* in pediatric medulloblastoma. *Nucleic Acids Research* 35.
- Pfister, S.M., Korshunov, A., Kool, M., Hasselblatt, M., Eberhart, C., Taylor, M.D., 2010. Molecular diagnostics of CNS embryonal tumors.
- Phoenix, T.N., Patmore, D.M., Boop, S., Boulos, N., Jacus, M.O., Patel, Y.T., Roussel, M.F., Finkelstein, D., Goumnerova, L., Perreault, S., Wadhwa, E., Cho, Y.J., Stewart, C.F., Gilbertson, R.J., 2016. Medulloblastoma Genotype Dictates Blood Brain Barrier Phenotype. *Cancer Cell* 29, 508–522.
- Pickles, J.C., Hawkins, C., Pietsch, T., Jacques, T.S., 2018. CNS embryonal tumours: WHO 2016 and beyond. *Neuropathology and Applied Neurobiology* 44, 151–162.
- Pierson, J., Hostager, B., Fan, R., Vibhakar, R., 2008. Regulation of cyclin dependent kinase 6 by microRNA 124 in medulloblastoma. *Journal of neuro-oncology* 90, 1–7.
- Pietsch, T., Haberler, C., 2016. Update on the integrated histopathological and genetic classification of medulloblastoma – a practical diagnostic guideline. *Clinical Neuropathology* 35, 344–352.
- Pietsch, T., Schmidt, R., Remke, M., Korshunov, A., Hovestadt, V., Jones, D.T., Felsberg, J., Kaulich, K., Goschzik, T., Kool, M., Northcott, P.A., Von Hoff, K., Von Bueren, A.O., Friedrich, C., Mynarek, M., Skladny, H., Fleischhack, G., Taylor, M.D., Cremer, F., Lichter, P., Faldum, A., Reifenberger, G., Rutkowski, S., Pfister, S.M., 2014. Prognostic significance of clinical, histopathological, and molecular characteristics of medulloblastomas in the prospective HIT2000 multicenter clinical trial cohort. *Acta Neuropathologica* 128, 137–149.
- Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W., Albertson, D.G., 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20, 207–211.
- Pizer, B.L., Clifford, S.C., 2009. The potential impact of tumour biology on improved clinical practice for medulloblastoma: progress towards biologically driven clinical trials. *British journal of neurosurgery* 23, 364–375.

- Powers, D.M.W., 2011. EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS CORRELATION. *Journal of Machine Learning Technology* .
- Pruitt, K., Zinn, R.L., Ohm, J.E., McGarvey, K.M., Kang, S.H.L., Watkins, D.N., Herman, J.G., Baylin, S.B., 2006. Inhibition of SIRT1 reactivates silenced cancer genes without loss of promoter DNA hypermethylation. *PLoS Genetics* 2, 0344–0352.
- Qian, B.Z., Pollard, J.W., 2010. Macrophage Diversity Enhances Tumor Progression and Metastasis. *Cell* 141, 39–51.
- Ramaswamy, V., Remke, M., Adamski, J., Bartels, U., Tabori, U., Wang, X., Huang, A., Hawkins, C., Mabbott, D., Laperriere, N., Taylor, M.D., Bouffet, E., 2016a. Medulloblastoma subgroup-specific outcomes in irradiated children: Who are the true high-risk patients? *Neuro-Oncology* 18, 291–297.
- Ramaswamy, V., Remke, M., Bouffet, E., Bailey, S., Clifford, S.C., Doz, F., Kool, M., Dufour, C., Vassal, G., Milde, T., Witt, O., von Hoff, K., Pietsch, T., Northcott, P.A., Gajjar, A., Robinson, G.W., Padovani, L., André, N., Massimino, M., Pizer, B., Packer, R., Rutkowski, S., Pfister, S.M., Taylor, M.D., Pomeroy, S.L., 2016b. Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. *Acta Neuropathologica* 131, 821–831.
- Ramaswamy, V., Remke, M., Bouffet, E., Faria, C.C., Perreault, S., Cho, Y.J., Shih, D.J., Luu, B., Dubuc, A.M., Northcott, P.A., Schüller, U., Gururangan, S., McLendon, R., Bigner, D., Fouladi, M., Ligon, K.L., Pomeroy, S.L., Dunn, S., Triscott, J., Jabado, N., Fontebasso, A., Jones, D.T.W., Kool, M., Karajannis, M.A., Gardner, S.L., Zagzag, D., Nunes, S., Pimentel, J., Mora, J., Lipp, E., Walter, A.W., Ryzhova, M., Zheludkova, O., Kumirova, E., Alshami, J., Croul, S.E., Rutka, J.T., Hawkins, C., Tabori, U., Codispoti, E.T., Packer, R.J., Pfister, S.M., Korshunov, A., Taylor, M.D., 2013. Recurrence patterns across medulloblastoma subgroups: an integrated clinical and molecular analysis. *Lancet Oncol* 14, 1200–1207.
- Ramaswamy, V., Taylor, M.D., 2017. Medulloblastoma: From myth to molecular. *Journal of Clinical Oncology* 35, 2355–2363.
- Ransohoff, K.J., Sarin, K.Y., Tang, J.Y., 2015. Smoothed inhibitors in sonic hedgehog subgroup medulloblastoma.
- Rao, R.C., Dou, Y., 2015. Hijacked in cancer: The KMT2 (MLL) family of methyltransferases. [arXiv:15334406](https://arxiv.org/abs/15334406).

- Rebholz, C.E., Reulen, R.C., Toogood, A.a., Frobisher, C., Lancashire, E.R., Winter, D.L., Kuehni, C.E., Hawkins, M.M., 2011. Health care use of long-term survivors of childhood cancer: the British Childhood Cancer Survivor Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 29, 4181–8.
- Relton, C.L., Davey Smith, G., 2010. Epigenetic Epidemiology of Common Complex Disease: Prospects for Prediction, Prevention, and Treatment. *PLoS Medicine* 7, e1000356. [arXiv:e1000356](https://arxiv.org/abs/e1000356).
- Remke, M., Hielscher, T., 2011a. Adult medulloblastoma comprises three major molecular variants. *Journal of Clinical ...* 29, 2717–2723.
- Remke, M., Hielscher, T., 2011b. FSTL5 is a marker of poor prognosis in non-WNT/non-SHH medulloblastoma. *Journal of Clinical ...* 29.
- Richon, V.M., Sandhoff, T.W., Rifkind, R.A., Marks, P.A., 2000. Histone deacetylase inhibitor selectively induces p21WAF1 expression and gene-associated histone acetylation. *Proceedings of the National Academy of Sciences* 97, 10014–10019.
- Robbins, D.J., Fei, D.L., Riobo, N.A., 2012. The hedgehog signal transduction network.
- Robertson, P.L., Muraszko, K.M., Holmes, E.J., Sposto, R., Packer, R.J., Gajjar, A., Dias, M.S., Allen, J.C., 2006. Incidence and severity of postoperative cerebellar mutism syndrome in children with medulloblastoma: a prospective study by the Children's Oncology Group. *Journal of Neurosurgery: Pediatrics* 105, 444–451.
- Robinson, G., Parker, M., Kranenburg, T., Lu, C., 2012. Novel mutations target distinct subgroups of medulloblastoma. *Nature* 488, 43–48.
- Robinson, G.W., Kaste, S.C., Chemaitilly, W., Bowers, D.C., Laughton, S., Smith, A., Gotardo, N., Partap, S., Bendel, A., Wright, K.D., Orr, B.A., Warner, W.C., Onar-Thomas, A., Gajjar, A., 2017. Irreversible growth plate fusions in children with medulloblastoma treated with a targeted hedgehog pathway inhibitor. *Oncotarget* 8, 69295–69302.
- Robinson, G.W., Orr, B.A., Wu, G., Gururangan, S., Lin, T., Qaddoumi, I., Packer, R.J., Goldman, S., Prados, M.D., Desjardins, A., Chintagumpala, M., Takebe, N., Kaste, S.C., Rusch, M., Allen, S.J., Onar-Thomas, A., Stewart, C.F., Fouladi, M., Boyett, J.M., Gilbertson, R.J., Curran, T., Ellison, D.W., Gajjar, A., 2015. Vismodegib exerts targeted efficacy against recurrent sonic hedgehog - Subgroup medulloblastoma: Results from phase II Pediatric Brain Tumor Consortium studies PBTC-025B and PBTC-032. *Journal of Clinical Oncology* 33, 2646–2654.

- Robinson, G.W., Rudneva, V.A., Buchhalter, I., Billups, C.A., Waszak, S.M., Smith, K.S., Bowers, D.C., Bendel, A., Fisher, P.G., Partap, S., Crawford, J.R., Hassall, T., Indelicato, D.J., Boop, F., Klimo, P., Sabin, N.D., Patay, Z., Merchant, T.E., Stewart, C.F., Orr, B.A., Korbel, J.O., Jones, D.T., Sharma, T., Lichter, P., Kool, M., Korshunov, A., Pfister, S.M., Gilbertson, R.J., Sanders, R.P., Onar-Thomas, A., Ellison, D.W., Gajjar, A., Northcott, P.A., 2018. Risk-adapted therapy for young children with medulloblastoma (SJYC07): therapeutic and molecular outcomes from a multicentre, phase 2 trial. *The Lancet Oncology* 19, 768–784.
- Rose, S.R., Danish, R.K., Kearney, N.S., Schreiber, R.E., Lustig, R.H., Burghen, G.A., Hudson, M.M., 2005. ACTH deficiency in childhood cancer survivors. *Pediatric Blood and Cancer* 45, 808–813.
- Roussel, M.F., Stripay, J.L., 2018. Epigenetic Drivers in Pediatric Medulloblastoma. *Cerebellum* 17, 28–36.
- Rubinfeld, B., Souza, B., Albert, I., Müller, O., Chamberlain, S.H., Masiarz, F.R., Munemitsu, S., Polakis, P., 1993. Association of the APC gene product with β -catenin. *Science* 262, 1731–1734.
- Rutkowski, S., Von Hoff, K., Emser, A., Zwiener, I., Pietsch, T., Figarella-Branger, D., Giangaspero, F., Ellison, D.W., Garre, M.L., Biassoni, V., Grundy, R.G., Finlay, J.L., Dhall, G., Raquin, M.A., Grill, J., 2010. Survival and prognostic factors of early childhood medulloblastoma: An international meta-analysis. *Journal of Clinical Oncology* 28, 4961–4968.
- Salaroli, R., Ronchi, A., Buttarelli, F.R., Cortesi, F., Marchese, V., Della Bella, E., Renna, C., Baldi, C., Giangaspero, F., Cenacchi, G., 2015. Wnt activation affects proliferation, invasiveness and radiosensitivity in medulloblastoma. *Journal of Neuro-Oncology* 121, 119–127.
- Salk, J.J., Fox, E.J., Loeb, L.A., 2010. Mutational Heterogeneity in Human Cancers: Origin and Consequences. *Annual Review of Pathology* 5, 51.
- Sandoval, J., Heyn, H.A., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., Esteller, M., 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* .
- Santen, G.W.E., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M., Kant, S.G., Snoeck, I.N., Peeters, E.A.J., Hilhorst-Hofstee, Y., Wessels, M.W., Den Hollander, N.S., Ruivenkamp, C.A.L., Van Ommen, G.J.B., Breuning, M.H., Den Dunnen, J.T., Van Haeringen, A., Kriek, M., 2012. Mutations in SWI/SNF chromatin remodeling complex

- gene ARID1B cause Coffin-Siris syndrome. *Nature Genetics* 44, 379–380.
- Schneider, R., Grosschedl, R., 2007. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev* 21, 3027–3043.
- Schroeder, K., Gururangan, S., 2014. Molecular variants and mutations in medulloblastoma. *Pharmacogenomics and personalized medicine* 7, 43–51.
- Schüller, U., Heine, V.M., Mao, J., Kho, A.T., Dillon, A.K., Han, Y.G., Huillard, E., Sun, T., Ligon, A.H., Qian, Y., Ma, Q., Alvarez-Buylla, A., McMahon, A.P., Rowitch, D.H., Ligon, K.L., 2008. Acquisition of Granule Neuron Precursor Identity Is a Critical Determinant of Progenitor Cell Competence to Form Shh-Induced Medulloblastoma. *Cancer Cell* 14, 123–134.
- Schwalbe, E.C., Hicks, D., Rafiee, G., Bashton, M., Gohlke, H., Enshaei, A., Potluri, S., Matthiesen, J., Mather, M., Taleongpong, P., Chaston, R., Silmon, A., Curtis, A., Lindsey, J.C., Crosier, S., Smith, A.J., Goschzik, T., Doz, F., Rutkowski, S., Lannering, B., Pietsch, T., Bailey, S., Williamson, D., Clifford, S.C., 2017a. Minimal methylation classifier (MIMIC): A novel method for derivation and rapid diagnostic detection of disease-associated DNA methylation signatures. *Scientific Reports* 7, 1–8.
- Schwalbe, E.C., Lindsey, J.C., Nakjang, S., Crosier, S., Smith, A.J., Hicks, D., Rafiee, G., Hill, R.M., Iliasova, A., Stone, T., Pizer, B., Michalski, A., Joshi, A., Wharton, S.B., Jacques, T.S., Bailey, S., Williamson, D., Clifford, S.C., 2017b. Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study. *The Lancet Oncology* 18, 958–971.
- Schwalbe, E.C., Williamson, D., Lindsey, J.C., Hamilton, D., Ryan, S.L., Megahed, H., Garami, M., Hauser, P., Dembowska-Baginska, B., Perek, D., Northcott, P.A., Taylor, M.D., Taylor, R.E., Ellison, D.W., Bailey, S., Clifford, S.C., 2013. DNA methylation profiling of medulloblastoma allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. *Acta Neuropathologica* 125, 359–371.
- Seisenberger, S., Peat, J.R., Hore, T.A., Santos, F., Dean, W., Reik, W., 2012. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20110330–20110330.
- Sengupta, S., Pomeranz Krummel, D., Pomeroy, S., 2017. The evolution of medulloblastoma therapy to personalized medicine. *F1000Research* 6, 490.

- Shahi, M., Afzal, M., Sinha, S., Eberhart, C., Rey, J., Fan, X., Castresana, J., 2011. Human hedgehog interacting protein expression and promoter methylation in medulloblastoma cell lines and primary tumor samples. *Journal of Neuro-Oncology* 103.
- Shahi, M.H., Afzal, M., Sinha, S., Eberhart, C.G., Rey, J.A., Fan, X., Castresana, J.S., 2010. Regulation of sonic hedgehog-GLI1 downstream target genes PTCH1, Cyclin D2, Plakoglobin, PAX6 and NKX2.2 and their epigenetic status in medulloblastoma and astrocytoma. *BMC Cancer* 10.
- Shannon-Lowe, C., Rickinson, A.B., Bell, A.I., 2017. Epstein-Barr virus-associated lymphomas. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 372, 20160271.
- Sharma, P., Allison, J.P., 2015. The future of immune checkpoint therapy. *Science* 348, 56–61.
- Sharma, S., Kelly, T.K., Jones, P.A., 2009. Epigenetics in cancer. *Carcinogenesis* 31, 27–36. [arXiv:PMC3410582](https://arxiv.org/abs/PMC3410582).
- Sharma, T., Schwalbe, E.C., Williamson, D., Sill, M., Hovestadt, V., Mynarek, M., Rutkowski, S., Robinson, G.W., Gajjar, A., Cavalli, F., Ramaswamy, V., Taylor, M.D., Lindsey, J.C., Hill, R.M., Jäger, N., Korshunov, A., Hicks, D., Bailey, S., Kool, M., Chavez, L., Northcott, P.A., Pfister, S.M., Clifford, S.C., 2019. Second-generation molecular subgrouping of medulloblastoma: an international meta-analysis of Group 3 and Group 4 subtypes. *Acta Neuropathologica* 138, 309–326.
- Sharp, A.J., Locke, D.P., Mcgrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R., Oseroff, V.V., Albertson, D.G., Pinkel, D., Eichler, E.E., 2005. Segmental Duplications and Copy-Number Variation in the Human Genome , 78–88.
- Shih, D.J., Northcott, P.A., Remke, M., Korshunov, A., Ramaswamy, V., Kool, M., Luu, B., Yao, Y., Wang, X., Dubuc, A.M., Garzia, L., Peacock, J., Mack, S.C., Wu, X., Rolider, A., Morrissy, A.S., Cavalli, F.M., Jones, D.T., Zitterbart, K., Faria, C.C., Schüller, U., Kren, L., Kumabe, T., Tominaga, T., Ra, Y.S., Garami, M., Hauser, P., Chan, J.A., Robinson, S., Bognár, L., Klekner, A., Saad, A.G., Liau, L.M., Albrecht, S., Fontebasso, A., Cinalli, G., De Antonellis, P., Zollo, M., Cooper, M.K., Thompson, R.C., Bailey, S., Lindsey, J.C., Di Rocco, C., Massimi, L., Michiels, E.M., Scherer, S.W., Phillips, J.J., Gupta, N., Fan, X., Muraszko, K.M., Vibhakar, R., Eberhart, C.G., Fouladi, M., Lach, B., Jung, S., Wechsler-Reya, R.J., Fèvre-Montange, M., Jouvét, A., Jabado, N., Pollack, I.F., Weiss, W.A., Lee, J.Y., Cho, B.K., Kim, S.K., Wang, K.C., Leonard, J.R., Rubin, J.B., De

- Torres, C., Lavarino, C., Mora, J., Cho, Y.J., Tabori, U., Olson, J.M., Gajjar, A., Packer, R.J., Rutkowski, S., Pomeroy, S.L., French, P.J., Kloosterhof, N.K., Kros, J.M., Van Meir, E.G., Clifford, S.C., Bourdeaut, F., Delattre, O., Doz, F.F., Hawkins, C.E., Malkin, D., Grajkowska, W.A., Perek-Polnik, M., Bouffet, E., Rutka, J.T., Pfister, S.M., Taylor, M.D., 2014. Cytogenetic prognostication within medulloblastoma subgroups. *Journal of Clinical Oncology* 32, 886–896.
- Shlien, A., Malkin, D., 2009. Copy number variations and cancer. [arXiv:1512.00567](https://arxiv.org/abs/1512.00567).
- Sigal, A., Rotter, V., 2000. Oncogenic Mutations of the p53 Tumor Suppressor: The Demons of the Guardian of the Genome. *CANCER RESEARCH* 60, 6788–6793.
- Simon, N., Friedman, J., Hastie, T., Tibshirani, R., 2011. Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software, Articles* 39, 1–13.
- Singh, S., Hawkins, C., Clarke, I., Squire, J., 2004. Identification of human brain tumour initiating cells. *Nature* 432, 396–401.
- Sporn, M.B., 1996. The war on cancer. *The Lancet Oncology* 347, 1377–81.
- Stanton, B.Z., Hodges, C., Calarco, J.P., Braun, S.M., Ku, W.L., Kadoch, C., Zhao, K., Crabtree, G.R., 2017. Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. *Nature Genetics* 49, 282–288. [arXiv:15334406](https://arxiv.org/abs/15334406).
- Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* .
- Stratton, M.R., Campbell, P.J., Futreal, P.A., 2009. The cancer genome. *Nature* 458, 719–724.
- Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.A., Jones, D.T., Konermann, C., Pfaff, E., Tönjes, M., Sill, M., Bender, S., Kool, M., Zapatka, M., Becker, N., Zucknick, M., Hielscher, T., Liu, X.Y., Fontebasso, A.M., Ryzhova, M., Albrecht, S., Jacob, K., Wolter, M., Ebinger, M., Schuhmann, M.U., van Meter, T., Frühwald, M.C., Hauch, H., Pekrun, A., Radlwimmer, B., Niehues, T., Von Komorowski, G., Dürken, M., Kulozik, A.E., Madden, J., Donson, A., Foreman, N.K., Drissi, R., Fouladi, M., Scheurlen, W., von Deimling, A., Monoranu, C., Roggendorf, W., Herold-Mende, C., Unterberg, A., Kramm, C.M., Felsberg, J., Hartmann, C., Wiestler, B., Wick, W., Milde, T., Witt, O., Lindroth, A.M., Schwartzentruber, J., Faury, D., Fleming, A., Zakrzewska, M., Liberski, P.P., Zakrzewski, K., Hauser, P., Garami, M., Klekner, A., Bogнар, L., Morrissy, S., Cavalli, F., Taylor, M.D., van Sluis, P., Koster, J., Versteeg, R., Volckmann, R., Mikkelsen, T., Aldape, K.,

- Reifenberger, G., Collins, V.P., Majewski, J., Korshunov, A., Lichter, P., Plass, C., Jabado, N., Pfister, S.M., 2012. Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell* .
- Sudarov, A., Joyner, A.L., 2007. Cerebellum morphogenesis: The foliation pattern is orchestrated by multi-cellular anchoring centers. *Neural Development* 2. [arXiv:NIHMS150003](https://arxiv.org/abs/150003).
- Swartling, F.J., Grimmer, M.R., Hackett, C.S., Northcott, P.A., Fan, Q.W., Goldenberg, D.D., Lau, J., Masic, S., Nguyen, K., Yakovenko, S., Zhe, X.N., Flynn Gilmer, H.C., Collins, R., Nagaoka, M., Phillips, J.J., Jenkins, R.B., Tihan, T., Vandenberg, S.R., James, C.D., Tanaka, K., Taylor, M.D., Weiss, W.A., Chesler, L., 2010. Pleiotropic role for MYCN in medulloblastoma. *Genes and Development* 24, 1059–1072.
- Swartling, F.J., Savov, V., Persson, A.I., Chen, J., Hackett, C.S., Northcott, P.A., Grimmer, M.R., Lau, J., Chesler, L., Perry, A., Phillips, J.J., Taylor, M.D., Weiss, W.A., 2012. Distinct Neural Stem Cell Populations Give Rise to Disparate Brain Tumors in Response to N-MYC. *Cancer Cell* 21, 601–613.
- Szczepny, A., Rogers, S., Jayasekara, W.S.N., Park, K., McCloy, R.A., Cochrane, C.R., Ganju, V., Cooper, W.A., Sage, J., Peacock, C.D., Cain, J.E., Burgess, A., Watkins, D.N., 2017. The role of canonical and non-canonical Hedgehog signaling in tumor progression in a mouse model of small cell lung cancer. *Oncogene* 36, 5544–5550.
- Sze, C.C., Shilatifard, A., 2016. MLL3/MLL4/COMPASS family on epigenetic regulation of enhancer function and cancer. *Cold Spring Harbor Perspectives in Medicine* 6.
- Tamagawa, Y., Ishikawa, K., Ishikawa, K., Ishida, T., Kitamura, K., Makino, S., Tsuru, T., Ichimura, K., 2002. Phenotype of DFNA11: A Nonsyndromic Hearing Loss Caused by a Myosin VIIA Mutation. *The Laryngoscope* 112, 292–297.
- Tamayo, P., Cho, Y.J., Tsherniak, A., Greulich, H., Ambrogio, L., Meeteren, N.S.V., Zhou, T., Buxton, A., Kool, M., Meyerson, M., Pomeroy, S.L., Mesirov, J.P., 2011. Predicting relapse in patients with medulloblastoma by integrating evidence from clinical and genomic features. *Journal of Clinical Oncology* 29, 1415–1423.
- Taylor, M.D., Northcott, P.A., Korshunov, A., Remke, M., Cho, Y.J., Clifford, S.C., Eberhart, C.G., Parsons, D.W., Rutkowski, S., Gajjar, A., Ellison, D.W., Lichter, P., Gilbertson, R.J., Pomeroy, S.L., Kool, M., Pfister, S.M., 2012. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathologica* 123, 465–472.
- Thompson, E.M., Hielscher, T., Bouffet, E., Remke, M., Luu, B., Gururangan, S., McLendon, R.E., Bigner, D.D., Lipp, E.S., Perreault, S., Cho, Y.J., Grant, G., Kim, S.K., Lee,

- J.Y., Rao, A.A., Giannini, C., Li, K.K.W., Ng, H.K., Yao, Y., Kumabe, T., Tominaga, T., Grajkowska, W.A., Perek-Polnik, M., Low, D.C., Seow, W.T., Chang, K.T., Mora, J., Pollack, I.F., Hamilton, R.L., Leary, S., Moore, A.S., Ingram, W.J., Hallahan, A.R., Jouvét, A., Fèvre-Montange, M., Vasiljevic, A., Faure-Contier, C., Shofuda, T., Kagawa, N., Hashimoto, N., Jabado, N., Weil, A.G., Gayden, T., Wataya, T., Shalaby, T., Grotzer, M., Zitterbart, K., Sterba, J., Kren, L., Hortobágyi, T., Klekner, A., László, B., Pócza, T., Hauser, P., Schüller, U., Jung, S., Jang, W.Y., French, P.J., Kros, J.M., van Veelen, M.L.C., Massimi, L., Leonard, J.R., Rubin, J.B., Vibhakkar, R., Chambless, L.B., Cooper, M.K., Thompson, R.C., Faria, C.C., Carvalho, A., Nunes, S., Pimentel, J., Fan, X., Muraszko, K.M., López-Aguilar, E., Lyden, D., Garzia, L., Shih, D.J., Kijima, N., Schneider, C., Adamski, J., Northcott, P.A., Kool, M., Jones, D.T., Chan, J.A., Nikolic, A., Garre, M.L., Van Meir, E.G., Osuka, S., Olson, J.J., Jahangiri, A., Castro, B.A., Gupta, N., Weiss, W.A., Moxon-Emre, I., Mabbott, D.J., Lassaletta, A., Hawkins, C.E., Tabori, U., Drake, J., Kulkarni, A., Dirks, P., Rutka, J.T., Korshunov, A., Pfister, S.M., Packer, R.J., Ramaswamy, V., Taylor, M.D., 2016. Prognostic value of medulloblastoma extent of resection after accounting for molecular subgroup: a retrospective integrated clinical and molecular analysis. *The Lancet Oncology* 17, 484–495. [arXiv:15334406](https://arxiv.org/abs/15334406).
- Thompson, M., Fuller, C., 2006. Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *Journal of Clinical ...* 24.
- Tsao, M.S., Sakurada, A., Cutz, J.C., Zhu, C.Q., Kamel-Reid, S., Squire, J., Lorimer, I., Zhang, T., Liu, N., Daneshmand, M., Marrano, P., da Cunha Santos, G., Lagarde, A., Richardson, F., Seymour, L., Whitehead, M., Ding, K., Pater, J., Shepherd, F.A., 2005. Erlotinib in lung cancer - molecular and clinical predictors of outcome. *New England Journal of Medicine* 353, 133–144.
- Tsurusaki, Y., Okamoto, N., Ohashi, H., Kosho, T., Imai, Y., Hibi-Ko, Y., Kaname, T., Naritomi, K., Kawame, H., Wakui, K., Fukushima, Y., Homma, T., Kato, M., Hiraki, Y., Yamagata, T., Yano, S., Mizuno, S., Sakazume, S., Ishii, T., Nagai, T., Shiina, M., Ogata, K., Ohta, T., Niikawa, N., Miyatake, S., Okada, I., Mizuguchi, T., Doi, H., Saitsu, H., Miyake, N., Matsumoto, N., 2012. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nature Genetics* 44, 376–378.
- Vaillant, C., Monard, D., 2009. SHH pathway and cerebellar development.
- Venkatraman, E.S., Olshen, A.B., 2007. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 23, 657–663.

- Vihinen, M., 2012. How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC genomics* 13 Suppl 4, S2.
- Vo, B.H.T., Li, C., Morgan, M.A., Theurillat, I., Finkelstein, D., Wright, S., Hyle, J., Smith, S.M., Fan, Y., Wang, Y.D., Wu, G., Orr, B.A., Northcott, P.A., Shilatifard, A., Sherr, C.J., Roussel, M.F., 2017. Inactivation of Ezh2 Upregulates Gfi1 and Drives Aggressive Myc-Driven Group 3 Medulloblastoma. *Cell Reports* 18, 2907–2917.
- Von Bueren, A.O., Von Hoff, K., Pietsch, T., Gerber, N.U., Warmuth-Metz, M., Deinlein, F., Zwiener, I., Faldum, A., Fleischhack, G., Benesch, M., Krauss, J., Kuehl, J., Kortmann, R.D., Rutkowski, S., 2011. Treatment of young children with localized medulloblastoma by chemotherapy alone: Results of the prospective, multicenter trial HIT 2000 confirming the prognostic impact of histology. *Neuro-Oncology* 13, 669–679.
- Wagner, J.A., 2002. Overview of biomarkers and surrogate endpoints in drug development.
- Wang, J., Wen, S., Fraser Symmans, W., Pusztai, L., Coombes, K.R., 2009a. The bimodality index: A criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. *Cancer Informatics* 7, 199–216.
- Wang, T., Pei, X., Zhan, J., Hu, J., Yu, Y., Zhang, H., 2012. FERM-containing protein FRMD5 is a p120-catenin interacting protein that regulates tumor progression. *FEBS letters* 586, 3044–3050.
- Wang, Z., Gerstein, M., Snyder, M., 2009b. RNA-Seq: a revolutionary tool for transcriptomics 10, 57–63.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., Schübeler, D., 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics* 37, 853–862.
- Weil, D., Kussel, P., Blanchard, S., Levy, G., Levi-Acobas, F., Drira, M., Ayadi, H., Petit, C., 1997. The autosomal recessive isolated deafness, DFNB2, and the Usher 1B syndrome are allelic defects of the myosin-VIIA gene. *Nature genetics* 16, 191–193.
- Williams, M.E., Wilke, S.A., Daggett, A., Davis, E., Otto, S., Ravi, D., Ripley, B., Bushong, E.A., Ellisman, M.H., Klein, G., Ghosh, A., 2011. Cadherin-9 regulates synapse-specific differentiation in the developing hippocampus. *Neuron* 71, 640–655.
- Wu, J.I., 2012. Diverse functions of ATP-dependent chromatin remodeling complexes in development and cancer.
- Xue, W., Kitzing, T., Roessler, S., Zuber, J., Krasnitz, A., Schultz, N., Reville, K., Weissmueller, S., Rappaport, A.R., Simon, J., Zhang, J., Luo, W., Hicks, J., Zender, L., Wang,

- X.W., Powers, S., Wigler, M., Lowe, S.W., 2012. A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions. *Proceedings of the National Academy of Sciences* 109, 8212–8217.
- Yam, P.T., Charron, F., 2013. Signaling mechanisms of non-conventional axon guidance cues: The Shh, BMP and Wnt morphogens.
- Yang, Z.J., Ellis, T., Markant, S.L., Read, T.A., Kessler, J.D., Bourbonoulas, M., Schüller, U., Machold, R., Fishell, G., Rowitch, D.H., Wainwright, B.J., Wechsler-Reya, R.J., 2008. Medulloblastoma Can Be Initiated by Deletion of Patched in Lineage-Restricted Progenitors or Stem Cells. *Cancer Cell* 14, 135–145.
- Yi, J., Wu, J., 2018. Epigenetic regulation in medulloblastoma. *Molecular and Cellular Neuroscience* 87, 65–76.
- Yock, T.I., Yeap, B.Y., Ebb, D.H., Weyman, E., Eaton, B.R., Sherry, N.A., Jones, R.M., MacDonald, S.M., Pulsifer, M.B., Lavally, B., Abrams, A.N., Huang, M.S., Marcus, K.J., Tarbell, N.J., 2016. Long-term toxic effects of proton radiotherapy for paediatric medulloblastoma: a phase 2 single-arm study. *The Lancet. Oncology* 17, 287–98.
- Yokota, J., 2000. Tumor progression and metastasis. *Carcinogenesis* 21, 497–503.
- Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhang, C.Z., Wala, J., Mermel, C.H., Sougnez, C., Gabriel, S.B., Hernandez, B., Shen, H., Laird, P.W., Getz, G., Meyerson, M., Beroukhi, R., 2013. Pan-cancer patterns of somatic copy number alteration. *Nature Genetics* 45, 1134–1140. [arXiv:15334406](https://arxiv.org/abs/15334406).
- Zeltzer, P.M., Boyett, J.M., Finlay, J.L., Albright, A.L., Rorke, L.B., Milstein, J.M., Allen, J.C., Stevens, K.R., Stanley, P., Li, H., Wisoff, J.H., Geyer, J.R., McGuire-Cullen, P., Stehbens, J.A., Shurin, S.B., Packer, R.J., 1999. Metastasis stage, adjuvant treatment, and residual tumor are prognostic factors for medulloblastoma in children: Conclusions from the Children's Cancer Group 921 randomized phase III study. *Journal of Clinical Oncology* 17, 832–845. [arXiv:arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Zhang, H., Zhang, H., Zhang, Y., Ng, S.S., Ren, F., Wang, Y., Duan, Y., Chen, L., Zhai, Y., Guo, Q., Chang, Z., 2010. Dishevelled-DEP domain interacting protein (DDIP) inhibits Wnt signaling by promoting TCF4 degradation and disrupting the TCF4/ β -catenin complex. *Cellular Signalling* 22, 1753–1760.
- Zhukova, N., Ramaswamy, V., Remke, M., Pfaff, E., Shih, D.J., Martin, D.C., Castelo-Branco, P., Baskin, B., Ray, P.N., Bouffet, E., Von Bueren, A.O., Jones, D.T., Northcott,

P.A., Kool, M., Sturm, D., Pugh, T.J., Pomeroy, S.L., Cho, Y.J., Pietsch, T., Gessi, M., Rutkowski, S., Bogner, L., Klekner, A., Cho, B.K., Kim, S.K., Wang, K.C., Eberhart, C.G., Fevre-Montange, M., Fouladi, M., French, P.J., Kros, M., Grajkowska, W.A., Gupta, N., Weiss, W.A., Hauser, P., Jabado, N., Jouvett, A., Jung, S., Kumabe, T., Lach, B., Leonard, J.R., Rubin, J.B., Liao, L.M., Massimi, L., Pollack, I.F., Ra, Y.S., Van Meir, E.G., Zitterbart, K., Schüller, U., Hill, R.M., Lindsey, J.C., Schwalbe, E.C., Bailey, S., Ellison, D.W., Hawkins, C., Malkin, D., Clifford, S.C., Korshunov, A., Pfister, S., Taylor, M.D., Tabori, U., 2013. Subgroup-specific prognostic implications of TP53 mutation in medulloblastoma. *Journal of Clinical Oncology* 31, 2927–2935.