# An evaluation of cancer subtypes and glioma stem cell characterisation

Unifying tumour transcriptomic features with cell line expression and chromatin accessibility

**Ewan Roderick Johnstone**

EMBL-EBI, Darwin College

University of Cambridge

This dissertation is submitted for the degree of

*Doctor of Philosophy*

Darwin College

December 2016

Dedicated to Klaudyna.

# Declaration

- I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.
- This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.
- This dissertation is typeset in LaTeX using one-and-a-half spacing, contains fewer than 60,000 words including appendices, footnotes, tables and equations and has fewer than 150 figures.

Ewan Roderick Johnstone
December 2016

# Acknowledgements

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC, Ref:1112564) and supported by the European Molecular Biology Laboratory (EMBL) and its outstation, the European Bioinformatics Institute (EBI).

I have many people to thank for assistance in preparing this thesis. First and foremost I must thank my supervisor, Paul Bertone for his support and willingness to take me on as a student. My thanks are also extended to present and past members of the Bertone group, particularly Pär Engström and Remco Loos who have provided a great deal of guidance over the course of my studentship. I must also thank the members of my thesis advisory committee, Steve Pollard, John Marioni and Jan Korbel for their support and comments. Last but not least I thank my wife to be, Klaudyna Johnstone nèe Schmidt, for your significant patience and support.

A number of people contributed to the analysis set out in this thesis. Remco Loos contributed with comments and notes work presented in Chapter 2. Harry Bulstrode contributed cell culture and ATAC-seq library preparation for ATAC-seq datasets used in Chapter 4. Steven Pollard provided advice on glioma and GNS biology as well as supervision for H. Bulstrode. Paul Bertone contributed supervision, funding, manuscript revision alongside cell culture and microarray data preparation for GNS and NS exon array data represented in Chapter 3.

# Abstract

The observation of significant cellular heterogeneity between tumours of the same type has inspired efforts to identify gene expression based signatures representative of this variation. Large sample size expression datasets have been used to describe discrete clusters of tumour samples that are intended to represent functionally and clinically divergent tumour subtypes. While many of these studies have identified reproducible subtypes, the definition of clear expression-based subtypes in glioma have been particularly elusive. Relating these tumour subtypes to expression signatures and phenotypes in cancer stem cells has also been difficult.

Set out within this thesis I apply a novel coexpression analysis method to identify independent subtypes within independent coexpression modules. These modules relate to intuitive biological features enabling module specific variation to be identified independently of a transcriptome wide subtype. This methodology is used to evaluate established subtypes in breast ductal carcinoma and glioma. In breast carcinoma the basal, luminal, Her2-enriched and claudin-low cancer subtypes are replicated revealing functional expression differences that define each type. In glioma, dominant expression variation presents a grade associated axis of proneural to mesenchymal expression. This axis is also present within individual tumours suggesting classification of individual tumours as discrete subtypes should not be assumed. Analysis of glioma derived stem cell lines similarly reveals distinct proneural and mesenchymal clusters in both gene expression and chromatin accessibility. These signatures also unify phenotypes described in previous glioma stem cell analysis. Proneural signature genes suggest these lines are similar to normal glial progenitor cells while mesenchymal expression largely relates inflammatory and immune responses.

Differential chromatin accessibility of signatures genes enables the analysis of epigenetic control of subtype signature transcriptional networks. Complementing subtype analysis I also compare between glioma derived stem cells and neural stem cells to identify glioma specific features. Novel methods for ATAC-seq analysis are also described for the examination of chromatin accessibility. These findings will further assist the translation of tumour subtypes to the clinic alongside deeper characterisation glioma's persistent and heterogeneous cancer stem cell population.

# Table of contents

# List of figures

# List of tables

# Nomenclature

**Acronyms / Abbreviations**

| | |
|---|---|
| ATAC-seq | Assay for transposase-accessible chromatin using sequencing |
| A, T, C and G | DNA nucleotides adenine, thymine, cytosine and guanine |
| BRCA | Breast invasive ductal carcinoma |
| CHC | Consensus hierarchical clustering |
| CMC | Correlation marker clustering |
| DNase-seq | DNase I hypersensitive sites followed by sequencing |
| FAIRE-seq | Formaldehyde-Assisted Isolation of Regulatory Elements sequencing |
| GDAC | Broad Institute genome data analysis center |
| GSC | Glioma stem cell |
| GNS | Glioma neural stem cell |
| KIRC | Kidney renal clear cell carcinoma |
| LUAD | Lung adenocarcinoma |
| MBC | Metaplastic breast carcinoma |
| MNase-seq | Micrococcal nuclease chromatin digestion followed by sequencing |
| MNE | Module normalised expression |
| NS | Neural stem cell |
| NMF | Nonnegative matrix factorization |

| | |
|---|---|
| OV | Ovarian serous cystadenocarcinoma |
| PCR | Polymerase chain reaction |
| PNN | Perineuronal net |
| TBS | Transposase binding site |
| TCGA | The cancer genome atlas |
| TFAP | Transcription factor accessibility profile |
| WGCNA | Weighted gene correlation network analysis |

# Chapter 1

# Introduction

## 1.1 Characterising glioma: tumours and stem cells

Central nervous system residing glioma tumours are named after their apparent similarity to the brain's normal glial cells. Gliomas are responsible for over 80% of malignant tumours and 55% of which are classed as the highest grade (glioma grade IV, glioblastoma multiforme) [80]. Low grade gliomas have a 5-year survival rate of approximately 59% [238] while patient survival to 5-years for high grade glioblastoma is a very poor 5% [190]. Despite the low overall occurrence of glioma, the impact of diagnosis is devastating for patients as glioma remains essentially incurable. The infiltrative nature of glioma cells and malignant progression of low grade tumours to high grade tumours causes significant damage in the brain leading to horrific cognitive deficits. Progress in understanding cancer development and progression has led to significant improvements in patient survival time over the last few decades as knowledge of cancer specific mechanisms and weaknesses are further exploited to target the disease. The application of genomics approaches to tumour characterisation have supplemented the traditional tumour histological grade as valuable clinical information. One example of this, brought with ubiquitous DNA sequencing, is a catalog of genomic aberrations that are thought to drive neoplastic cell proliferation that have been thoroughly investigated by projects like the cancer genome atlas (TCGA) [170] and the COSMIC database [63]. While these genome focused efforts made great strides in describing candidate transformative mutations, they did little to explain the phenotypic and histological differences between and within tumour types. In parallel, methods for profiling gene expression have revealed the transcriptomic diversity found in tumour samples. In an attempt to reduce the complexity of within tumour type gene expression variation, it was proposed that individual tumours could be classified into discrete subtypes. It's hoped that these subtypes could be accurately related back to other features like tumour grade, differentiated cell type, survival and response to therapy. While some of these cancer subtype studies have been a distinct success, a lack of consensus in other cancer types, including glioma, has become apparent. Resolution of transcriptional diversity and the status of cancer subtypes would further assist in the search for neoplastic related mechanisms, treatments and prognostic information in both glioma and other cancer types.

### Glioma stem cells

A subset of cells from glioma tumours capable of expanding in culture and recapitulating the tumours *in vivo* have been described as glioma stem cells for their similarity to other normal stem cell populations sharing multipotent differentiation potential and expression of stem cell markers like *SOX2* and *NES*. These glioma stem cells do however harbour extensive genetic abnormalities and are able to recapitulate tumours *in vivo*. In order to prevent tumour recurrence, suppression of this cellular population is key, initiating a great deal of effort to characterise the functional properties of these cells, like response to therapeutics and tumourigenic capacity etc. Attempts to translate tumour derived subtype signatures to glioma stem cell expression data have been hindered by the many differences in cellular content, microenvironment and degree of differentiation between *in vitro* culture and tumour tissue. Although the status of glioma subtypes themselves lack consensus, some studies have identified subtype-like expression in glioma stem cell cultures, however the relationship between tumour and cell line signatures is not clear. Resolution of glioma subtype-like expression may assist in the characterisation of glioma stem cell lines and enable self-renewing explant cultures to represent the diversity of neoplastic cells found in glioma tumours. Exploitation of these well characterised glioma stem cell cultures may help screen for drugs that target particular phenotypes or identify the critical and potential therapeutically targetable features found in all glioma cells.

### Glioma epigenetics

While expression studies provide profiles that present a somewhat representative silhouette of their phenotype, they do not comprehensively explain how gene expression itself is controlled. Many factors that may modulate transcription such as DNA methylation, histone modification and chromatin structure are united under the term *epigenetics*. The field of epigenetics is growing fast with many novel methods in development to study different aspects of transcriptional control. An important epigenetic mechanism is to regulate the physical accessibility of the coding DNA sequence either blocking transcription through nucleosome dense heterochromatin or through enabling access to gene promotors and other regulatory features at open euchromatin. A number of methods have been described that exploit high throughput sequencing to identify regions of open chromatin including DNase-seq and MNase-seq. A recently described method called ATAC-seq (Assay for transposase-accessible chromatin using sequencing) has become a popular alternative to many long-standing methods, however many of the methods and unique biases that are associated with this new method have not been thoroughly explored. Application of ATAC-seq to characterise glioma stem cells, coupled with gene expression analysis, may also further the clinical and research potential of these lines.

## 1.2 An overview glioma cellular diversity

One of the defining characteristics of glioma tumours is the diversity of cell types that compose them. Glioma cells can present the morphology of normal glial cells with oligodendrogliomas presenting oligodendrocyte like morphology and astrocytomas presenting astrocyte like morphology. Tumours presenting both oligodendrocyte and astrocyte like cells as well as a significant population of anaplastic cells are known as oligoastrocytomas reflecting their mixed lineage potential. Astrocytomas are characterised by cells with morphological similarities to astrocytes and the trauma responsive reactive astrocyte phenotype. Histological markers for astrocytomas and astrocytes alike include GFAP, YKL-40 (*CHI3L1*) and ApoE [223]. Oligodendrogliomas present markers for oligodendrocyte precursor cells but not mature oligodendrocytes including Olig2, Ng2 (*CSPG4*), and PDGFR$\alpha$ [221, 236]. Brain tumours are also classified by grade with more aggressive forms of glioma being assigned a higher grade. The highest grade of glioma, glioblastoma multiforme grade IV (GBM) presents cells with an immature glial morphology and mostly astrocytic differentiation [158]. In concert with traditional classification processes, other methods of examining the diversity found in these tumours have exploited molecular biology techniques to identify key features that differentiate some tumours from others. Genomic aberrations like 1p/19q deletions and loss of function IDH1/2 mutations have played a significant role in delineating tumours into different groups [53, 217]. Similarly gene expression and epigenetic data has been exploited to further classify and characterise this intratumoural variation [30, 198, 270].

The origins of this cellular heterogeneity are the focus of a great amount of research questioning whether this diversity reflects the cells in which the originating oncogenic transformations occurred or if the tumour was initiated in a more stem like progenitor cell that is biased towards differentiation into one lineage in subsequent divisions. The lineage potential of glioma cells mirrors the glial potential of the brain's resident stem cells with various glial progenitors possessing variable linage potential including some glial progenitor cells that are capable of bipotent differentiation into both astrocytes and oligodendrocytes [210, 214] as well as more restricted progenitors like oligodendrocyte progenitor cells [182]. These progenitor cells persist throughout adulthood along with neural stem cells (NS) implicating them as potential cells of glioma origin [31, 227]. Some glial progenitor cells are able to differentiate into alternative lineages if induced suggesting significant plasticity in fate [31, 125]. While glioma cells may capable of differentiating into a diverse range of cell types, infiltration of glioma cells throughout the central nervous system intermixes cancer cells with non-neoplastic cells like neurons, microglia and tissues like vasculature adding to the complexity of cell content found between tumours. Non-neoplastic glia can be recruited to the tumour and induced to proliferate in a paracrine stimulated manner [4] and lineage tracing experiments suggest that recruited non-neoplastic cells can gather mutations and be transformed into another aberrant constituent of the tumour [61].

The location at which the initial neoplastic transformation of a cell occurs is difficult to define due to the infiltrative nature of these cells. Gliomas typically occur in the cerebral hemispheres and many of these proximal to the subventricular zone which is host to a resident population of adult NS cells [14]. Other central nervous system tumours are also regionally enriched potentially based on variation in progenitor population or microenvironmental signalling [76, 114]. The morphology of tumours arising in different brain regions suggest that tumours either respond to environmental queues to define their lineage specification or these cells adopt the potential lineage of the resident stem cell population before transformation.

Insights into the initiation of glioma can be found by examining the incidence of glioma in different age groups. The majority of gliomas occur in adults with the relative risk increasing with age. tumours that arise in younger patients are usually of low grade and have features that distinguish them from tumours that tend to occur in older patients. Secondary GBMs that progress from a low to high grade tumour are also more likely to occur in younger patients. One of the most predictive prognostic features is age of diagnosis, which may suggest that tumours that emerge early in life are distinct [138]. One explanation for this is that the functional potential of progenitor cells changes with age. Aged neural progenitor cells were less likely to enter mitosis than their young equivalents but those that did would divide more often [248]. Another study found that neural progenitors tended towards astroglial differentiation compared to neural and oligodendrocyte lineages with increased age [84]. Other factors like growth factor responsiveness [56] and tumour suppressor expression [176] suggest the process of ageing has a unique role in the ontogenesis of glioma.

### 1.2.1 Cancer stem cells

When studying cancer the most notable cells in the tumour microenvironment are those that are capable of sustaining and recapitulating the tumour. Tissues are formed by a defined cellular hierarchy with multipotent stem cells giving rise to transient progenitor cells and eventually terminally differentiated cells. This theoretical framework of a cellular hierarchy can be transposed from normal neural development with a neural stem (NS) cell at its root, to a model of tumour development with a oncogenically transformed stem cell at its source and differentiated cancer cell progeny defining the tumour type (e.g. astrocytoma). The cells at the root of the tumour hierarchy are often described as cancer stem cells in reference to normal stem cell populations. The term 'cancer stem cell' has been used to associate these cells with various functional properties and origins. Concepts that are related and often conflated with cancer stem cells include the concept of a cancer initiating cell and the cancer propagating cell.

The concept of the cancer initiating cell describes the cell in which the initial neoplastic transformation takes place, for example a NS or glial progenitor cell is a good candidate for this in glioma. The distinction between cancer stem and cancer propagating cell is that a cancer stem cell must be able to propagate tumours with differentiated progeny representative of the parental tumour [136]. It is not fully understood whether glioma cells differentiate via intermediate progenitor states or directly from a more deeply rooted NS like cell but it is likely that the relative proportions of cells in different states of differentiation compose a significant proportion of the intertumoural heterogeneity observed in glioma. A fundamental property of the cancer stem cell is the ability to self-renew and prevent its own terminal differentiation. The ability of some glioma cells to dedifferentiate back into a cancer stem like cell suggests that glioma cells may have greater plasticity than their normal neural counterparts [65]. This ability to self-renew has enabled the *in vitro* culture of glioma stem cells using techniques developed for neural stem cells making them a valuable platform for exploring glioma biology [5, 42, 85]. Cells capable of forming neurospheres or adherent growth are also able to recapitulate tumours *in vivo* [106, 203, 205].

Great efforts have been expended to characterise markers that identify glioma stem cells (GSC) in attempts to isolate and quantify this subpopulation. Ideally GSC markers target stem cell exclusive features like self renewal to ensure the specificity of the marker. Many of the GSC markers in use were originally identified as markers of NS cells like *SOX2*, *NANOG*, *OLIG2*, *MYC* and *NES* [9, 96, 122, 150, 266]. Further to these transcriptional regulators, a number of cell surface markers, amenable to cell sorting methods were identified as GSC markers including CD133, CD44 and L1CAM [6, 96, 155]. The most commonly used cell surface marker for GSCs is CD133 however variation in presentation of CD133 between different GSC like populations and expression differences suggest it may not be an ideal GSC marker [7, 157]. Functional variation in GSC biology suggests that there may not be an ideal marker to identify all GSCs and instead subtypes of GSCs may be better identified by a panel of markers derived from *in vitro* cultured cells. As with many stem cell like populations, the regulation of GCSs is enacted through multiple intrinsic factors including genetic, epigenetic and metabolic control as well as extrinsic factors like microenvironmental factors, cell signalling and the immune component.

Through analysis of the common mutations that define the glioma cell we can infer causative mutational events and peek at the emergent aberrant cellular state. A comprehensive survey of these mutations by the cancer genome atlas revealed that deregulation of the RB, p53, RTK/RAS/PI3K pathways are critical components of the glioma genome [170]. Further to this 40% of GBMs were found to have mutations in chromatin modifying genes [19]. Some other recurrent mutations in glioma include the receptors *EGFR* and *PDGFRA* alongside *IDH1*, *HDM2*, *PIK3CA* and the established tumour suppressors *PTEN*, *TP53*, *CDKN2A*, *NF1* and *RB1* [19, 170]. Some mutations are associated with a particular tumour expression subtype with *NF1* mutations being associated with the mesenchymal subtype [270] and *IDH1* mutations linked to a proneural G-CIMP hypermethylation subtype (Glioma-CpG island methylator phenotype) [185]. Gliomas have also been shown to develop structural rearrangements through chromothripsis and chromoplexy mechanisms that form neochromosome-like amplifications at enriched genomic breakpoints commonly amplifying *CDK4* and *MDM2* [293]. The genetic composition within each tumour may also show significant variation. One study found that tumours can be separated into mono or polygenomic tumours composed of cells that could be expanded in spheroid culture and recapitulate tumours *in vivo* [247]. The monogenomic tumours were considered psudodiploid due to their approximately DNA content, whilst still harbouring landmark GBM mutations. The polygenomic tumours contained multiple highly aneuploid tumour clones and produced more aggressive xenotransplanted tumours.

The GSC state is maintained through epigenetic mechanisms via transcriptional regulation and chromatin organisation. Transcription factors that play key roles in the maintenance and specification of GSCs include c-Myc [277] alongside many transcription factors that also regulate NS cells like the markers *SOX2*, *OLIG2* and *NES* [96, 150, 266]. The overexpression of *FOXG1* in GSCs compared to NS cells identified a role for this transcription factor in glioma for this neural development associated transcription factor [55] which has subsequently been linked to GSC tumourigenesis in functional studies [269]. Using a network model and ChIP-seq analysis, roles for the neurodevelopmental transcription factors POU3F2, SOX2, SALL2 and OLIG2 were found in GSCs [252]. Control of transcription was demonstrated by the Polycomb complex component EZH2 which was shown to phosphorylate STAT3 and promote tumourigenesis [121]. Prolific growth causes regions of the tumour to be isolated from the supply of nutrients like oxygen and glucose. Cancer cells have been shown to tolerate and exploit their inhospitable environment through a metabolic shift called the Warburg effect [172]. The increased production of ATP via aerobic glycolysis leads to an accumulation of lactate. One of the consequences of this aberrant metabolic state is production of reactive oxygen species which increases the risk of further mutations [186]. Differential regulation of proliferation and migration by the pentose phosphate pathway induced by these hypoxic conditions suggest that metabolic factors play an crucial role in GSC regulation [120]. In glioma the role of *IDH1* mutations, a key feature in classifying gliomas [53], in metabolic homeostasis further acts to regulate epigenetics and differentiation. Mutations in *IDH1* typically lead to the accumulation of 2-hydroxyglutarate which inhibits the DNA and histone demethylation activity of TET1 and TET2 [160]. Interestingly these *IDH1* mutations are uniquely found in the proneural subtype of glioma hinting at functional differences between the subtypes. Comprehensive evaluation of metabolic regulation would help inform on glioma biology but also assist improving culture methods to better represent conditions *in vitro*.

Glioma stem cells present and exploit developmental programs reflecting their neural origins controlled via signalling and gene regulatory pathways. These pathways include regulation through Notch, BMP, NF-$\kappa$B, Wnt, TGF$\beta$ and cell surface receptors like EGFR, PDGFR$\alpha$ and MET. The Notch pathway promotes the maintanence of GSC growth as well as preventing neural differentiation [73, 126]. Activation of the Notch pathway through diverse routes as been observed in glioma including through transcription factor activity [112], nitric oxide production and response to radiation [208]. The importance of Notch signalling in GSCs is demonstrated by reduced neurosphere growth in response to $\gamma$-secretase inhibitors [58]. GSCs resident within the perivascular niche may exploit endothelial cell Notch ligands to promote self renewal [296]. BMP signalling directs NS cells towards astrocyte differentiation [91, 246] which led to suggestions that BMP could be utilised therapeutically to force GSCs towards terminal differentiation [199]. Differentiation based therapies my be limited by GSC mechanisms to prevent terminal differentiation like the Gremlin1 mediated inhibition of BMP signalling and p21 [285]. A few studies have described the importance of the NF$\kappa$B pathway in GSCs either as a response to radiation or TNF$\alpha$ [11, 101]. The Wnt signalling pathway also has regulatory roles in GSC development. Genomic amplification of *PLAGL2* has the effect of Wnt mediated supression of differentiation in both NS and GSCs [292]. The proneural transcription factor *ASCL1*, which is highly expressed in glioma, was shown to regulate Wnt signalling via the downregulation of the negative Wnt regulator *DKK1* [220].

Genomic amplification of receptor tyrosine kinases *EGFR*, *PDGFRA* and *MET* are a frequent event in gliomas and heterogeneity of receptor amplification has been observed between cell populations within the same tumour [240, 254]. Amplification, overexpression or constitutive activation of these receptors has the effect of promoting proliferation of GSCs [69, 145]. It has also been suggested that different GSC populations are dependent on alternative receptor tyrosine kinases for proliferative signalling [69]. Cells treated with an anti-EGFR therapy reduced their dependence on EGFR promoted proliferation to overexpress MET leading downstream expression of stem related transcription factors *POU5F1*, *NANOG* and *KLF4* [116]. MET overexpression was also shown to promote the resistance to radiation in GSCs [115]. TGF-$\beta$ signalling was demonstrated to upregulate *SOX2* via *SOX4* promoting GSC stemmness [107]. Moreover, TGF-$\beta$ inhibition was revealed to selectively target a CD44 high population of GSCs implying heterogeneity in GSC regulation and populations [3]. Variation in *EGFR* status was shown to influence a switch between infiltrative and angiogenic phenotypes [255]. Infiltrative cells typically showed high level amplification of *EGFR* alongside higher levels of activated pEGFR. Inhibition of EGFR in infiltrative cells reduced their ability to invade surrounding tissue and instead activate an angiogenic program. This switch to an angiogenic phenotype was accompanied by a selective pressure for the EGFRvIII variant. These results demonstrate how *in vivo* tumour evolution and heterogeneity can lead to phenotypic intratumoural variation.

Mechanisms by which cancer cells evade and suppress the immune mediated clearance of aberrant cells are a critical hallmark of tumour development. The central nervous system has unique immune surveillance mechanisms that interacts with the developing glioma tumour to modulate disease progression [212]. GSC secreted TGF-$\beta$ promotes an immunosuppressive environment, with macrophages induced to tumour supportive M2 type cells [283]. Other GSC derived factors that promote M2 macrophage recruitment include periostin (*POSTN*) and integrins [295].

### 1.2.2 Inter and intratumoural expression variation

Whilst significant attention is paid to glioma by studying GSCs both *in vitro* and *in vivo*, comparable efforts have been made to investigate glioma through analysis of tumour samples. Traditional methods of tumour classification like tumour grade and cellular differentiation have long been used to provide prognostic and therapeutic information to the clinician and patient alike. Relatively recent developments improving the high throughput interrogation of gene expression have enabled the production of large sample size tumour expression datasets. Early studies utilising microarray based platforms have now been extended or replaced by RNA sequencing (RNA-seq) based projects and often with associated other data types pliable for more integrated analysis such as miRNA expression, DNA methylation and genomic mutation analysis. Projects like the cancer genome atlas have generated a large amount of tumour analysis data and perhaps more importantly made this data available online for the scientific community at large [257]. These datasets are produced with the intent to provide a survey of the variation between different tumour samples.

The variance of expression in tumour samples is often expected to identify discrete clusters of highly similar tumour samples that can be distinguished from other tumour samples. Identification of these clusters is described under the umbrella term cancer subtype analysis however, it is often seen as a generic clustering problem tractable with established machine learning methods. As such a number of methods have been applied to the problem of identifying and validating discrete clusters of samples in tumour gene expression data [22, 86, 93, 156, 177, 184, 224, 234, 258]. When these discrete subtypes are established correlations with clinical aspects such as response to treatment and patient survival are investigated with the goal of making these subtype signatures clinically relevant. Classification of samples on clinical time schedules could then be used as a prognostic tool.

This subtype methodology has been applied to a large number of different tumour types with varying degrees of success. One of the earliest and perhaps the best example of cancer subtypes to date is breast ductal carcinoma (BRCA) [194, 195]. These early studies established the “intrinsic” BRCA subtypes as well as establishing that “measurements of gene expression based on total mRNA isolated from such a complex tissue can be interpreted in terms of the properties of specific cells (e.g., the carcinoma cells)" [194]. The dominant nature of BRCA-subtype like expression variation emerging from a complex multicellular background led to many other teams applying the same methodologies to other cancer types including medulloblastoma [183], renal cell carcinoma [44], epithelial ovarian cancer [256] and glioma [198, 270]. As the intended purpose of cancer subtypes is to identify transcriptional features of the whole tumour that can be associated with clinical features these studies have also largely relied on the assumption that a single tumour sample can be representative of the whole tumour. For this reason few studies have investigated the degree of expression variation to be found within individual tumours, i.e. intratumoural variation, by taking multiple samples from the same tumour.

For glioma different studies have identified slightly different subtype definitions [104, 166]. Phillips *et al.* initially described three subtypes of high grade glioma termed proneural, proliferative and mesenchymal [198]. Subsequently using a larger dataset Verhaak *et al.* identified four subtypes of high grade glioma, two of which were termed proneural and mesenchymal due to their similarity to the Phillips *et al.* equivalent subtypes [270]. The remaining two Verhaak *et al.* subtypes were named neural and classical after the functional annotation of their signature genes.

Combining data from the TCGA's split glioblastoma and low grade glioma projects a robust separation of glioma tumours into 3 molecular subtypes defined by consistent genomic aberrations and epigenomic features [34, 53]. A dominant component of these subtypes is the mutational status of the IDH1 and IDH2 genes with tumours presenting nonfunctional variants of these genes typically presenting low grade features. These IDH1/2 mutant tumours are also delineated based on a codeletion of both chromosome arms 1p and 19q with tumours lacking this codeletion typically presenting mutation of TP53 [53]. The final molecular subtype is composed of tumours that present wildtype IDH1/2 gene status, of which a large proportion are classified as grade IV glioblastomas. This separation of IDH1/2 wildtype tumours has led to further efforts to characterise the diversity within this molecular type. Using an integrated co-clustering method, Ceccarelli *et al.* divided IDH1/2 wildtype tumours into classical-like, mesenchymal-like, PA-lie (Pilocytic Astrocytoma) and LGm6-GBM types [34]. Of these the PA-like subtype presented the most significant clinical differences with improved survival and lower age of onset whilst other co-clustered subtypes present a mixture of the Verhaak *et al.* subtypes with a minimal enrichment of their namesake classical and mesenchymal subtypes.

The Verhaak *et al.* subtypes and discrete cluster assumptions have become a prominant methodology in glioblastoma since publication, however the reproducibility of these subtypes have previously been called into question both statistically through reanalysis [166] and through the generation of new data relating subtypes to intratumoural variation [174, 192, 245].

Difficulties with subtype discovery methods can also be found in the classical case of BRCA. The identification of a novel BRCA subtype, the claudin-low type, was only accomplished through co-clustering expression data from murine and human mammary carcinoma tissues, independently from the other BRCA subtypes [98]. This claudin-low subtype is identified experimentally in a separate classification process to the intrinsic BRCA subtypes suggesting some subtype-like features may be independent and distinct from variation representing other subtypes. The noted variation in cellular composition of tumour samples has influenced the study of tumour transcriptomes in different ways. tumour sample selection for RNA profiling may consider the cellular composition and reject those that display high levels of confounding variables like necrosis, or cellular infiltration. These efforts will however only exclude the most effected samples leaving variable residual levels of of these cells and processes confounding downstream expression analysis. Similarly variations in other, less prominent cell types like cancer associated fibroblasts, endothelial vascular derived cells and non-neoplastic bystander cells will vary from sample to sample. Other factors like rate of cell division may relate to a genuine tumour subtype but may however be expected to vary in an intratumoural fashion and thus be confounded by sampling error.

Some informatics methods have been applied to study these potentially critical sources of variation. Various forms of coexpression clustering has been applied to large sample size datasets in the past [50, 54, 134, 141, 175, 181, 215, 225]. These methods are typically used to infer larger interconnected networks consisting of thousands of genes. When applied to cancer expression data, these coexpression methods frequently identify coexpression networks relating to proliferation or immune cell processes suggesting expression analysis can identify variation in immune cell frequency and rate of cell division between tumour samples. While coexpression methods are are used during the cancer subtype discovery process, it is typically leveraged to characterise the functional significance of the subtype clusters rather than to help define them. Further integration of coexpression methods into the subtype discovery process may help deconvolute tumour expression signatures and enable the identification of more complex independent subtypes.

## 1.3 Aims for this thesis

The aims of this thesis are to identify common features and sources of both transcriptomic and epigenetic variation between gliomas, other tumour types, and glioma derived cell lines. The common components of expression variation shared between glioma and other tumour types will be explored in the context of cancer subtypes by a novel coexpression clustering method applicable to a broad range of cancers and datasets. Extending the glioma tumour analysis, subtype expression profiles in glioma derived stem cell lines (GNS lines) will be explored and compared to other glioma stem cell datasets. I will also compare GNS cells to karyotypically normal NS cells further characterising the differences between the brain's resident stem cells and their neoplastic counterparts. Novel analysis and methods are applied to a recently described method, ATAC-seq, for identifying accessible chromatin may be exploited to interrogate epigenetic mechanisms relating to GNS and NS functional variation.

# Chapter 2

# Methods

## 2.1 Coexpression methods for cancer subtype analysis

### Tumour data collection and classification

Normalised TCGA RNA-seq datasets were acquired from the Broad Institutes GDAC [21]. For the glioma analysis the LGG and GBM datasets were combined using the consensus genes and quantile normalised. BRCA samples were PAM50 classified using the “genefu” package in R and the “pam50.robust centroids". Claudin-low samples were identified by clustering the samples using claudin-low predictor genes [98], 1 - the Pearson’s correlation and Ward’s linkage (Figure 7.6). Glioma samples were classified using the Verhaak GBM subtype centroids and associated classification method [270] alongside classification into glioma molecular subtypes using IDH1/2 gene and 19/19q codeletion status also generated by the Broad Institute’s GDAC [21].

### Correlation Marker Clustering

Correlation marker clustering was performed on the 10,000 most variable genes from each dataset using one minus the Pearson’s correlation coefficient as the distance metric and the centroid (UPGMC) linkage method. As the centroid linkage method can produce inversions, the cluster dendrogram is cut with the added condition that all cluster heights for each feature be below the tree cut height parameter. In this way all subsequent branch merges during agglomerative clustering inherit the highest cluster merge distance found within the lower branches. The tree cut parameter for module agglomeration limits used was $h \leq 0.2$ for all tumour analysis with the exception of the immune cell to mesenchymal signature comparison in glioma where $h \leq 0.15$ was used to discriminate lower-sized and more highly correlated features (Table 7.5). Modules that consist of less than 3 genes were also omitted from further investigation as potentially uninformative aggregations. Z-score values for modules were calculated by mean centring gene expression values and dividing these by their standard deviation. The mean z-score value across all transformed module genes was used to represent that samples module expression value (Module z-score).

Coexpression of modules in different cancer types was determined by calculating the mean pairwise correlation (MPC) between genes for each cancer type. Module/cancer type pairs were considered coexpressed if the mean pairwise correlation was above 0.5 (Table 3.1).

### Correlation Marker Subclustering

Clustering within coexpression modules was achieved by converting each gene within a feature into a z or standard score and subtracting the sample mean z-score across all module genes. This matrix then represents the degree to which each gene or sample varies from the expected expression level as predicted by the module z-score. From this matrix only genes and samples with a variance greater than half the mean variance were carried forward to reduce the dimensionality of the clustering operation. This reduced matrix was clustered using the Euclidean distance metric and Ward's linkage. Module subclusters were defined by cutting the dendrogram at half the maximum branch height to form the dominant subclusters within that module. Genes that are significantly differentially expressed for each module subcluster are identified using the Welch Two-Sample $t$-test ($p < 0.001$) to compare module-normalised expression between subcluster and non-subcluster samples.

### Tumour microarray analysis methods

Intratumoural multi-sample data from Sottoriva et al. was retrieved from ArrayExpress (E-MTAB-1129). Glioma microarray expression data generated by Phillips *et al.* [198] was downloaded from ArrayExpress (E-GEOD-4271) and RMA normalised using the “Affy" library in R. The data were quantile normalised and low variance probes were removed from further investigation (Variance $\geq$ mean variance $\times$ 2). The minimum number of probes used per module was 55 (glioma Interferon/within-Mesenchymal module).

### Cluster stability measures

In order to test for stability in established subtype classifications the silhouette method from the R library "cluster" and the connectivity measure from R library "clValid" was applied with standard parameters.

### Other analysis

GO term analysis was performed using the R library “GOstats” for each coexpressed module where all genes included in the correlation matrix were used as the gene universe. Principal component analysis was completed using the "prcomp" function in R. Pearson’s correlation and tests for significance of association were calculated using the “cor.test” function in R. All analysis was performed in “R”.

## 2.2 Transcriptomic analysis of glioma derived neural stem cells

### RNA sample processing

RNA was extracted with the Trizol method (Invitrogen) followed by treatment with TURBO DNase (Ambion), further phenol/chloroform extraction and ethanol precipitation. RNA quality was assessed on the Agilent 2100 Bioanalyzer.

### Gene expression profiling

Samples were processed for microarray hybridisation according to the GeneChip whole-transcript sense target labeling assay (Affymetrix). Briefly, 2 $\mu$g of each sample was depleted of ribosomal RNA (RiboMinus, Invitrogen). Double-stranded cDNA was synthesized using random hexamers tagged with a $5'$ T7 primer, and synthesis products were amplified with T7 RNA polymerase to generate antisense cRNA. Reverse transcription was performed on the cRNA template using SuperScript III to yield ssDNA, substituting dUTPs for dTTPs, and cRNA was subsequently degraded via RNase H digestion. cDNA products were then nicked with uracil DNA glycosylase (UDG) and apurinic/apyrimidinic endonuclease 1 (APE 1) at sites of first-strand dUTP incorporation, followed by biotin labeling with terminal deoxynucleotidyl transferase (TdT). Affymetrix Exon Array 1.0 ST arrays were hybridized for 16 h at 45°C, washed, stained with streptavidin-phycoerythrin (SAPE) conjugate on a FS450 automated fluidics station, and imaged on a GCS3000 7G scanner (Affymetrix). Feature extraction was performed using Command Console 3.2.3, and hybridisation quality was assessed with Expression Console 1.1.2 (Affymetrix).

### Exon Microarray normalisation

Exon arrays were processed using v1.28 of the xps Bioconductor package. Background correction, quantile normalisation and calculation of probeset expression values from fluorescence data was performed using the Robust Multi-chip Average (RMA) method [108], and probesets were summarised by median polish in xps. Where a gene was represented by multiple splice variants, the transcript model having the maximal value was taken as the dominant isoform.

### Microarray and differential expression analysis

Differential analysis of gene expression between GNS versus NS and proneural versus mesenchymal samples was completed using the R libraries "limma" and "affy" where the eBayes function was used for differential expression. GSC microarray expression data generated by Günther *et al* [87], Bhatt *et al* [11] and Lee *et al* [143] was downloaded from ArrayExpress (E-GEOD-23806, E-GEOD-49161 and E-GEOD-4536) and RMA normalised using the “Affy" library in R. In situations where multiple GSC microarray probes match to a gene the mean of all probe values was used to represent that gene’s expression. Gene ontology enrichment analysis was completed using the "GOstats" library in R using the combined GNS and NS differential gene sets as the gene universe. CMC analysis of GNS data was accomplished using the same methodology as was used for tumour data in Chapter 1 (3000 most variable genes clustered, tree cut parameter = 0.15). Calculations of z-score values in GNS CMC modules excluded NS samples for the data mean and standard deviation calculations. Principal component analysis was completed using the "prcomp" function in R. Pearson’s correlation and tests for significance of assocciation wre calculated using the “cor.test” function in R.

### Centroid based Verhaak *et al.* subtype classifier

Centroid values for the subtypes defined by Verhaak *et al.* were matched to genes expression values in the GNS dataset. For each sample the spearman’s correlation to each subtype centroid set was calculated and the highest correlation was used to identify that samples’s subtype as described by Verhaak *et al.* [270].

### Z-score based Verhaak *et al.* subtype classifier

GNS expression data for Verhaak *et al.* [270] subtype centroid genes were extracted on a per subtype basis. Within each subtype each gene is converted into a z-score by subtracting the mean expression across all samples and dividing by the standard deviation. For each sample the average subtype marker z-score is calculated and the highest average subtype z-score was used to identify that sample's subtype.

### Identification of GNS subtype modules

Associations between these GNS modules and glioma proneural and mesenchymal CMC modules described in Chapter 2 were found by identifying GNS modules with intersecting gene sets and a positive correlation to the respective glioma module ($\geq$ 3 shared genes).

## 2.3 ATAC-seq analysis and its application to GNS and NS cells

### Cell culture and ATAC-seq library preparation

Cell culture and ATAC library preparation was undertaken by H. Bulstrode according to published protocols [24]. Cells were cultured in identical culture conditions growing as adherent cultures on laminin coated plates in media supplemented with EGF and FGF as has been described in previously published work [202]. These cells were lysed and the nuclei were pelleted then suspended in buffer and treated with the Tn5 transposase for 30 minutes. The DNA was cleaned up using MinElute columns, followed by 12 cycles of PCR amplification using NEBNext High Fidelity Master Mix and custom primers with sequencing adapter sequences as previously specified [24]. The resulting libraries were cleaned up on MinElute columns once more. These were pooled based on quality control and quantification data from Bioanalyser Tapestation and Qubit analysis, then sequenced on the Illumina HiSeq 2500 (50 bases paired end) (Methods provided by H. Bulstrode).

### ATAC-seq analysis

Paired end reads were trimmed for sequencing primers using the tool "Cutadapt" and aligned to hg19 using Bowtie2 and a maximum fragment length (-X) of 3kb. Aligned reads were filtered for duplicates and split into two 27bp intervals and adjusted 18bp into the read and 9bp beyond the read to represent the transposase binding site and 9bp replicated region using an “awk” script. Regions of the genome enriched for transposase loci were identified using F-seq [17] using both broad and narrow parameters (Narrow: -l 600 -f 28, -t 8. Broad: -l 2000, -f 28, -t 3.). Broad and narrow peaks were called separately in each library and merged using samtools merge [146] into combined broad and combined narrow peak sets and concatenated together into one final set of intervals. Loci that overlapped regions blacklisted for functional genomics analysis by the ENCODE consortium were removed from further analysis [259]. Transposase binding bias nucleotide frequency was calculated by selecting 10,000 random paired end reads from two libraries and extracting the reference sequence $+/-$ 60bp from ends of both paired ends. Nucleotide frequencies were calculated and visualised using “R". Insert size distributions were extracted using "CollectInsertSizeMetrics" from Picard tools. For the analysis of insert size bias on accessibility estimates, an “awk" script was used to filter reads into 40-100bp or 170-230bp insert size sets, converted to transposase binding site footprints and then used to calculate accessibility estimates per loci. Transposase access site counts for the final peak loci were generated using bedtools ‘intersect". Loci counts were normalised for GC bias using conditional quantile normalization [94] and the CQN offsets provided to DESeq2 as the normalisation factors for differential testing. Loci with adjusted p-values below 0.05 were called as significantly differentially accessible. PCA plots and heatmaps were generated using CQN normalised data. Data visualised as heatmaps in figures 5.6, 5.8 and 5.10 were clustered using one minus the Pearson’s correlation coefficient as the distance metric and Ward’s linkage method. Visualisation and clustering was performed in “R". Predicted nucleosome density was estimated by extracting reads with insert size of between 180-247bp, 315-473bp or 558-616bp and converting them into single, double and triple nucleosome intervals of length 150bp centred at $1/2$, $1/3$ and $1/4$ intervals within the mapped read respectively.

#### Expression comparison to ATAC-seq profile

Genes differentially expressed (adjusted p value < 0.05) between for GNS/NS and proneural/mesenchymal types were examined in ATAC-seq libraries. Transcription start sites for all transcripts associated with differential genes were extracted using the R library "biomaRt" and extended 200bp into the gene body and 1kb upstream. ATAC-seq TBS counts indicating accessibility were extracted and CQN normalised for all TSS loci and the most accessible locus for each gene was used to represent the gene’s TSS. The log2 mean CQN normalised ATAC-seq score for each condition was calculated (i.e. GNS or NS) and converted into a ratio by dividing the difference between conditions by the overall mean of that gene over the complete dataset. The regression F-test was used to check for the significance of association between microarray expression and ATAC-seq accessibility estimates.

### Transcription factor motif analysis

Enrichment of transcription factor motifs in differentially accessible chromatin was accomplished using the MEME suite (4.10.2) tool “AME" [169] using sequences extracted from all merged open chromatin regions as the control sequences. The motif databases used were “HOCOMOCO version 9", “Jaspar core 2014" and “jolma2013". Motif instances in open chromatin were found using the MEME suite tool FIMO [83]. For motif frequency counts, motif loci were merged across strands to avoid counting symmetrical motifs twice. Number of intersecting motifs with differential loci was calculated using bedtools “intersect". For TFAPs, strand information was preserved however accessibility estimates were normalised by subtracting the mean, division by the standard deviation then converted to a motif average by taking the median normalised accessibility for each bp across all libraries presented here. TFAPs were extracted using bwtools aggregate [201].

# Chapter 3

# Coexpression methods for cancer subtype analysis

## 3.1 Introduction: The cancer subtype problem

Considerable effort has been made to identify discrete classes or subtypes of tumour samples based on large sample size gene expression datasets. Once clear subtypes are established, it is hoped that these divisions relate to clinical or biological features such as cell of origin, prognosis or the response to treatment, continuing the progress towards personalized therapies. This process has been particularly successful in breast cancer where clear subtypes with genomic copy number and tumour immunohistological correlates have been shown to provide prognostic and therapeutic value [48, 207]. Despite these successes, consistent and reproducible expression derived subtypes have been difficult to define for other tumour types [104, 127, 166].

A major feature of cancer as an intrinsically complex disorder is the degree of morphological and transcriptional variation present, both between and within each tumour. Neoplastic growth in the same tissue can lead to tumours resembling different cell types [253] or presenting varying levels of differentiation. tumour samples are themselves a heterogeneous mix of cell types including immune cell infiltrates, cancer-associated fibroblasts (CAF) and vascular endothelial cells alongside the target neoplastic cells. Different populations of neoplastic cells within a tumour can display distinct differences in molecular features like genomic mutations [75, 240, 245], expression [192, 245], metastatic ability or response to therapy [167]. In both breast invasive ductal carcinoma (BRCA) and glioblastoma (GBM), samples have been separated into discrete subtypes based on consensus hierarchical clustering of gene expression data [29, 270]. For glioma and BRCA tumours the consistent definition of subtypes has been problematic with variation between studies [48, 166] but also new subtypes identified using further datasets [98, 270]. GBM was last divided into four subtypes named proneural, neural, classical and mesenchymal [270]. Subsequent studies however suggest that the proneural and mesenchymal classes are most easily reproduced across studies and that the mesenchymal class is only present in high grade, poor prognosis tumours [104]. Using an integrative approach Shen *et al.* found three subtypes that approximately match Verhaak *et al.*'s proneural, classical and mesenchymal types omitting any neural-like subtype [234]. Studies using multiple samples from a single tumour [245] and through the application of single cell transcriptomics [174, 192] have found a diversity of subtype-like intratumoural variation. Analysis of a combined low and high grade glioma data set reveals that glioma tumors can largely be split into IDH1/2 wildtype and IDH1/2 mutant types with a further subdivision of IDH1/2 samples into 1p/19q codeletion or those that typically present a TP53 mutation [34, 53]. In BRCA the definitions of subtypes between studies also vary. However, a set of consensus signatures has emerged, known as the intrinsic subtypes, consisting of luminal A/B, Her2-enriched, basal and normal-like subtypes [242]. The basal subtype was identified with early microarray experiments and tends to co-occur with the triple-negative receptor subtype [29]. After the initial classifications a new subtype, Claudin-low, emerged when co-clustering human and mouse expression data [98].

The lack of consensus between studies is a key concern for the transition of these subtypes to the clinic. The established methods used for subtype discovery include consensus hierarchical clustering (CHC) and non-negative matrix factorization (NMF). These methods are most commonly used to classify tumour samples into a few ($\approx$3-7) discrete categories of tumours to avoid overfitting the data. Whole transcriptome data representing tens of thousands of genes is commonly limited to a reduced number of highly variable genes ($\approx$1000-2000) therefore omitting a significant amount of potentially valuable information. Identification of the desirable number of clusters is also a somewhat subjective process with many differing metrics and methods for deciding on the optimum. Different studies using these and other similar methods in GBM have failed to identify consistent classes other than the proneural and mesenchymal-like signatures [166] and present differing opinions on the optimum division of samples in other tumour types [183, 207]. Similarly, multiple groups present expression-based prognostic signatures that have little overlap or are difficult to interpret biologically [166]. In this chapter I investigate the transcriptional heterogeneity and emergence of discrete subtypes in BRCA and glioma tumours through coexpression analysis.

## 3.2 Results

To identify robust subtypes I considered that independent biological features may identify feature specific subtype-like clusters. As transcriptional variation derived from biological features like cell division, niche specific differentiation or stromal cell content can be thought of as independent features, clusters of tumour samples may present variation representative of clinically or developmentally distinct subdivisions. Using the example of niche specific differentiation, neoplastic growth can present variable differentiation of cells towards a particular lineage. The level of differentiation as well as non-neoplastic lineage cell content will vary between tumour samples. Aberrant modulation of the transcriptional network that guides this differentiation may initiate tumour development or provide evolutionary advantages to the affected cells. This aberrant modulation may take the form of subtype specific overexpression of a lineage promoting transcription factor or altered expression of cell signalling components etc. As lineage content varies between tumours, aberrations or subtype specific variations within the expression network must be identified relative to the baseline of lineage-associated expression.

### 3.2.1 Correlation marker clustering: Coexpression analysis for the identification of independent subtypes

To identify subtype-like modulation of these independent transcriptional features, I devised a custom coexpression analysis methodology, correlation marker clustering (CMC), to identify independent components of transcriptional variation representative of biological signatures by aggregating sets of highly correlated genes. By clustering module expression values relative to the module mean, samples that display consistent variation from the module mean can be identified as module specific subtypes. As these coexpression modules are independent from each other, clusters of samples that present module specific variation can be identified independently in comparison to global transcriptome subtype. To describe this in different terms, samples can then be assigned to different subtypes based on module specific variation.

Coexpression modules can be derived using hierarchical clustering using one minus the Pearson's correlation as a distance metric, followed by the "cutting" of the aggregated tree at a defined height or distance. The clustered modules can then be extracted by pruning away small sized clusters as well as unclustered single genes. The choice of linkage criteria affects the size, expansion and internal consistency of coexpression modules. Two of the most effective linkage criteria for producing variably sized and highly correlated clusters are UPGMA (Unweighted pair-group method with arithmetic mean) and UPGMC (Unweighted pair-group method centroid mean) [239]. These methods can produce similar and highly overlapping coexpression modules when their clustering trees are cut at different heights. For the purposes of identifying independent subtype coexpression modules, the desirable characteristics include reliable identification of significant components of the tumour transcriptome, the association of many genes to related clusters to facilitate module subclustering and a reduced number of small and potentially less informative clusters. As UPGMC linkage is non-monotonically increasing and can produce so called *inversions*, where clusters may merge at lower distances than previous merge operations, I add the additional requirement that all cluster merges for a subcluster must be below the tree cut height. This requirement avoids the inclusion of relatively uncorrelated clusters that were previously merged at much higher distances safeguarding the consistency of the module. Comparing between UPGMA and UPGMC linkage I find that the centroid based UPGMC linkage produces fewer and larger sized clusters than UPGMA for different tree cut heights (Figure 3.1). Both linkage methods produced similar clusters after adjusting for distance differences between methods. Regardless of linkage method choice, consistent module subclusters can be identified in modules derived with both methods.

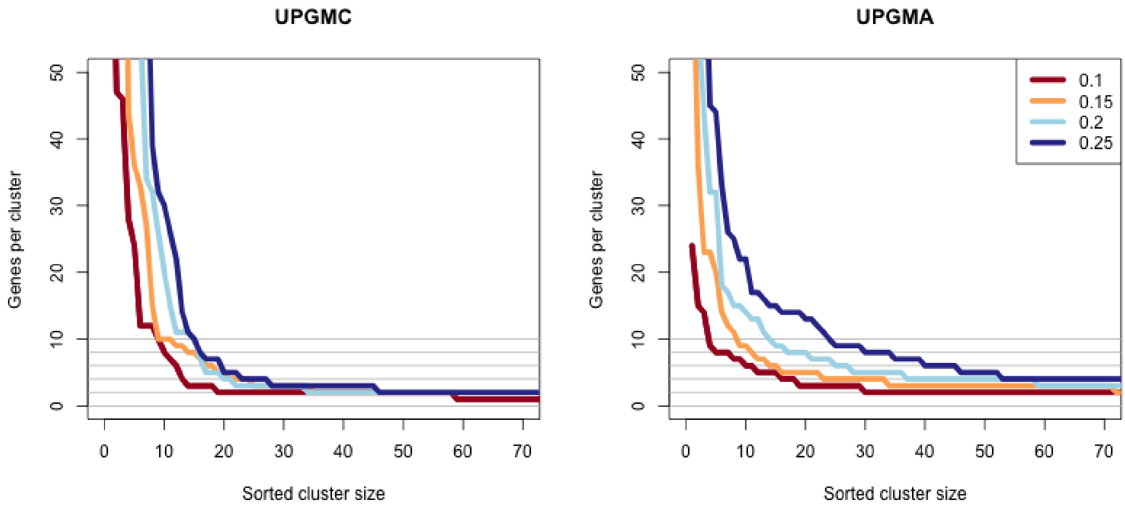


Fig. 3.1 Comparison of UPGMC (Left) and UPGMA (Right) linkage criteria in test breast cancer RNA-seqexpression data (10,000 genes by 1026 samples, Log2 FPKM). The centroid based UPGMC finds fewer and larger clusters than UPGMA. The UPGMA clustering also finds a large number of small clusters (n $\leq$ 10 genes) increasing the number of clusters that may be brought forward for intramodule clustering. The distance metric used is one minus Pearson's r.

Since this methodology demands the use of a hard threshold to limit the clustering, a range of cut off values were tested ranging from 0.1 to 0.5 and examined with a variety of metrics (Figure 3.2). As the cut off height increases the average number of genes within a module rises and the average correlation to the module z-score decreases. An ideal cut off height would include enough genes within each module to enable robust within-module clustering yet also avoid over expansion of each module to include poorly correlated genes. An arbitrary cut off height of 0.2 was selected as balance between module size (Average 27 genes) and the mean of the minimum correlation of module genes to its parent module (r = 0.85). Similarly where tumour subtypes are replicated using average linkage derived modules (Figure 7.1), a cut off height of 0.45 was used to approximate the size of UPGMC derived modules and as compromise between module size, gene to module correlation as well as total number of modules (Figures 3.1 and 7.2).

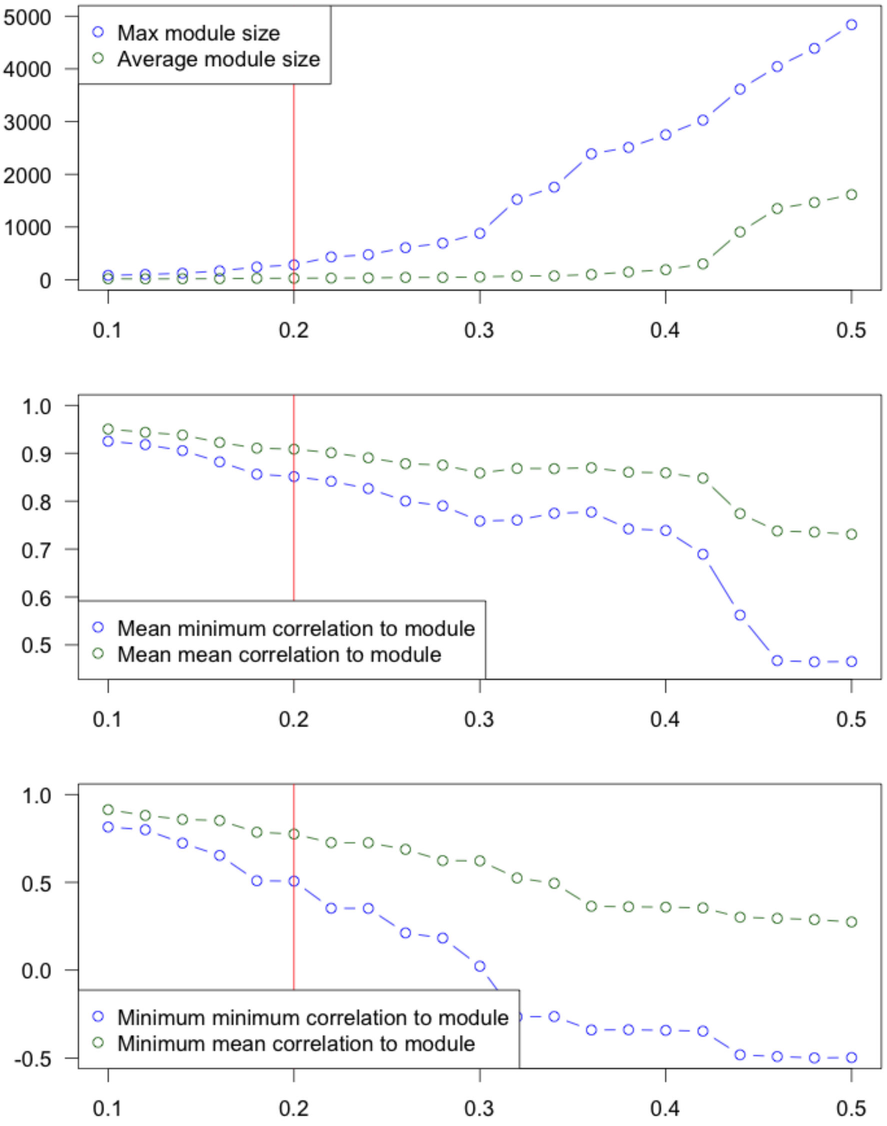


Fig. 3.2 Centroid linkage module metrics across a range of cut off heights. Coexpression modules derived using the UPGMC linkage method are examined using metrics for average and maximum number of genes per module (Top panel) as well as correlation to module z-score summarised as the mean (Middle panel) or minimum (Bottom panel) of all module metrics. The vertical red line indicates the chosen cut off height of 0.2 for UPGMC derived modules which provides a compromise between module size (average 27 genes) and within module consistency. Data and clustering were identical to that used in figure 3.1.

In order to cluster samples within modules the expression of genes were normalised as z-scores and the sample mean z-score was subtracted from the gene z-score values. After removing low variance genes and samples, this matrix of distances from the module mean enables the identification of consistent subtype like variation via hierarchical clustering using Euclidian distance and Ward's method for the linkage. I refer to this combination of coexpression module extraction via UPGMC linkage, tree cut height limit and module subclustering as *correlation marker clustering* (CMC), after the distinct components of transcriptional variation found to be representative of established biological features.

### 3.2.2 Correlation marker clustering analysis identifies both general and cancer-specific features in tumour transcriptomes

Correlation marker clustering was applied to RNA-seq data generated by The Cancer Genome Atlas [257] and preprocessed by the Broad Institute's Genome Data Analysis Center [21]. The focus was turned initially towards breast invasive ductal carcinoma (BRCA) and glioma, as these tumour types have been the subjects of substantial subtype analysis. A total of 1026 BRCA samples, and combined low-grade glioma and high-grade glioblastoma multiforme data into a multi-grade glioma dataset consisting of 468 low-grade and 166 glioblastoma samples were examined using the CMC methodology. A total of 33 expression modules in the BRCA dataset and 44 modules in the glioma dataset were identified. To test for similarities between the signatures derived for BRCA and glioma, I identified those modules that showed overlap in their gene sets. Two modules in each dataset consisted of at least 25 genes and shared more than 25% of the larger module's genes. Taking the intersecting genes between these modules as consensus features I applied gene ontology (GO) term enrichment analysis to identify potential functional significance. These two modules were enriched for GO terms relating to mitosis/cellular division and immune cell function respectively (Tables 7.1 and 7.2). In agreement with this functional association this mitosis-associated feature contains immunohistological markers of mitotic activity including *MKI67* [48, 74, 207] alongside cell cycle regulators like *PLK1*, *FOXM1*, *CCNB1* and *CDK1*. Likewise the immune function-associated module is enriched for markers of immune cells including *CD4* [226], *CD163* and *CD68* [75, 137, 240, 245, 272] alongside complement pathway components *C1QA*,*B*,*C*, chemokine receptors *CCR1*,*2*,*5* and numerous major histocompatibility complex components. Based on this functional association these modules were labeled as the consensus mitosis and immune cell modules.

Given the general nature of the processes involved, we hypothesised that the same signatures may also be present in other cancer types. Expression of genes contained in the consensus modules was investigated in a panel of three additional tumour types including kidney renal clear cell carcinoma (KIRC), ovarian serous cystadenocarcinoma (OV) and lung adenocarcinoma (LUAD) retrieved from the Broad Institute's GDAC [21]. The expression of genes in both consensus modules is highly correlated across all examined cancer types (Table 3.1). Moreover, for all cancer types, samples present a gradient of expression of these signatures (Figure 3.3). Based on this analysis I propose that these modules represent biological signatures, which are independent of cancer type, and can be thought of as transcriptional proxies for variation in proliferation and immune cell infiltration within the tumour sample.

Next I examined the BRCA and glioma features individually to determine if their genes are co-expressed in other tumour types (Table 3.1). A notable signature enriched across multiple tumour types includes a module representative of the Interferon type I response (*STAT1*, *HERC5*,*6* and *OAS1*,*2*,*3*,*L*.). Other signatures were restricted to a given tumour type, including signatures representative of established cancer subtypes. In BRCA a module that contains numerous genes attributed to the luminal subtype [29, 167, 195] is broadly co-expressed as a gradient of expression where in other cancer types these genes show no consistent relationship (Figure 3.3). Also found here are a number of glioma-specific modules including neural-like, oligodendrocyte-like and the glioma proneural subtype (Figure 3.3) [198, 270]. These tumour-specific modules refer to tumour type or niche-specific co-expression. Many of these genes or cell types have previously been associated with cancer subtypes. Alongside the identification of broadly continuous signatures, CMC establishes expression-based proxies for discrete signatures like the presence of Y chromosome gene expression in male glioma samples. Cancer-specific discrete modules like expression proxies for the *ERBB2*/Her2 amplicon in BRCA and the *CDK4* amplicon in glioma are also found using CMC (Figure 7.3).

I conclude that the transcriptome of tumour samples can be divided into independent components through correlation marker clustering. These components can be shared across tumour types or restricted to a single cancer type. Many modules can be interpreted as expression-based proxies for biological signatures or processes within each sample, which can be continuous or discrete in nature. To examine whether disentangling these signatures can lead to a more robust understanding of expression-based cancer subtypes, I investigated in detail how our modules relate to previously described subtypes in BRCA and glioma.

### 3.2.3 Breast ductal carcinoma expression modules allow precise identification of established tumour subtypes

Previous expression studies agree on the existence of the luminal and basal expression-based signatures. Other subtypes such as claudin-low, Her2-enriched and the normal type have

| Feature Origin | Module | Glioma | BRCA | KIRC | OV | LUAD |
|---|---|---|---|---|---|---|
| Consensus Mitosis | Cellular proliferation / mitosis | ✓ | ✓ | ✓ | ✓ | ✓ |
| Consensus Immune | Immune cell infiltration | ✓ | ✓ | ✓ | ✓ | ✓ |
| BRCA | AP-1 TF | ✓ | ✓ | ✓ | ✓ | ✓ |
| BRCA | Interferon type I response | ✓ | ✓ | ✓ | ✓ | ✓ |
| BRCA | BRCA Luminal feature | ✗ | ✓ | ✗ | ✗ | ✗ |
| Glioma | Glioma Proneural | ✓ | ✗ | ✗ | ✗ | ✗ |
| Glioma | Oligodendrocyte | ✓ | ✗ | ✗ | ✗ | ✗ |
| Glioma | Neural | ✓ | ✗ | ✗ | ✗ | ✗ |

Table 3.1 Table of modules discussed detailing if the module is coexpressed in a panel of five different cancer types. Modules coexpressed in all RNA-seq datasets are presented in contrast to modules enriched only in one tumour type. A module is considered to be coexpressed if the module presents high mean pairwise correlation (MPC) (mean $r \geq 0.5$).

been reported in individual studies, but have not been reproduced in others [29, 123]. Here I examine the expression modules to see if these subtypes can be identified and characterised in more detail. Through CMC analysis, I identified two modules that contain markers for the established basal or luminal subtypes, including *EGFR*, *FOXC1* and *SFRP1* for basal, and *ESR1*, *FOXA1*, *GATA3*, *SPDEF* and *XBP1* for luminal. These basal and luminal signatures are not present in other cancer types (Glioma, KIRC, OV and LUAD), indicating these modules are specific to BRCA samples (Table 3.1). These genes also have established roles in non-metaplastic basal/myoepithelial or luminal progenitor differentiation and cell fate [10, 23, 37, 113, 168, 262, 276]. To see how the expression levels of modules are related across tumour samples, I calculated a module score that represents the global expression level of the module in each sample (z-score). The basal and luminal module scores reveal two distinct clusters that replicate the distinction between basal and other luminal-like subtypes using PAM50 classification (Figure 3.4). Curiously, the basal samples did not present the highest levels of the basal module. In fact, these samples are better distinguished by low expression of the luminal module rather than high basal module expression (Figure 3.4). The Her2-enriched samples present a slightly reduced luminal signature compared to the luminal samples. Samples presenting a claudin-low like signature on average express the highest levels of the basal module alongside a broad range of luminal module expression. The claudin-low samples are distinguished from the remaining types by high expression of both the basal and an additional module composed of genes that have previously been associated with the basal type like *ZEB1* and *EGFR* (Figure 3.4b). Taken together these features broadly support the separation of the luminal, basal and claudin-low subtypes but raise further questions regarding the strict delineation, characterisation and unique transcriptional differences between BRCA subtypes.

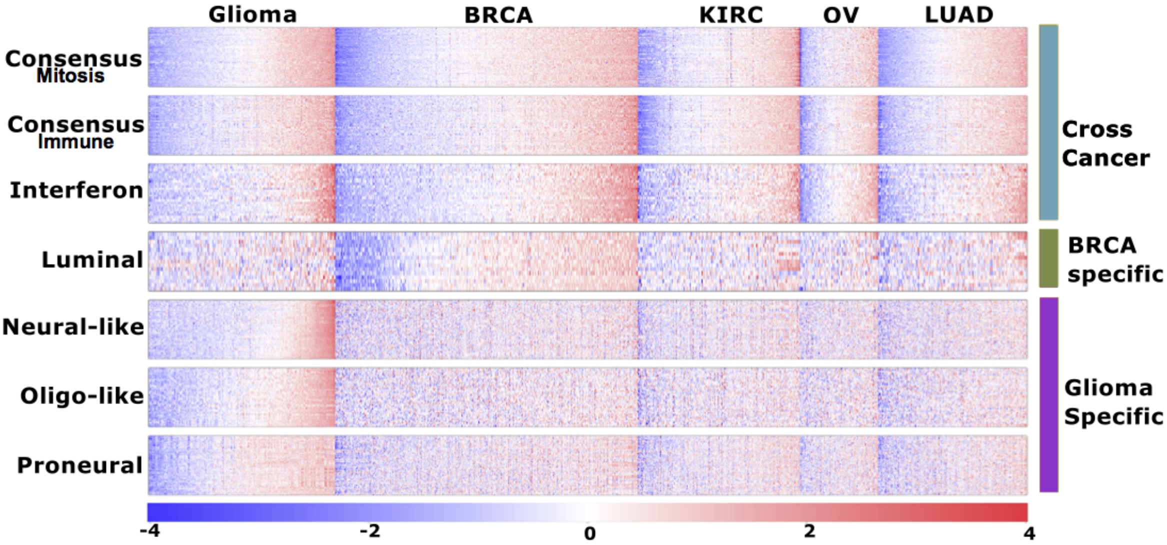
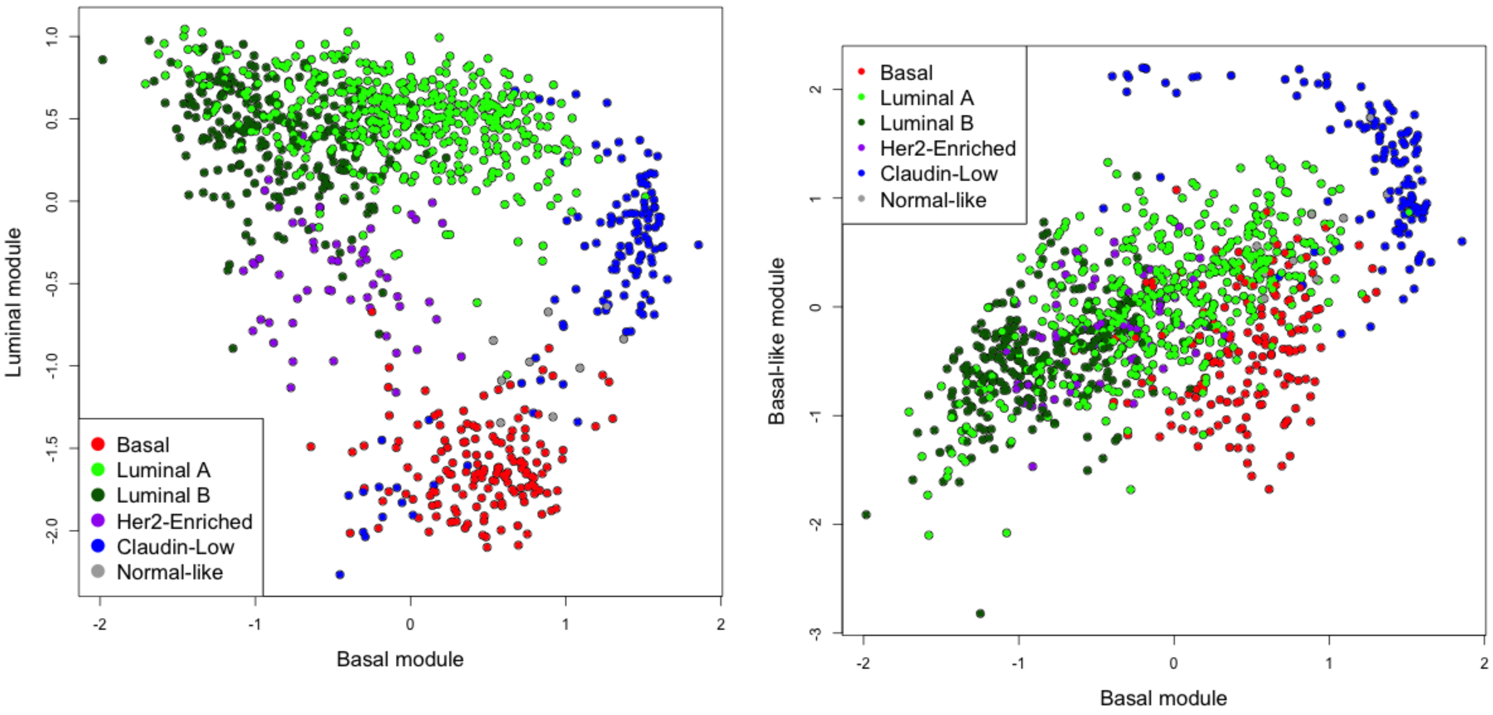


Fig. 3.3 Heatmap visualising coexpression modules derived from cross-cancer, BRCA or glioma RNA-seq tumour data presented as z-score log2 FPKM expression. The cross-cancer consensus modules are seen to be co-expressed and present a continuous gradient of expression in Glioma, BRCA, KIRC, OV, and LUAD tumour types. The luminal feature is only identified in the BRCA data. The neural-like, oligodendrocyte-like and neural-like features are only identified in the glioma data. All together these results mirror table 3.1.



(a)

(b)

Fig. 3.4 BRCA sample scatterplot of basal and luminal modules using PAM50 intrinsic subtypes or claudin-low classification (a). Basal samples (Red) show low luminal module expression. Claudin-low samples (Blue) show high basal module expression. Her2-enriched samples show considerable variation in the luminal module. High expression of the basal module and a secondary basal-like module distinguish claudin-low samples (b). Comparable features and subtypes can be defined using alternative linkage methods (Figure 7.1). Units are z-score normalised log2 FPKM.

### 3.2.4 Coexpression module subclusters define established breast cancer subtypes and reveal their distinctive transcriptional features

Untangling distinct components of transcriptional variation permits the identification of clusters within modules. Since marker modules describe independent properties of tumour samples, these clusters can be directly tied to tumour features. Technically, within-module clusters are characterised by consistent deviation from the mean module expression (or z-score) for a subset of samples. Modules are composed of genes highly correlated in expression, but within these genes, some show up- or downregulation with respect to the module as a whole for subsets of samples. These sample subsets can be identified as a functionally distinct subcluster, along with the genes that distinguish them from other subclusters. In a sense these subclusters can be described as module dependent subtypes.

I find that subclusters within multiple modules replicate the established PAM50 basal, Her2-enriched and claudin-low classifications of these samples (Figure 3.5, Table 3.2). PAM50 basal classified samples are replicated by a subcluster in the basal module, Her2-enriched as within the luminal module and claudin-low samples are replicated in subclusters within an additional stromal-like module. We define these subclusters as the basal, Her2-enriched and claudin-low subclusters respectively. Other modules are divided into subclusters that are independent of PAM50 classification that reveal module-specific differences without significantly intersecting PAM50 classification. By comparing module gene expression relative to the overall module expression the genes that distinguish subclusters can be identified (module normalised expression, MNE).

Applying this methodology to the basal subcluster identifies the genes that distinguish these samples. Specifically, the basal subcluster is most distinctly characterised by unusually high expression of *FOXC1* ($p = 4.54 \times 10^{-51}$, MNE) (Figure 3.6) [216]. It is important to note that it is not the direct expression level of these genes that distinguishes the basal subcluster. Indeed, the other cluster consisting of mostly luminal and Her2-Enriched subtypes shows a full range of expression of *FOXC1*. However, expression of *FOXC1* in these samples broadly follows the expression of the basal module, whereas the basal subcluster presents higher levels of expression than would be expected for a gene in the module. Similarly, the basal subcluster was found to overexpress other established basal markers including *SFRP1* ($p = 1.68 \times 10^{-24}$, MNE), *GABRP* ($p = 1.85 \times 10^{-41}$, MNE), *SOX10* ($p = 1.57 \times 10^{-14}$, MNE) and *KRT6B* ($p = 3.24 \times 10^{-18}$, MNE) relative to the basal module alongside novel associations of *ROPN1* ($p = 3.07 \times 10^{-55}$, MNE), *ROPN1B* ($p = 1.03 \times 10^{-29}$, MNE) and *RGMA* ($p = 2.86 \times 10^{-35}$, MNE) (Figure 3.5). These findings re-enforce the significance of *FOXC1* and *SFRP1* (Figures 3.6 and 7.4) in the etiology of basal-like breast cancer but establishes that it is not the absolute expression level of these genes that characterises divergent cancer samples, but rather their aberrant expression relative to similar genes in a minority of samples. This comparison to the module expression level is reflective of heterogeneous cellular content of tumour samples where canonically non-basal samples are expected to have some basal-like cell content.

Claudin-low like samples are identified as a subcluster within a module enriched for stromal associated genes including multiple collagens and matrix metalloproteinases (Stromal-atypical subcluster). These samples express low levels of *COL10A1* ($p = 2.41 \times 10^{-60}$, MNE), *COL11A1* ($p = 8.64 \times 10^{-34}$, MNE), *MMP11* ($p = 1.40 \times 10^{-74}$, MNE) and *PPAPDC1A* ($p = 2.97 \times 10^{-31}$, MNE) compared to their module mean (Figure 3.7, Table 7.3). Many of the samples classified as the Normal-type are comparatively re-classified as members of the Stromal-atypical subcluster (24/34 samples) suggesting similarity between these two signatures.

A subcluster within the luminal module is similar to the Her2-enriched subtype but is instead distinguished based on the loss or retention of particular luminal transcriptional regulators. The luminal module divides the samples into three clusters the smallest of which largely consists of samples classified as Her-2 enriched using the PAM50 centroids (50/81, Her2-enriched samples) [29]. We found that while 35/81 subcluster samples were labeled as Her2 receptor positive these samples were best distinguished by retention of *SPDEF* ($p = 5.29 \times 10^{-28}$, MNE) and *FOXA1* ($p = 4.33 \times 10^{-33}$, MNE) expression (Figure 7.5) alongside progressive loss of *ESR1* expression ($p = 2.50 \times 10^{-18}$, MNE) (Figures 3.8 and 7.5, 37/81 are also ER receptor negative). As many of these samples display broad range of *ESR1* expression it is possible that these tumours have progressed from an early luminal-type towards a Her2-like signature.

Here I show that established BRCA subtypes can be replicated and rephrased using CMC coexpression analysis and intra-module CMC subclustering. Module-based analysis of cancer types can provide more precise distinctions between clusters of samples and pinpoint the miss regulated genes characteristic of different types. Expression of basal-associated genes is not confined to basal samples and coexpression of these genes can be demonstrated in luminal samples. The basal, claudin-low and Her2-enriched-like samples are identified as distinct from the otherwise indistinct remaining luminal-like BRCA samples by variation within independent features. The basal subtype is best characterised by low expression of luminal genes and relatively high expression of a subset of established basal marker genes with *FOXC1* as exemplary. Her2-enriched-like samples are characterised by the loss of *ESR1* and retention of luminal markers like *FOXA1* and *SPDEF*. Variation in these modules may be either inter or intratumoural in origin.

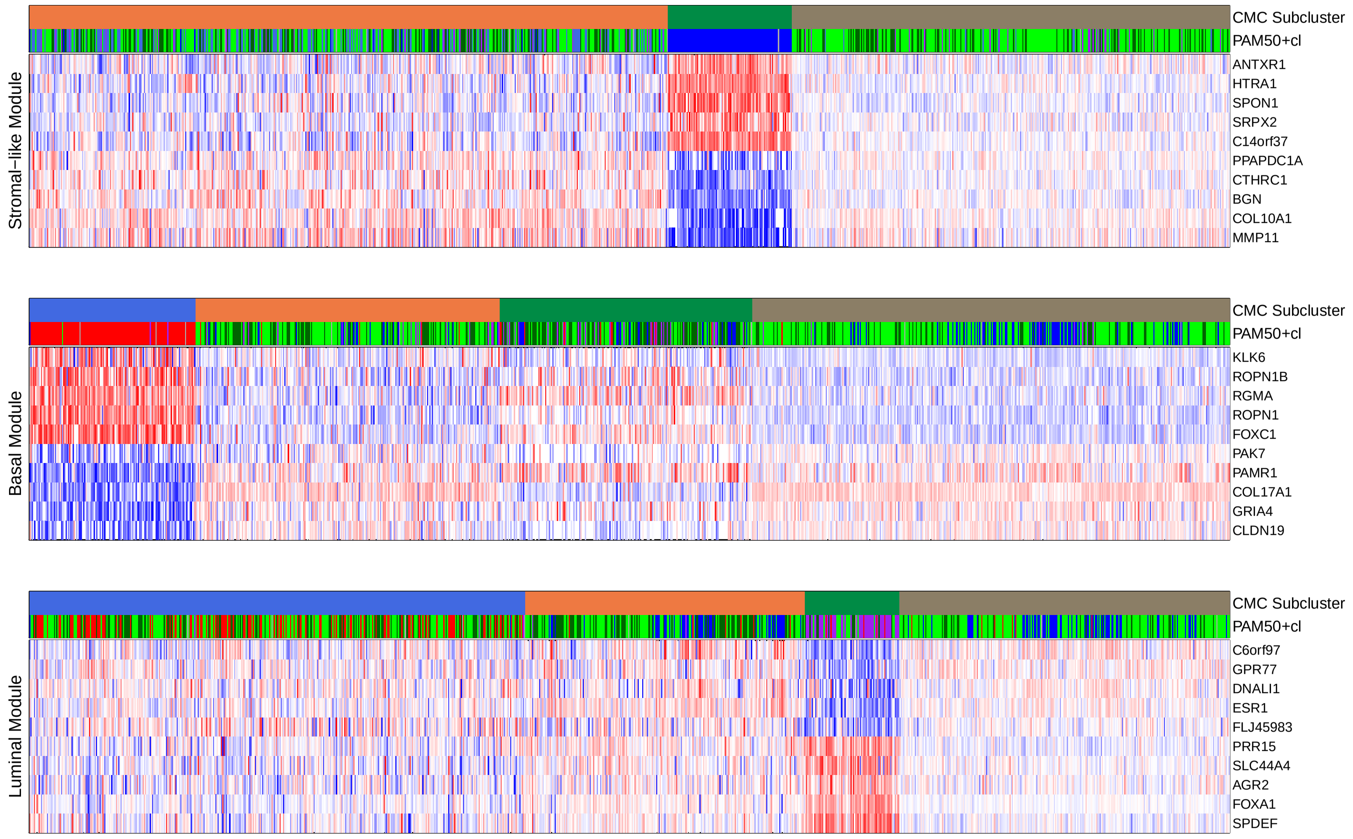


Fig. 3.5 Independent intramodule variation identifies consensus subtypes in BRCA samples. Clustering within the stromal-like, basal and luminal modules reveals sample clusters equivalent to the established claudin-low, basal and Her2-enriched subtypes respectively. Ten differentially expressed genes for each cluster and module are shown. Column markers for each heatmap represent both module dependent CMC subcluster (top) and PAM50/claudin-low subtype classifications (bottom). Units are row z-score normalised log2 FPKM with column means subtracted to highlight within module expression differences.

| PAM50 subtype | Basal | Claudin-low | Her2-enriched | Luminal A | Luminal B |
|---|---|---|---|---|---|
| Intersection (% of subcluster) | 95 | 99 | 62 | 65 | 43 |
| Intersection (% of subtype) | 90 | 89 | 91 | 69 | 55 |
| Subcluster module | Basal | Stromal-like | Luminal | BRCA 28 | Basal |

Table 3.2 Table showing the intersection of CMC subclusters with established PAM50/claudin-low classification. Both basal and claudin-low show comparable subtype and subcluster classification (between 89% and 99%). A CMC subcluster intersects 91% of samples classified as Her2-enriched yet Her2-enriched samples compose only 62% of CMC subcluster samples suggesting the Her2-enriched PAM50 classification may misidentify many samples. The distinction between the luminal A and B PAM50 subtypes is not well defined by CMC subclustering with relatively poor intersected sample sets.

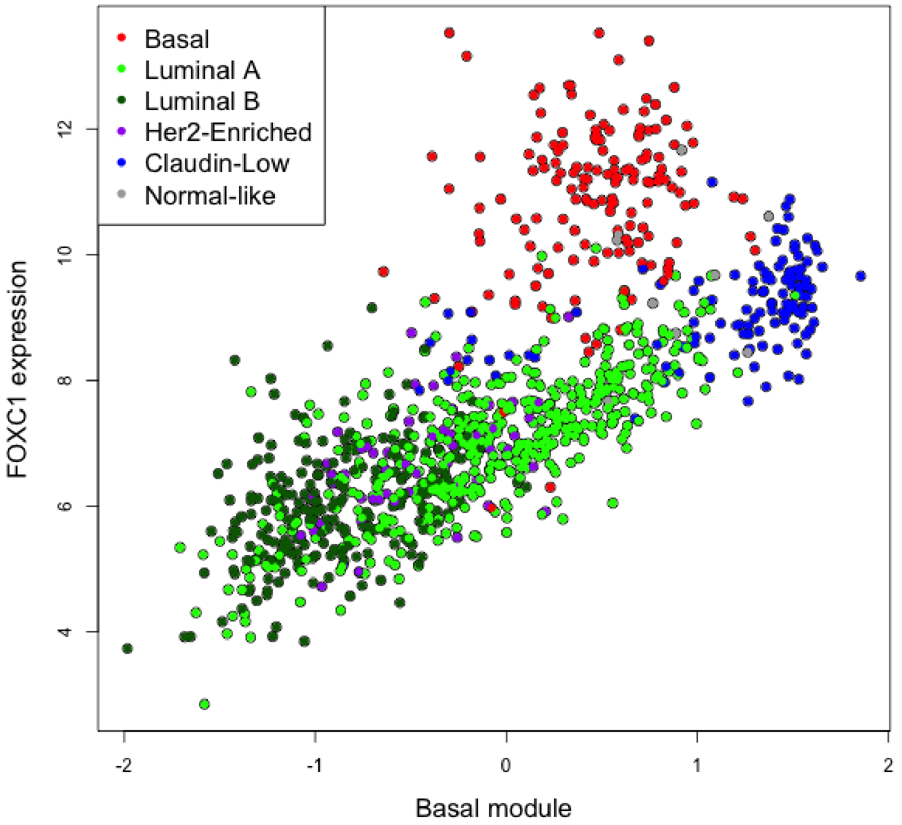
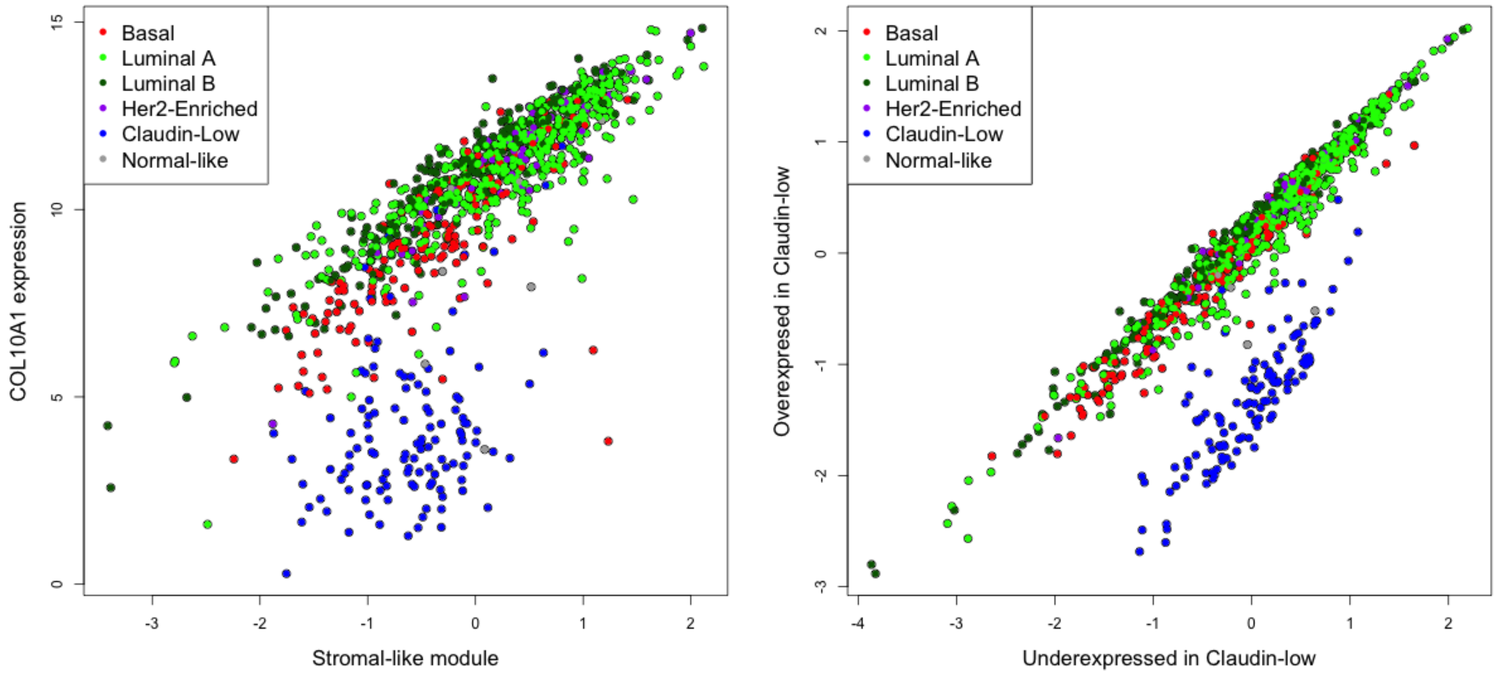


Fig. 3.6 Basal module versus *FOXC1* expression showing basal module expression variation in luminal samples and significantly higher than expected *FOXC1* expression in the basal samples (p = $4.54 \times 10^{-51}$, MNE). While basal samples do express high levels of *FOXC1*, these samples are better distinguished by high *FOXC1* expression relative to other basal module genes. Colours indicate PAM50/claudin-low classification. Units are log2 FPKM for the y axis and z-score normalised log2 FPKM on the x axis.



(a) (b)

Fig. 3.7 Expression of genes that distinguish claudin-low samples. While other BRCA subtypes show expression of COL10A1 consistent with other stromal-like module genes, claudin-low samples are distinguished by particularly low expression (p = 2.41 × $10^{-60}$, MNE) (a). Low expression of other stromal-like module genes further distinguish and characterise claudin-low samples (b and Table 7.3). Units are z-score normalised log2 FPKM.

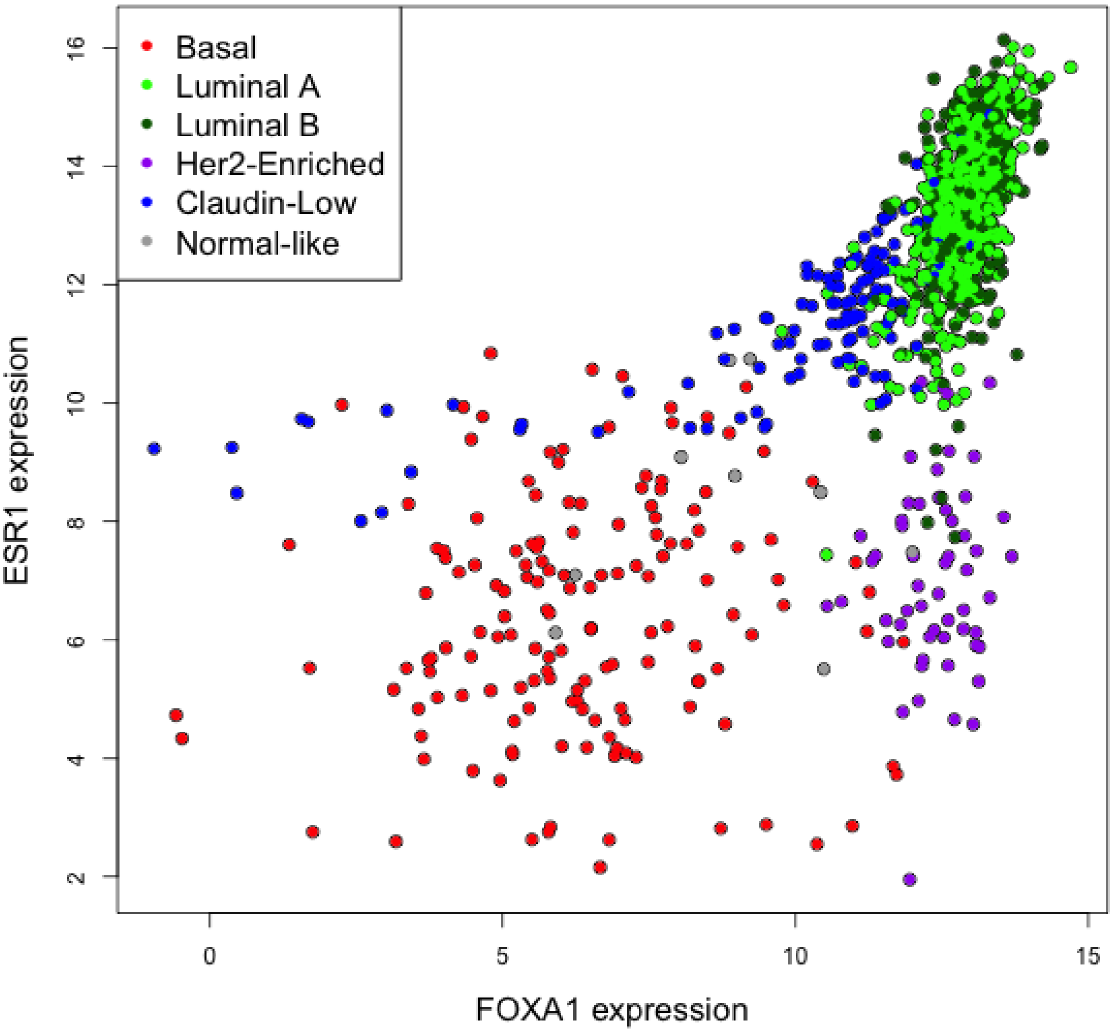


Fig. 3.8 Expression of *FOXA1* and *ESR1* distinguish Her2-enriched samples. Low expression or genomic loss of *ESR1* (p = $2.50 \times 10^{-18}$, MNE) and retention of luminal subtype level expression of *FOXA1* (p = $4.33 \times 10^{-33}$, MNE) characterises the Her2-enriched subtype more accurately than Her2 amplification (35/81 samples Her2 receptor positive). Units are log2 FPKM. Colours indicate PAM50 classification.

### 3.2.5 Glioma specific signatures identify components of intratumoural variation including a Proneural to Mesenchymal axis

Having established that CMC module subclustering can replicate and expand upon established BRCA subtypes I considered if these methods may help resolve the lack of subtype consistency in glioma [166]. Applying CMC to glioma, numerous glioma or brain specific modules that are not reproduced in other cancer types (Figure 1) including mature neural and oligodendrocyte like expression signatures are identified. Alongside these I find modules enriched for genes previously associated with the proneural and mesenchymal subtypes [198, 270]. The proneural module was identified by the expression of common Proneural marker genes including *OLIG1/2*, *BCAN*, and *DLL3* (26% shared genes with Verhaak centroids, Table 7.4). The mesenchymal module was likewise identified with expression of *CHI3L1/2*, *CD44* and *CEBPB* (57% shared genes with Verhaak centroids, Table 7.4).

The proneural and mesenchymal modules are anti-correlated to each other (Figure 3.9, Pearson's r = 0.89, $p < 2.2 \times 10^{-16}$) as well as being highly correlated to the second principal component of the same data (proneural to PC1 Pearson's r = -0.98, $p < 2.2 \times 10^{-16}$) (Figure 3.9). The anti-correlated proneural and mesenchymal features combine to compose a dominant axis of proneural to mesenchymal expression. Samples previously classified by Verhaak *et al.* [270] as discrete subtypes are distributed along this proneural to mesenchymal axis without discrete separation into distinct subtypes (Figure 3.9). In comparison to BRCA tumours, subclusters within any of the glioma modules do not replicate any of the four Verhaak *et al.* [270] subtype clusters (Table 3.3).

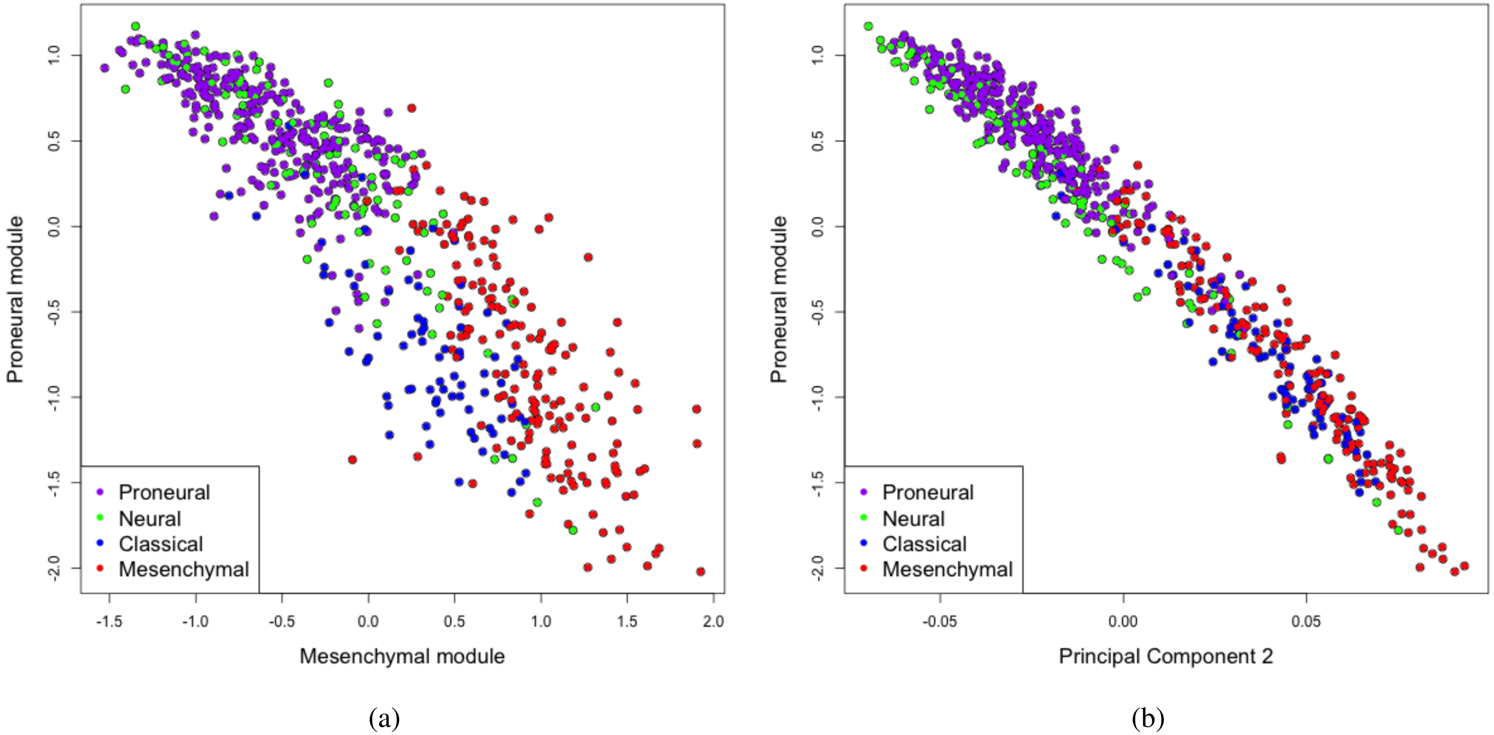


Fig. 3.9 Glioma samples present a dominant proneural to mesenchymal axis. A mix of high and low grade samples classified using the centroids described by Verhaak *et al.* [270] display anti-correlated proneural and mesenchymal modules (a). The proneural module is also highly correlated to the second principal component of this dataset (b). Colouring indicates Verhaak *et al.* [270] subtype classification. Units are z-score normalised log2 FPKM.

| Verhaak subtype | Proneural | Mesenchymal | Classical | Neural |
|---|---|---|---|---|
| Intersection (% of subcluster) | 78 | 61 | 35 | 71 |
| Intersection (% of subtype) | 70 | 44 | 66 | 27 |
| Subcluster module | Neural-like | Neural-like | Glioma 14 | Glioma 20 |

Table 3.3 Table showing the intersection of CMC subclusters with Verhaak *et al.* classification. Poor intersection of subclusters with Verhaak *et al.* subtypes indicate a potential instability of Verhaak *et al.* and CMC subtypes relative to subtype/subcluster assignments in BRCA (Table 3.2).

In comparison to the Verhaak *et al.* subtypes, the intersection between CMC subclusters and the recently described IDH/codeletion molecular subtypes [30, 34] of glioma is more consistent (Table 3.4). Replication of the IDHwt molecular subtype is relatively strong compared to the Verhaak *et al.* subtypes with 88% of subtype and subcluster samples overlapping. The overlap for the IDH mutant subtypes is comparatively poor to the IDHwt subtype however the high intersection of the subcluster in non-codeletion (96%) and subtype intersection in the 1p/19q codeletion subtype (99%) indicates a consistent segregation of subtypes and subclusters. This suggests further substructure within the module-normalised data that is not captured by a single division of the module into *k* clusters. This may be due to the lack of extreme subtype differences for each module subcluster in glioma. In comparison the BRCA basal subtype can be robustly identified due to the consistent expression of basal-like genes as a coexpression module in all samples yet present aberrant expression of *FOXC1* within the basal subtype. Investigation of this within-module substructure, with robust subcluster extraction to more powerfully identify subtle expression differences, may be the focus of future work with these coexpression methods.

Comparing IDH/codeletion molecular subtypes to the Verhaak *et al.* classifications I find that proneural subtypes are most frequently IDH1/2 mutants (96% of proneural samples) where the majority of mesenchymal (73%) and classical (99%) tumours are IDH wild type (Table 3.5). Only the classical type is highly segregated into the single IDH/codel molecular subtype that lacks mutations in either IDH1 or IDH2. In order to test the robustness of both subtype classifications in the TCGA RNA-seq dataset, the silhouette and connectivity tests were applied. The silhouette test asks to what degree of confidence can we assign to the cluster identity of each observation with values increasing towards 1 indicating better clustering performance. Applying this test finds the IDH/codeletion subtypes have the highest average silhouette width at 0.08 compared to 0.06 in the Verhaak *et al.* subtypes. The minimum silhouette width for the IDH/codeletion subtypes was 0.05 for the IDH/non-codel subtype yet in the Verhaak *et al.* subtypes, the Mesenchymal subtype presents a negative minimum silhouette width. The connectivity test determines to what extent observations are placed in the same subtype as their nearest samples in the dataset where lower values suggest a better partitioning. For the IDH/codeletion subtypes the connectivity test gives a value of 141.2 in comparison to 465.5 for the Verhaak *et al.* subtypes. Therefore overall in both tests the IDH/codeletion subtypes presented the most robust separation compared to the Verhaak *et al.* subtypes suggesting this classification is more reflective of the transcriptional diversity found in this combined high and low-grade glioma dataset.

| Molecular subtype | IHDwt | IDHmut non-codel | IDHmut codel |
|---|---|---|---|
| Intersection (% of subcluster) | 88 | 96 | 46 |
| Intersection (% of subtype) | 88 | 48 | 99 |
| Subcluster module | Glioma 32 | Mesenchymal | Glioma 14 |

Table 3.4 Table showing the intersection of CMC subclusters with IDH/codeletion molecular subtypes [30, 34]. These intersections are greatly improved in comparison to Verhaak *et al.* subtypes (Table 3.3) suggesting an improved consensus between IDH/codeletion molecular subtypes and CMC analysis. Incomplete intersection of subclusters with IDH/codeletion molecular subtypes [30, 34] may indicate inadequately subclustered structure in the CMC analysis.

| | IDHwt | IDHmut non-codel | IDHmut codel | Total |
|---|---|---|---|---|
| Proneural | 11 (3.9) | 148 (51.9) | 126 (44.2) | 285 |
| Mesenchymal | 117 (72.7) | 40 (24.8) | 4 (2.5) | 161 |
| Classical | 74 (98.7) | 1 (1.3) | 0 (0) | 75 |
| Neural | 28 (28.9) | 44 (45.4) | 25 (25.8) | 97 |
| Total | 230 | 233 | 155 | 618 |

Table 3.5 Table showing the intersection of IDH/codel molecular subtypes with the Verhaak *et al.* classifications. Samples classified as the classical type are typically IDHwt tumours. Similarly IDH1/2 mutant tumours are most commonly classified as members of the proneural subtype. Percentage of each molecular subtype per Verhaak *et al.* subtype shown in brackets.

The mesenchymal module includes many genes identified in the cross-cancer consensus immune and interferon response modules. Lowering the cut-off parameter for module clustering to 0.15 from 0.2 allows for the separation of interferon, immune cell and mesenchymal-like modules highlighting the interplay between interferon signature, immune response and mesenchymal subtype expression (Table 7.5). Comparing these reduced mesenchymal and immune cell modules separates mesenchymal and classical samples with classical samples showing a significantly reduced immune cell related expression (Figure 3.10).

This proneural to mesenchymal axis was confirmed in two independent datasets [198, 245] (Figure 3.11). Low-grade glioma samples tend to present a more proneural signature and the high-grade glioblastoma samples present a more mesenchymal signature with no clear separation of the samples based on clinically defined grade (Figure 7.7). Examination of the CMC derived proneural and mesenchymal modules within an intratumourally sampled dataset [245] allowed me to trace this relationship within individual tumours. Remarkably we found the proneural and mesenchymal modules to vary both within and between tumours whilst mirroring the anti-correlated relationship found in the TCGA datasets (Figure 3.11, Pearson's r = -0.74, p-value = $1.76 \times 10^{-09}$). The presence of proneural to mesenchymal variation within individual tumours and its characterisation as non-discrete gradient across high and low grade glioma samples suggests this axis is associated with glioma progression rather than distinct patient tumour subtypes.

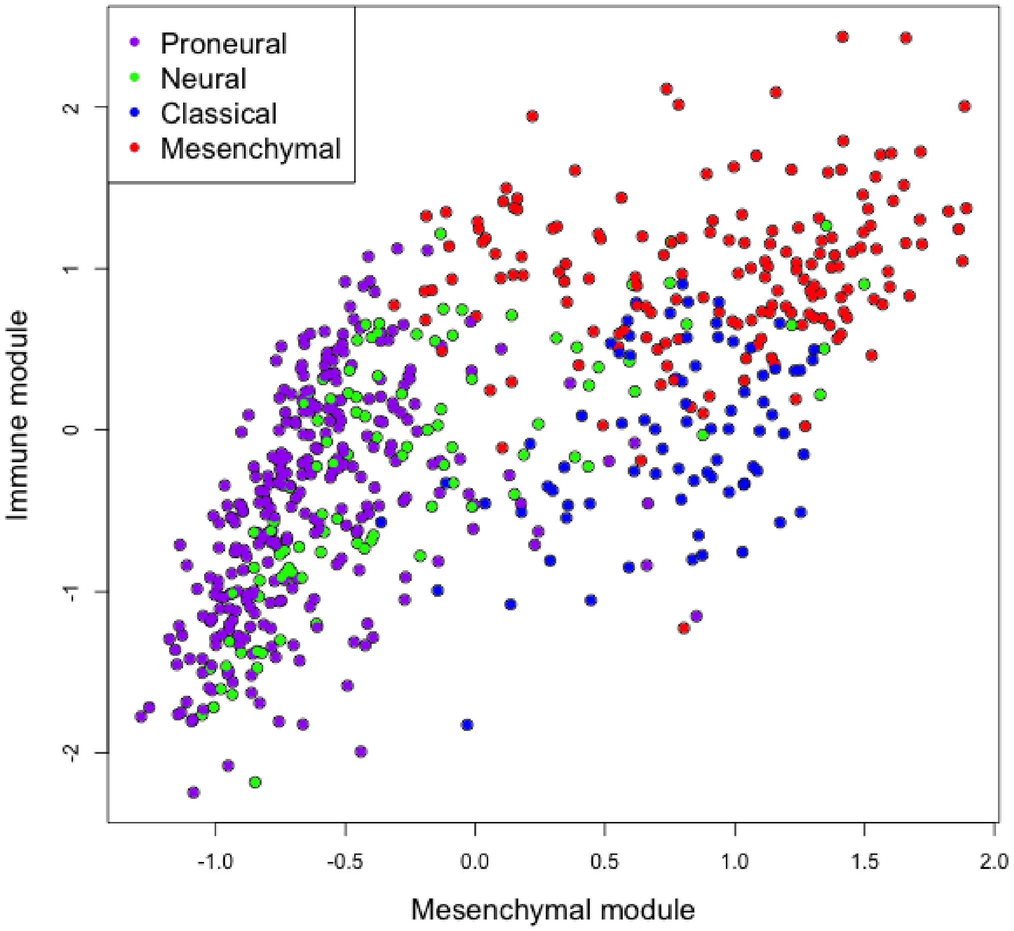
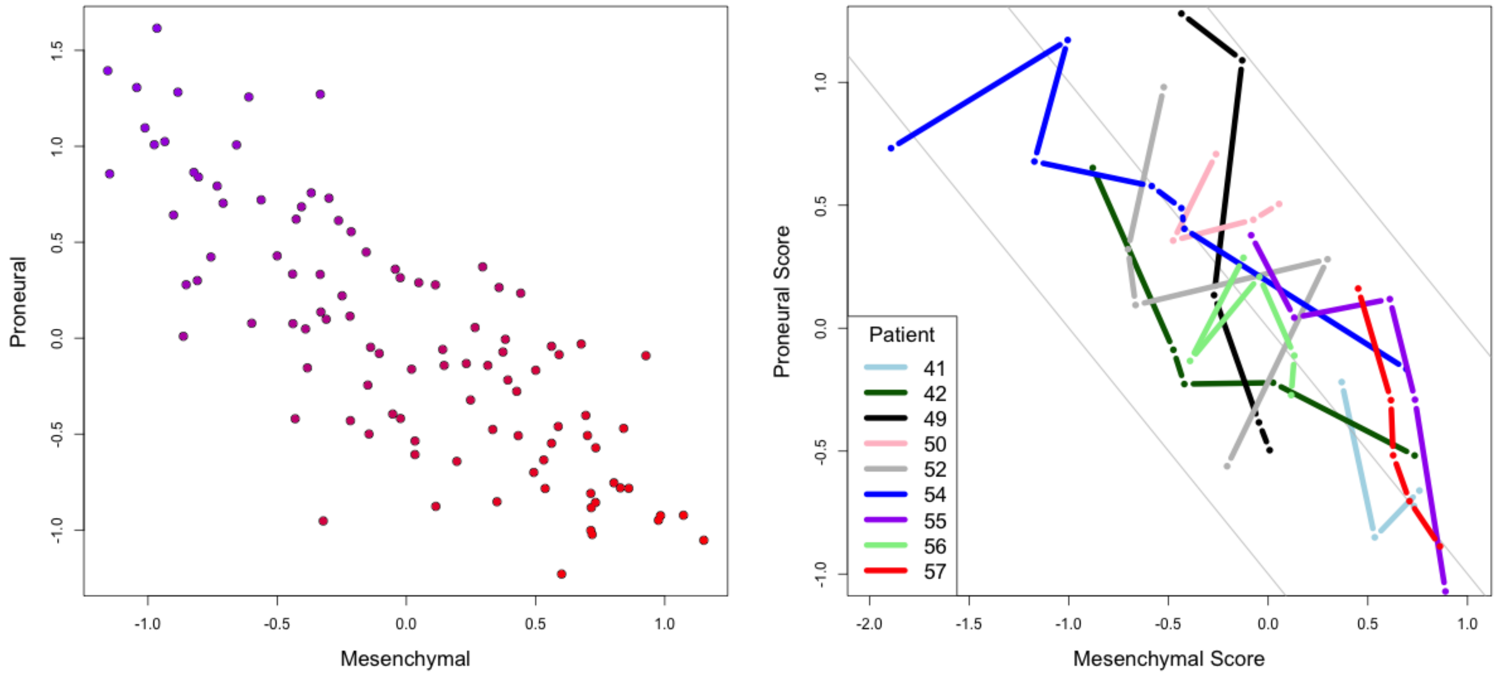


Fig. 3.10 Expression of reduced mesenchymal and immune cell modules distinguishes mesenchymal and classical subtype samples. Subsetting the larger mesenchymal module and extracting submodules that represent immune cell and mesenchymal like expression (Table 7.5) allows for the distinct separation of the Verhaak *et al.* classical and mesenchymal samples from the proneural type. The neural type presents a mixed relationship between expression of these two submodules. Units are z-score normalised log2 FPKM.



(a) (b)

Fig. 3.11 The proneural to mesenchymal axis can be found in other datsets. Expression of proneural and mesenchymal modules in a microarray derived glioma expression dataset described by Phillips *et al.* (a) [198] alongside an intratumourally sampled dataset described by Sottoriva *et al.* (b) [245]. Coloured lines indicate samples extracted from independent tumours. Dot chart coloring indicates relative proneural versus mesenchymal expression (Left panel). Units are z-score normalised log2 microarray intensity units.

Comparing other glioma derived modules between samples taken from the tumour margin and the main tumour bulk allows the study of these features in an intratumoural context. Samples taken from the tumour to normal brain margin have a significantly higher expression of the proneural (p = $1.08 \times 10^{-04}$), oligodendrocyte (p = $1.86 \times 10^{-06}$) and neural (p = $7.83 \times 10^{-04}$) modules in comparison to the main tumour bulk and considerably lower expression of the mitosis (p = $4.15 \times 10^{-15}$) and mesenchymal associated modules (multiple values, $p \leq 0.01$, Figure 3.12). The differential expression of features between the tumour margin and bulk supports the association of the mesenchymal and Interferon feature with the tumour bulk alongside identifying the neural and oligodendrocyte modules as enriched within likely post-mitotic normal brain cells on the tumour margin.

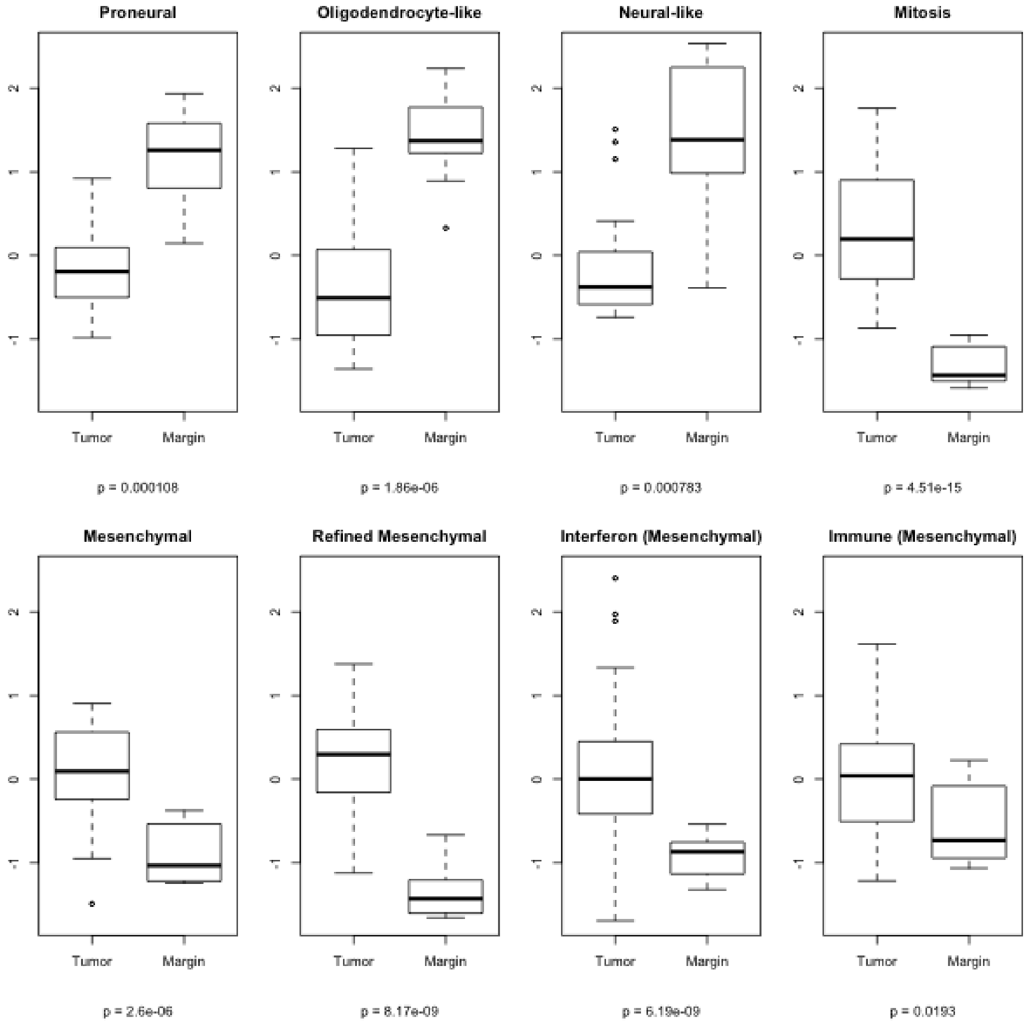


Fig. 3.12 Expression of CMC modules between the tumour bulk and margin (Data from Sottoriva *et al.* [245]). Proneural and mesenchymal modules show differential expression between the tumour bulk and margin. Significant overexpression of the neural-like and oligodendrocyte-like modules alongside low expression of the mitosis module may represent an enrichment of post-mitotic non-neoplastic cells along the tumour margin. Mesenchymal submodules are described in Table 7.5. Differential expression of module expression was tested using a Welch's two sample T-test ($p \leq 0.05$).

In summary distinct from previous studies I describe a non-discrete dominant gradient of proneural to mesenchymal expression. This gradient is present intratumourally and is associated with a mesenchymal phenotype, immune cell infiltration and tumour grade. In contrast to BRCA tumours the established discrete subtypes set out by Verhaak *et al.* [270] were not identified as distinct subtypes in any module subclusters. The intersection of CMC subclusters with the IDH/codeletion molecular subtypes suggests substructure within the module clustering may not be effectively representing the underlying subtile biological signal in glioma.

## 3.3 Discussion

Here I describe independent components of transcriptional variation that represent established tumour processes and relate to previously described cancer subtypes. Identified here, components of the tumour transcriptome shared across multiple tumour types that are representative markers of independent biological features, such as cell division and immune cell infiltration, which can be expected to vary in an intratumoural fashion. While previous studies have identified similar cross cancer type features using more broadly defined coexpression networks [41, 50, 179, 280] I seek here to extend this intratumoural heterogeneity context to tumour subtypes. I recast the cancer subtype problem moving away from methods that impose reductive and assumed transcriptome-wide discrete subtypes towards the identification of independent coexpression modules followed with intramodule clustering to identify discrete within-module subtypes. Following this methodology I replicate and redefine the luminal, basal, claudin-low and Her2 enriched subtypes in BRCA and set out the novel identification of an intratumoural non-discrete proneural to mesenchymal axis in glioma.

Establishing consensus reproducible tumour subtypes has been difficult for many tumour types [104, 127, 166]. Established subtype methods focus on identifying the global expression differences between tumour samples and classifying them into discrete categories, reducing all the apparent complexity of tumour biology into a subset of signatures where each tumour can be classified as a single subtype. As an alternative I focus on the independent components of transcriptional variation found in all samples and between tumour types. In this way, signatures that are unique to a single tumour type, or samples that show a distinct variation in a common expression signature, can be identified and reproduced independently of a reductive global transcriptome subtype.

Here I show that coexpression analysis via CMC can replicate and improve upon tumour subtypes derived with established methods, also demonstrating how clustering within a coexpression module can reveal distinct subpopulations independently from subtype classifications using other features. In the case of BRCA, I identify modules that replicate the division of samples into the luminal and basal discrete subtypes established by TCGA analysis [29]. Curiously the basal samples did not express the highest levels of our basal module but instead were identified based on variation within the basal module and low luminal module expression. The samples that did however express the highest levels of the basal module were identified as analogs of the claudin-low subtype. This claudin-low subtype, defined independently by variation within the stromal-like module, shows a distinct similarity to metaplastic breast carcinoma (MBC) [281] and could represent a transcriptional signature of MBCs as discussed previously [206, 207]. Basal BRCA samples are identified primarily by the unusually high expression of *FOXC1* compared to other basal-like genes. Likewise I show the Her2-enriched subtype is characterised by retention of luminal transcription factors including *SPDEF* [23] and *FOXA1* [168, 260] and reduced *ESR1* expression. This critical difference may be of significant importance to distinguishing, treating and developing therapeutics for tumours of this type.

By contrast, in glioma, no comprehensive discrete separation of samples into the established subtypes is found using CMC analysis and instead I find a gradient of proneural to mesenchymal expression. This proneural to mesenchymal axis is mirrored by the enrichment of the mesenchymal signature in high-grade gliomas. I show using intratumourally sampled data that this proneural to mesenchymal variation is present within multiple tumours. Based on these findings it is easy to suggest that the proneural to mesenchymal axis represents a signature of progression that is intrinsic to glioma biology rather than a signature that can be attributed to a subset of tumours. This hypothesis is supported by the observation that the mesenchymal type tends to present a worse prognosis, which might be expected should the mesenchymal signature represent tumour progression and high grade. The association of immune cell module as correlated to, and contained within, the mesenchymal module further reinforces the mesenchymal progression hypothesis. The mesenchymal signature has also been associated with other markers of progression like necrosis but also within regions of the same tumour which can present different patterns of subtype signature expression [43]. The observation that recurrent tumours tend to drift towards the mesenchymal type is also of note here [198] further supporting the progression model. Similarly the association of *NF1* mutations with the mesenchymal type may suggest that this mutation is only beneficial to cells expressing a mesenchymal progression program potentially explaining why tumours induced via *TP53* and *NF1* deletions gave rise to proneural-like tumours [154]. Recent work suggests that most glioma tumours may evolve from a proneural like precursor cell with subsequent mutations in *NF1* [191].

When CMC derived subtype definitions are compared to both the IDH/codeletion and Verhaak *et al.* subtypes a greater overlap is detected for the IDH/codeletion subtypes. This finding combined with improved performance in the silhouette and connectivity tests suggests this established genome focused subtype definition is more reflective of the biological reality. The difficulty CMC analysis has with deriving robust subtypes may be reduced through a process of testing coexpression modules at a range of cut off heights and varying the number of within module clusters to avoid ignoring potentially important within-module substructure. Combining these results into a psudo-bootstrapped consensus CMC subtype. With future work on the structure of these coexpression methods it may be possible to derive or validate established glioma subtypes.

In conclusion I propose that the discrete separation of samples into subtypes should not be assumed without prior investigation into the nature of the tumour specific variation. The assumption of discrete subtypes presents misrepresentative overview of glioma and potentially other tumour subtypes. While some tumor types, like breast invasive ductal carcinoma, present expression variation that readily supports their division into distinct subtypes, other tumour types like glioma present a more convoluted route to defining discrete subtype signatures. Moreover high or low expression of a coexpression module, which can dominate a standard transcriptome-wide clustering, can give the misleading impression of a discrete subtype, such as in the case of proneural and mesenchymal modules in glioma, which is not matched by differential coexpression of module component genes. The definition of independent markers of biological signatures as expression modules and the samples that behave differently within these modules presents a battery of potential targets and markers for further investigation, whilst also providing a more intuitive intratumoural context to tumour subtype analysis.

# Chapter 4

# Transcriptomic analysis of glioma derived neural stem cells

## 4.1 Introduction

Cancer stem cells are a critical component of the tumour mass that are capable of repropagating or recapitulating the tumour. As such, the characterisation of cancer stem cells is of particular significance. For glioma, cancer stem cells have been derived from tumour samples and sustained in culture using a number of methods. The ability to culture cancer stem cells provides a flexible *in vitro* research model platform and enables high throughput drug discovery screens. The culture of glioma stem cells has been enabled by parallel development of methods developed for normal neural stem cells (NS cells). Early methods for NS cell culture established the expansion of cell clusters as neurospheres in suspension with the growth factor EGF [219]. Later methods added FGF-2 to the media improving the isolation of multi-lineage potential NS cells [5, 85]. Culture of NS cells in neurosphere suspension has a number of problems including the heterogeneous phenotype of neurosphere cluster cells and the progressive loss of self renewal and differentiation capacity [218]. As an alternative to neurosphere culture the growth of NS cells adherent to laminin coated surfaces suplemented with EGF and FGF-2 growth factors was found to enable sustained symmetrical self renewal without differentiation [42, 203, 251]. These adherently grown NS cells can be established from both adult and fetal forebrain tissue and well as differentiated from embryonic stem cells *in vitro* [203]. Further work went on to show that these adherent NS cells are capable of differentiating into the three primary lineages of central nervous system cells, showing the potential for oligodendrocyte differentiation which had previously remained enigmatic [251].

### 4.1.1 Glioma stem cell culture

The derivation of cell lines from glioma tumours has a long history similarly to NS cell culture methods. Early work established that a subset of cells from dissociated tumour biopsy tissues were able to self renew when suspended in fetal calf serum containing media [205]. Expansion of these cells in serum contrasts with NS cells which have been shown to irreversibly differentiate in serum containing media [70]. While many glioma lines tolerate serum, the long term effects of this media are not well established and may cause these cells to shift away from their initial cultured state. This may affect how representative the cell line is of the glioma stem cells *in vivo* [263]. Improved methods for culturing NS cells, i.e. removing serum and using EGF and FGF-2 growth factors, were subsequently applied to glioma derived stem cells [71, 106, 143, 237]. Likewise the addition of adherent culture methods to glioma derived cells enabled similar advantages to adherent culture of NS cells with reduced differentiation and greatly improved stability (Figure 4.1) [204]. These adherent EGF and FGF growth factor cultured cells were described as glioma neural stem cells (GNS) due to their distinct phenotypic similarities to normal NS cells. While GNS cells have similarities to NS cells, GNS cells are able to recapitulate the tumour mass following xenotransplantation and retain the hallmark genomic aberrations. Adult NS cells are also a strong candidate for the cell in which the initial neoplastic activation takes place, transforming it into a GNS counterpart [151].

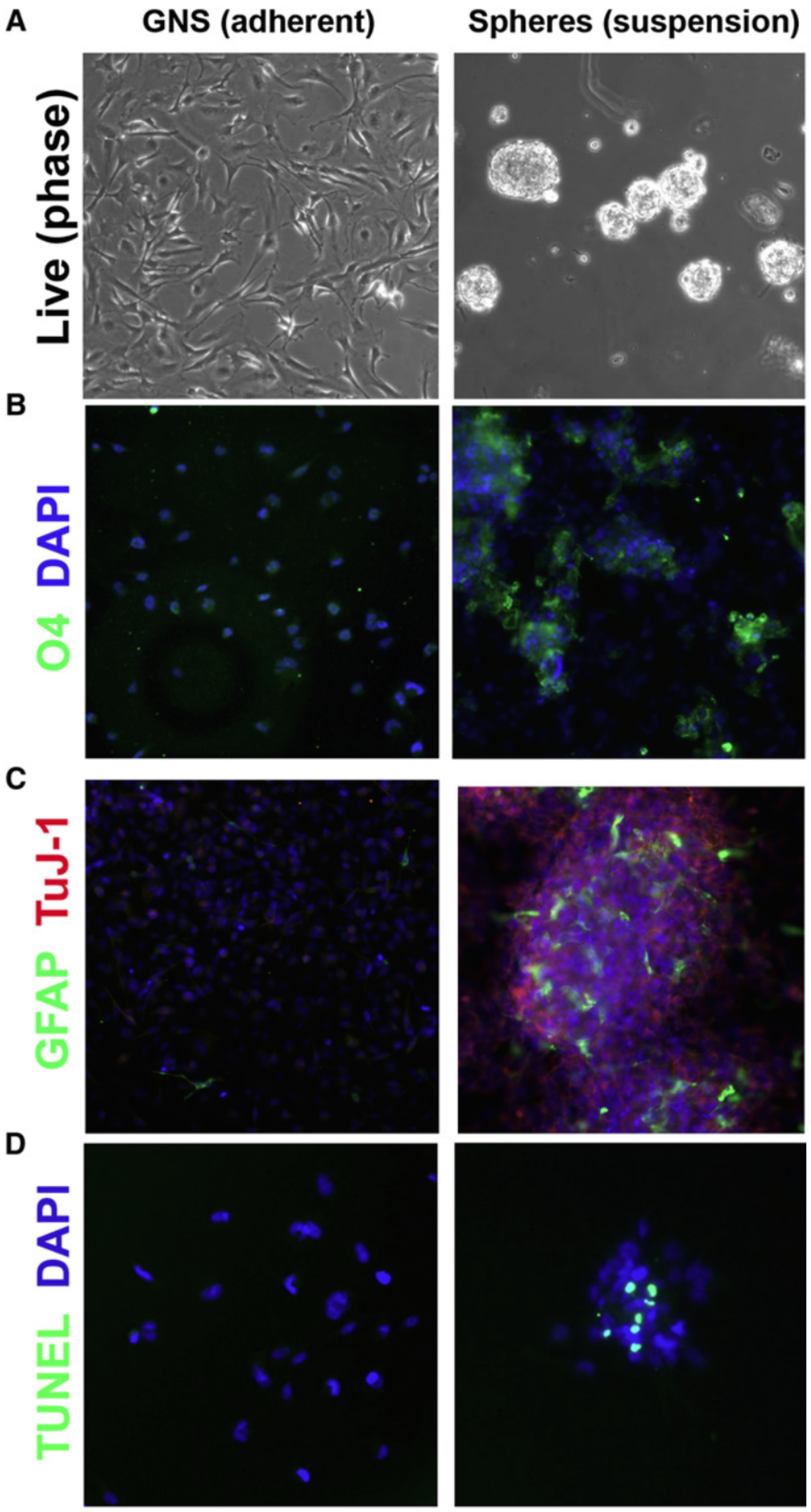


Fig. 4.1 A comparison of neurosphere vs adherent culture in G144 cells. Immunocytochemical staining for neural differentiation markers as well as TUNEL assays for apoptosis in adherent (Top row) and neurosphere (Bottom row) culture [204]. Neurosphere cultured cells show signs of astrocyte (GFAP), neural (TuJ-1) and Oligodendrocyte (O4) differentiation alongside increased apoptotic activity compared to the adherent cultured cells.

With greater *in vitro* consistency and stability brought about by improved culture methods, efforts to characterise and exploit glioma derived cell lines expanded. tumour derived stem cells have been shown to closely mirror the phenotype and genotype of their parental primary tumours [143, 237]. In these studies, serum culture of these cell lines, compared to EGF and FGF-2 growth factors, was shown to induce morphological, transcriptomic and glial lineage differentiation alongside decreased tumourigenicity *in vivo* for early passages. Curiously late passage serum cultured cells were able to form xenograft tumours at an increasing rate with further passages [143].

### 4.1.2 The phenotypic diversity of glioma stem cells

The specific features and markers that define these glioma stem cells has been a contentious question with multiple potential definitions of what defines this diverse population. In an early study, CD133 was identified as a marker of stem cells derived from CNS tumours, that can proliferate as neurospheres in growth factor media and also show differentiation towards similar lineages as their parental tumours [237]. CD133 negative glioblastoma cells were found to grow as adherent clumps before terminally differentiating. Expression of CD133 is absent from a majority of normal adult neural stem-like populations but is expressed in embryonic neural stem cells [197]. Using a limited panel of qPCR targets diversity within glioma stem cells (GSC) indication of functional diversity was observed [71], however it was not until transcriptome wide assay methods like microarrays became ubiquitous that large panel studies of both tumours and tumour derived cell lines became commonplace [8, 87, 194, 198, 204]. Examination of cells derived from both primary and secondary glioblastomas found that no significant *in vitro* growth was observed for secondary glioma cells [8]. In contrast cells derived from primary glioblastomas readily proliferated long term either as CD133+ neurospheres or CD133- adherent spheres (Figure 4.2). The distinction between CD133+ neurospheres and adherent, mostly CD133- cultures was replicated by gene expression clustering [87] revealing some similarity to the previously described proneural and mesenchymal tumour subtypes [198]. The more proneural subtype-like CD133+ cells comparatively over expressed neural development associated genes like *OLIG2*, *BCAN*, *DLL3* and *NES*. By comparison, the CD133- mesenchymal looking cells overexpressed *MET*, *LOX*, *CAV1* and a number of extra cellular matrix associated genes [87]. These two GSC types were inferred to relate to the differences between adult and fetal neural stem cells, further suggesting that these normal populations may be the different cells of origin for different tumour types [157]. Significantly this study also showed that

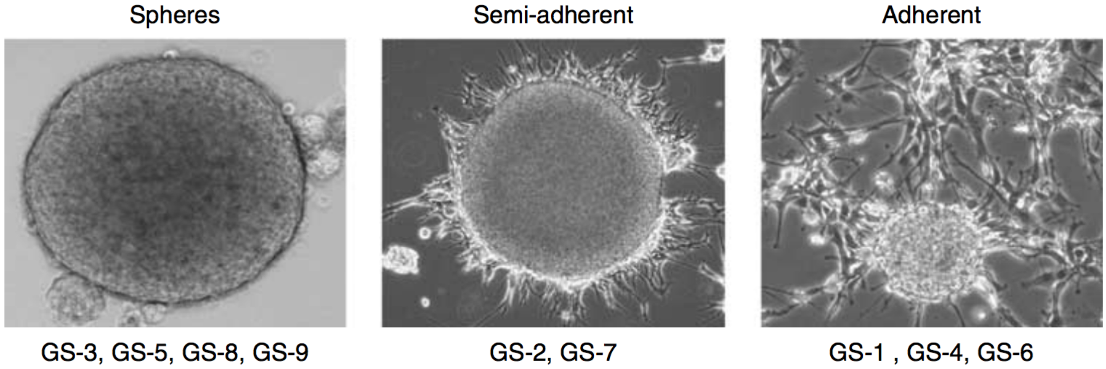


Fig. 4.2 CD133+ GSC neurospheres and CD133- adherent cell morphology. Morphological differences between the CD133+ proneural-like neurospheres (Left panel) and CD133- mesenchymal-like adherent and semi-adherent cells (Centre and right panels). [87].

adherent culture was capable of culturing and continuously expanding both GSC types with the aforementioned lower levels of differentiation.

Differences in morphology between GSC cultured in either serum or EGF and FGF-2 growth factors were observed previously, with an apparent preference towards adherent growth in serum culture [143]. Cell lines initially established as neurospheres in growth factor media switched towards an adherent fibroblast-like morphology when transferred to serum containing media. Transferal to serum containing media induced a temporary reduction of proliferation followed by a return to exponential growth after approximately 24 hours. As a counter point, GSCs lines established in serum culture were unable to subsequently expand in growth factor media. This shift towards CD133- mesenchymal-like GSCs has also been shown to be induced by radiation [11, 90, 164]. Mesenchymal-like GSCs were found to express high levels of aldehyde dehydrogenase genes and proneural-like cells, induced by radiation, shifted towards a mesenchymal expression signature alongside expression of the aforementioned aldehyde dehydrogenase genes [164]. Later cell lines derived from mesenchymal subtype tumours were shown to look comparatively proneural and only after exposure to radiation or TNF-$\alpha$ did these cell express a mesenchymal signature [11]. Mesenchymal-like GSCs were shown to have a poorer response to radiation in murine xenografts compared to their proneural counterparts. The key regulators of this mesenchymal shift were proposed to be NF-$\kappa$B and TNF-$\alpha$.

### 4.1.3 Characterisation of adherent GNS cells

Exploiting the comparative *in vitro* stability of adherent glioma derived cells, or glioma neural stem cells (GNS), allowed for detailed investigation of glioma stem cell diversity and relative consistency in high throughput chemical screening against multiple GNS lines [204]. Application of this high throughput methodology to a diverse panel of GNS and NS cell lines found that GNS lines were relatively susceptible to polo-like kinase1 (PLK1) inhibitors compared to the non-neoplastic NS cells [47].

The comparison between karyotypically normal NS and aneuploid GNS cells is an area of substantial interest. Comparing between NS and GNS cells may help identify the tumour enabling differences but also allow for the identification of GNS specific features that may be targeted therapeutically without affecting resident adult NS cells. Using Tag-seq data generated for four GNS and two fetal NS lines key transcriptional differences between GNS and NS cells were identified [55]. In particular this study identified the transcriptional regulators *FOXG1* and *CEBPB* as highly expressed in glioma alongside downregulation of *PTEN* and *TUSC3* amongst others. Expansion of this GNS to NS comparison to a wider panel could reveal further valuable differences. Likewise a larger panel of GNS lines may help explore the huge diversity in GNS cells as may be expected of their many histologically distinct tumour origins. The epigenetic plasticity of GNS cells was explored using induced pluripotent stem cell (iPSC) reprogramming techniques to reset DNA methylation patterns [249]. As GNS cells have natively high levels of SOX2 and C-MYC, transfection of an OCT4 and KLF4 inducible vector induced a shift of transcriptional and DNA methylation program towards an embryonic stem cell (ESC) like state. Differentiation of these ESC-like GNS cells either towards neural or mesodermal followed by xenotransplantation were able to form tumours, however cells directed towards the mesodermal lineage were less malignant and showed a reduced infiltrative capacity than their neural counterparts. This work demonstrates that significant changes in expression and DNA methylation acting within the confines of an aberrant cancer genome were unable to suppress the tumour propagating capacity of these cells.

### 4.1.4 Outlook

In this chapter a panel of 15 GNS lines and 4 NS lines are characterised and compared to other GSC lines in publicly available data. Conventional centroid based subtype classification of these lines is discussed and compared to CMC coexpression methods. The NS to GNS cell line comparison is revisited alongside a detailed investigation of the diversity found between GNS lines as representative of glioma subtypes.

## 4.2 Results

### 4.2.1 Comparing GNS to NS cells reveals glioma specific expression

To extend the analysis by Engström *et al* [55]. we produced expression data for 4 fetal NS lines and 15 GNS lines each with biological replicates using the Affymetrix GeneChip human exon ST array. Differential expression analysis found 335 genes overexpressed in GNS cells and 272 overexpressed in NS cells (Tables 7.6 and 7.7). Clustering expression values for these genes reveals that NS cells present a generally consistent expression pattern between different NS lines (Figure 4.3, mean pairwise correlation = 0.926). In comparison GNS lines show a remarkable diversity in expression of these genes between lines (mean pairwise correlation = 0.077). Restricting the data to genes that are differentially expressed in both this exon array data and the Tag-seq data from Engström *et al.* finds the same pattern of within NS line consistency and GNS diversity (Table 4.1, Figure 7.8). Analysis of gene ontology terms enriched in differentially expressed genes finds enrichment for many neuronal and differentiation related GO terms within NS overexpressed genes where GNS genes show an enrichment for RNA processing and metabolism terms (Tables 7.8 and 7.9). Examination of the genes differentially expressed between GNS and NS cells may help contextualise the functional differences that characterise these cells.

#### Genes overexpressed in GNS

A total of 1191 genes were found to be over expressed in GNS cells compared to NS (Table 7.6). Here we show in GNS cells the transcription factor *FOXG1* is highly over expressed in GNS cells in agreement with previous findings [55] (adj. p = $1.72 \times 10^{-13}$). *FOXG1* has established roles in neural development and disease including telencephalic hypoplasia [284], Rett syndrome [171] and maintenance of adult neurogenesis in the dentate gyrus [261]. Knockdown of *FOXG1* in primary cell lines resulted in reduced neurosphere formation and increaced survival in mouse xenografts [269]. *FOXG1* was identified as a regulator of glioma cell proliferation by binding to a FoxO-Smad complex [230] and was also found to be overexpressed in the non-Shh/Wnt subtypes of medulloblastoma [163]. Deeper functional characterisation of *FOXG1* may be critical to understand the aberrant GNS transcriptional network and its developmental context. Another transcription factor highly expressed in glioma is AP-2$\alpha$ (*TFAP2A*, adj. p = $1.29 \times 10^{-05}$). AP-2$\alpha$ was associated with low grade gliomas and expression was found to be reduced with higher grade [95]. Later work found that AP-2$\alpha$ attenuates expression of anti-apoptotic and pro-angiogenic genes suggesting a tumour suppressive role for this transcription factor [20]. More recently, hypermethylation of the AP-2$\alpha$ promotor was also associated with poor survival [235]. *RNF114* encodes a RING domain E3 ubiquitin ligase that can be introduced by interferons and dsRNA [12] (adj. p = $7.20 \times 10^{-06}$). A functional role for *RNF114*, also known as *ZNF313*, was found to be regulation of the G1-S phase transition of the cell cycle through degredation of WAF1 and destabilisation of KIP1 and KIP2 in tumour cells [92]. There is also some evidence that *RNF114* can regulate the NF-$\kappa$B pathway via stabilisation of the inhibitor A20 [222]. Heparan sulfates have a complex and critical role in development, homeostasis and disease acting as signalling cofactors and growth factor sinks [124]. Heparinase (*HPSE*, adj. p = $9.29 \times 10^{-05}$) is an endoglycosidase that cleaves heparan sulfate proteoglycans in the remodelling and degradation of the extracellular matrix [290]. Heparinase expression has been shown to increase cell infiltration and decrease proliferation as well as promote adhesive monolayer growth compared to large cellular aggregates [289]. Similarly in medulloblastoma cell lines *HPSE* expression was associated with *in vivo* infiltration [165]. *NKX2-2* is a key transcription factor for the development of oligodendrocyte precursor cells alongside *OLIG2* [67] which are both overexpressed in GNS lines (adj. p = $2.15 \times 10^{-06}$ and $5.55 \times 10^{-03}$ respectively). High expression of *NKX2-2* is also associated with oligodendroglial and astrocytic tumours suggesting a key role for lineage specific transcription factors in the development of the disease [221]. Overexpression of *NKX2-2* in NS cells was found to promote an oligodendrocyte precursor cell fate [275]. The role of *NKX2-2* in glioma is however relatively poorly understood. Another oligodendrocyte regulatory factor, *APCDD1* is identified as over expressed in GNS lines (adj. p = $8.42 \times 10^{-06}$). Regulation of differentiation by *APCDD1* was identified in different glial lineages with regulation via either the canonical Wnt pathway in oligodendrocytes or the non-canonical planar cell polarity Wnt pathway in astrocytes [142]. *LMO4* (adj. p = $2.87 \times 10^{-06}$) was identified as a cofactor of Snail2 (*SNAI2*, adj. p = $1.04 \times 10^{-04}$), which is also overexpressed in GNS, in cadherin repression and an epithelial to mesenchymal transition in both neural crest and neuroblastoma cells [60]. Disregulation of EGF receptor pathway either by *EGFR* mutation or over expression of TGF-$\alpha$ (*TGFA*, adj. p = $3.85 \times 10^{-06}$), one of EGFR's ligands, is a commonly found feature of glioma. Mature astrocytes exposed to TGF-$\alpha$ treatment proliferate and dedifferentiate into neural stem like cells via a neural progenitor like state [232]. The functional role of TGF-$\alpha$ in GNS cells is less well understood however it may well reflect the normal neural functioning and oncogenically act to promote proliferation and block differentiation. Mutations within the *OGFR* gene (adj. p = $9.17 \times 10^{-06}$) have been previously described potentially modulating the replication inhibitory activity of the

OGF/OGFr axis [130]. Activation of the receptor OGFR reduced proliferation of cultured astrocytes in a reversible and dose-dependent manner [27]. Although there is no published work describing the role of the OGF/OGFr axis in glioma, other cancers including pancreatic, breast and ovarian carcinomas have presented growth suppression on OGFr activation leading to a number of clinical trials [51, 286, 287]. With future work the OGF/OGFr axis may become a viable therapeutic target in glioma.

### Genes overexpressed in NS

A total of 1016 genes were found to be overexpressed in NS cells relative to GNS (Table 7.6). Here the functional roles of some of these genes is discussed. Overexpressed in NS relative to GNS are the importins *RANBP17* (adj. $p = 1.11 \times 10^{-22}$), *KPNA3* (IPOA4, adj. $p = 4.24 \times 10^{-05}$) and *IPO5* (RANBP5, adj. $p = 2.86 \times 10^{-06}$). Importins regulate the transport of macromolecules into the nucleus, activated by the GTPase Ran [250]. Different importins are know to transport different sets of cargo molecules, including transcription factors, suggesting differential expression of importins may lead to differential transport of regulatory factors to the nucleus [33]. Importins have recently been given a functional role in mitotic interphase [62]. Further investigation of importin differential expression may reveal critical regulatory differences between these cell types. The NS associated overexpression of tumour suppressor *EPB41L3* (adj. $p = 2.98 \times 10^{-09}$), a member of the protein 4.1 family, reinforces recent work describing it as prognostic biomarker in diffuse gliomas where *EPB41L3* was found to be hypermethylated in tumours [193]. Members of the protein 4.1 family are largely thought to act as adaptor proteins linking the cytoskeleton to the plasma membrane, however this family also has roles in signal transduction interacting with CD44, integrins and the PRMT family [278]. Protein family 4.1 members have also been shown to promote apoptosis and reduce cell motility and proliferation illustrating its role as a tumour suppressor. Little work has been done to describe the functional significance of this gene family in glioma. The gene *CELSR1* (adj. $p = 1.91 \times 10^{-07}$) encodes a receptor that has complex roles in neural planar cell polarity pathways [59]. Yet another receptor, *RGMB* (adj. $p = 1.32 \times 10^{-07}$), is implicated in neural development through BMP signalling [267].

NS overexpression, or rather GNS underexpression of several other tumour suppressor associated genes is identified here. Among this list of tumour suppressors perhaps the best known is *RB1* (adj. $p = 3.45 \times 10^{-05}$) which has been the focus of extensive research since first discovered [49]. Disrupted regulation of PTEN (adj. $p = 7.11 \times 10^{-4}$) and IGF signalling via the ubiquitin ligase *NEDD4* (adj. $p = 2.39 \times 10^{-06}$) has been identified in a number of cancer types [13]. *NEDD4* has also been shown to have dual oncogene/tumour suppressor function via either enhancing nuclear import or cytoplasmic degradation of PTEN [264]. Ras suppressor-1 (*RSU1*, adj. $p = 5.46 \times 10^{-08}$) was initially identified as a suppressor of Ras dependent oncogenesis [46] and in glioma deleterious mutation of *RSU1* is relatively common [40]. A apoptotic promoting role for *RSU1* and its interaction with the pro-survival adhesion protein PINCH-1 was also found [77]. Another established glioma tumour suppressor gene found over expressed in NS cells is the protein tyrosine phosphatase *PTPRD* (adj. $p = 6.86 \times 10^{-04}$) [241]. Heterozygous loss of PTPRD in human glioblastoma was demonstrated to promote tumourigenesis via Stat3, alongside *CDKN2A* deletion [189]. Mutations in the heparin sulphate glycosyltransferase *EXT1* (adj. $p = 5.92 \times 10^{-06}$) are frequent occurrences in different tumour types which may suggest tumour suppressor like activity [26]. The transcription factor *TCF7L2* (adj. $p = 1.732 \times 10^{-06}$) is a downstream effector of the canonical Wnt signalling pathway with critical roles in oligodendrocyte differentiation [88]. Reduced expression of *TFC7L2* in GNS cells may function to suppress GNS cell differentiation in response to Wnt signalling. Further investigation of this gene panel, while beyond the scope of this thesis, may identify further candidate glioma tumour suppressors or oncogenes.

### Analysis of differentially expressed genes in glioma expression data

As GNS cell lines are expanded in EGF/FGF culture conditions it is important to ensure these cells are representative of glioma stem cells *in vivo*. One way of testing this is to examine genes that are differentially expressed between GNS and NS lines in tumour derived expression data. Combining GNS and NS overexpressed genes into mean z-score values that represent the average expression of each gene set reveals an anti-correlated expression (Pearson's r = -0.78, $p < 2.2 \times 10{-}16$, Figure 4.4). This relationship is largely irrelevant of molecular subtype with all types dispersed across the axis. Ensure the GNS and NS gene sets are representing a real coexpressed relationship the correlation of each gene in the respective set was correlated to the component module mean z-score. Both GNS and NS gene sets were found to be more correlated to their mean z-score in comparison to random gene sets suggesting the relationship identified in the GNS/NS dataset is at least partially replicated in the glioma dataset (Figure 4.4). Restricting the GNS and NS modules to include only genes with high expression fold change improves the performance of these gene sets in the glioma dataset.

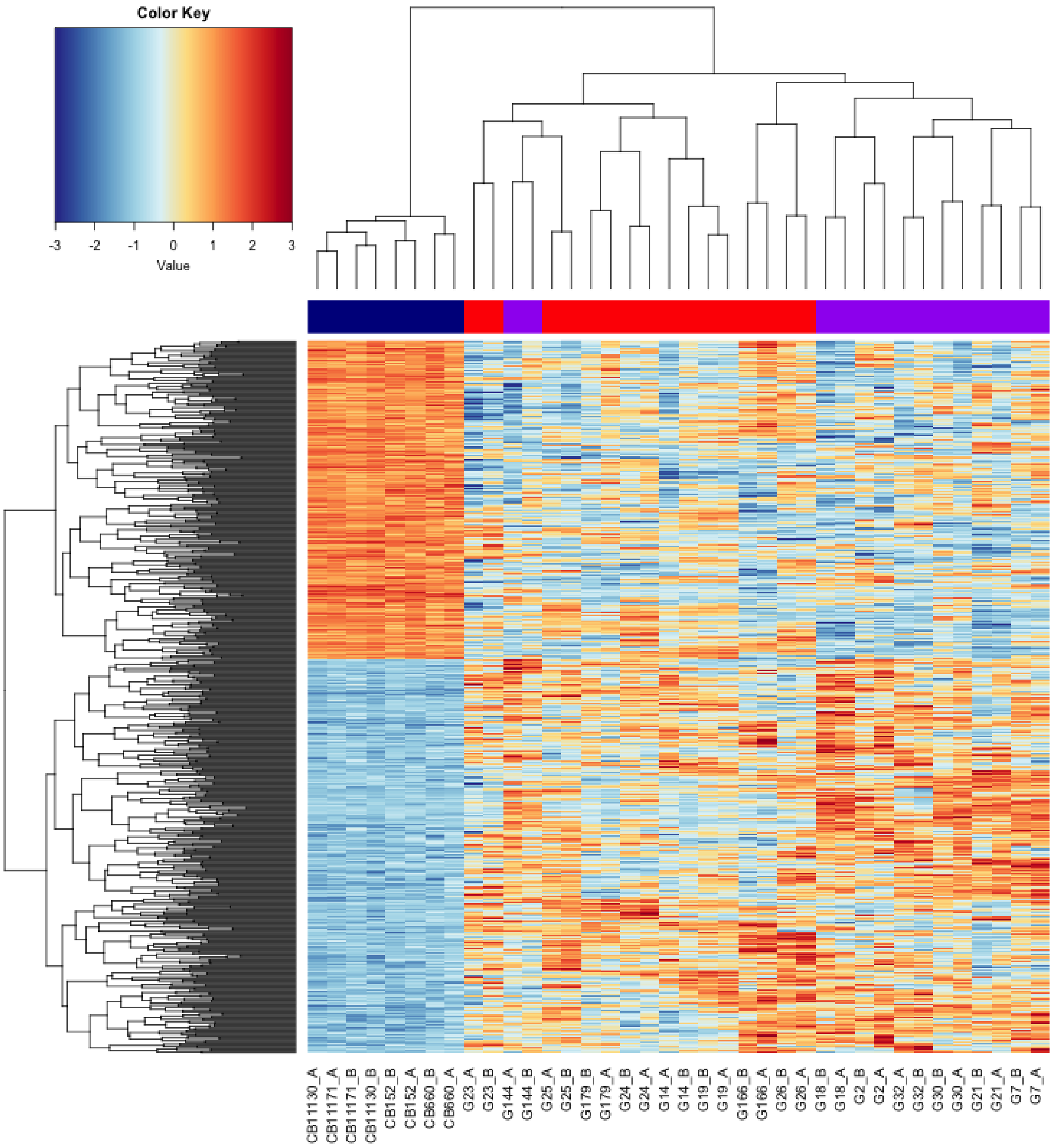


Fig. 4.3 Heatmap illustrating genes differentially expressed between NS and GNS. Genes over and underexpressed in NS cells (Tables 7.6 and 7.7) show relative consistency compared to GNS lines. Colours within the heatmap are representations of row mean normalised, log2 intensity units. Column colour makers represent the cell type including NS (blue) and GNS subtypes (Discussed in Figure 4.9, red = mesenchymal, purple = proneural).

| GNS | | | NS | | |
|---|---|---|---|---|---|
| Gene name | Log2 fold change | Adj. p-value | Gene name | Log2 fold change | Adj. p-value |
| FOXG1 | -2.75 | 1.72e-13 | RANBP17 | 2.53 | 1.11e-22 |
| NRN1 | -3.90 | 2.59e-10 | TES | 3.26 | 2.03e-14 |
| PCDHB9 | -2.50 | 5.42e-08 | RASGRF2 | 2.84 | 4.16e-14 |
| TTC39C | -2.04 | 5.97e-08 | EPHA7 | 5.03 | 4.16e-14 |
| ZFAND2A | -1.22 | 9.27e-08 | OTX2 | 4.20 | 1.01e-13 |
| LDHA | -1.90 | 2.98e-07 | TNFRSF10D | 3.09 | 4.84e-13 |
| KLHL13 | -1.78 | 5.52e-07 | CDCP1 | 3.71 | 1.55e-12 |
| FAM102A | -0.87 | 5.93e-07 | AFF2 | 1.70 | 2.92e-12 |
| DYNLL2 | -0.79 | 5.93e-07 | SYT1 | 4.37 | 1.37e-11 |
| MT2A | -1.57 | 6.08e-07 | ANO4 | 4.27 | 1.95e-11 |
| TNFRSF21 | -2.73 | 1.02e-06 | AK7 | 1.09 | 5.04e-11 |
| PMS2P3 | -0.67 | 1.28e-06 | BTBD11 | 3.45 | 1.03e-10 |
| NUDCD3 | -0.74 | 1.55e-06 | MCHR1 | 2.73 | 2.28e-10 |
| MTG2 | -0.66 | 1.58e-06 | CRHBP | 2.26 | 2.47e-10 |
| ADGRE5 | -1.27 | 1.68e-06 | GREB1L | 2.70 | 6.73e-10 |
| THY1 | -3.37 | 1.79e-06 | IGF2BP1 | 1.90 | 6.73e-10 |
| CD9 | -2.72 | 2.02e-06 | NELL2 | 4.74 | 9.44e-10 |
| NKX2-2 | -2.38 | 2.15e-06 | PBX3 | 2.37 | 9.44e-10 |
| LMO4 | -1.86 | 2.87e-06 | NEFM | 1.93 | 1.88e-09 |
| MT1L | -1.48 | 3.02e-06 | EPB41L3 | 2.63 | 2.98e-09 |
| WDR91 | -0.66 | 3.02e-06 | MGST1 | 3.55 | 3.18e-09 |
| C12orf66 | -0.80 | 3.02e-06 | NEGR1 | 2.24 | 7.97e-09 |
| SHOX2 | -0.84 | 3.25e-06 | LRRC7 | 2.44 | 8.47e-09 |
| TGFA | -2.66 | 3.85e-06 | WBSCR17 | 3.62 | 9.58e-09 |
| FAM122C | -0.88 | 5.09e-06 | GRPR | 3.12 | 1.34e-08 |
| ADAMTS9 | -2.34 | 6.20e-06 | RSU1 | 1.30 | 5.47e-08 |
| RNF114 | -0.96 | 7.20e-06 | RAB11FIP1 | 1.18 | 6.16e-08 |
| APCDD1 | -2.22 | 8.42e-06 | REC8 | 1.44 | 6.73e-08 |
| OGFR | -0.47 | 9.17e-06 | NECAB1 | 2.88 | 1.32e-07 |
| MR1 | -1.98 | 1.07e-05 | RGMB | 1.56 | 1.32e-07 |
| BCAM | -0.93 | 1.11e-05 | HS3ST3A1 | 1.49 | 1.49e-07 |
| HSF2BP | -0.66 | 1.13e-05 | NXN | 1.31 | 1.59e-07 |
| TFAP2A | -2.05 | 1.29e-05 | CELSR1 | 1.91 | 1.91e-07 |
| WBSCR22 | -0.72 | 1.43e-05 | SLC18A3 | 3.22 | 2.53e-07 |
| CNTNAP3 | -2.16 | 1.43e-05 | OCA2 | 1.20 | 2.83e-07 |
| GCC1 | -0.69 | 1.56e-05 | DAPK1 | 2.58 | 3.04e-07 |
| PPM1K | -1.83 | 1.76e-05 | MYO1B | 3.96 | 3.40e-07 |
| QSOX2 | -0.74 | 1.93e-05 | TLE4 | 2.05 | 3.59e-07 |
| MT1G | -0.97 | 2.52e-05 | DOCK2 | 1.58 | 3.80e-07 |
| MT1H | -1.44 | 2.52e-05 | OXTR | 1.51 | 6.08e-07 |

Table 4.1 Table of the top 40, most differentially expressed genes between NS and GNS ordered by p-value (Full tables 7.6 and 7.7)

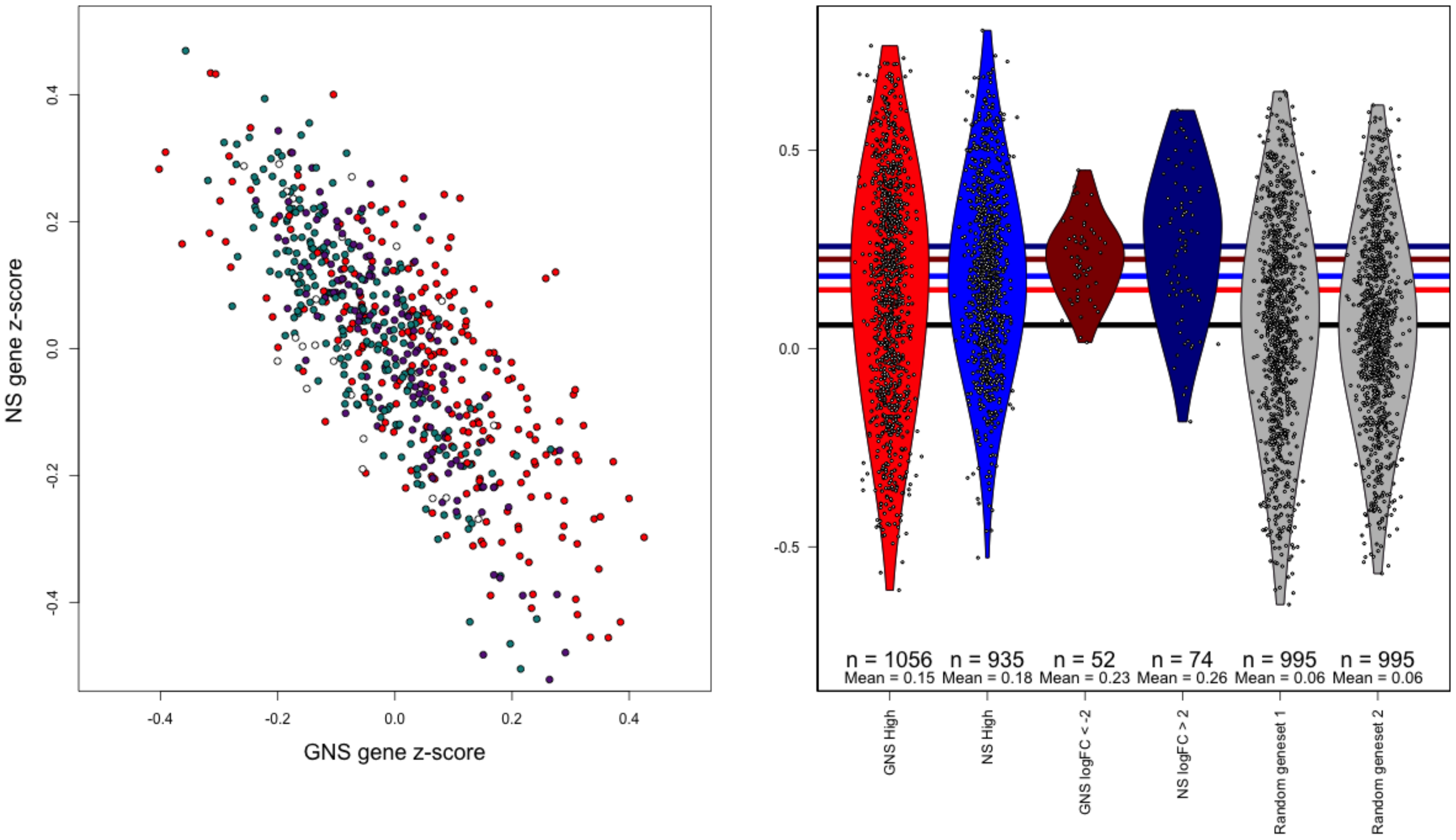
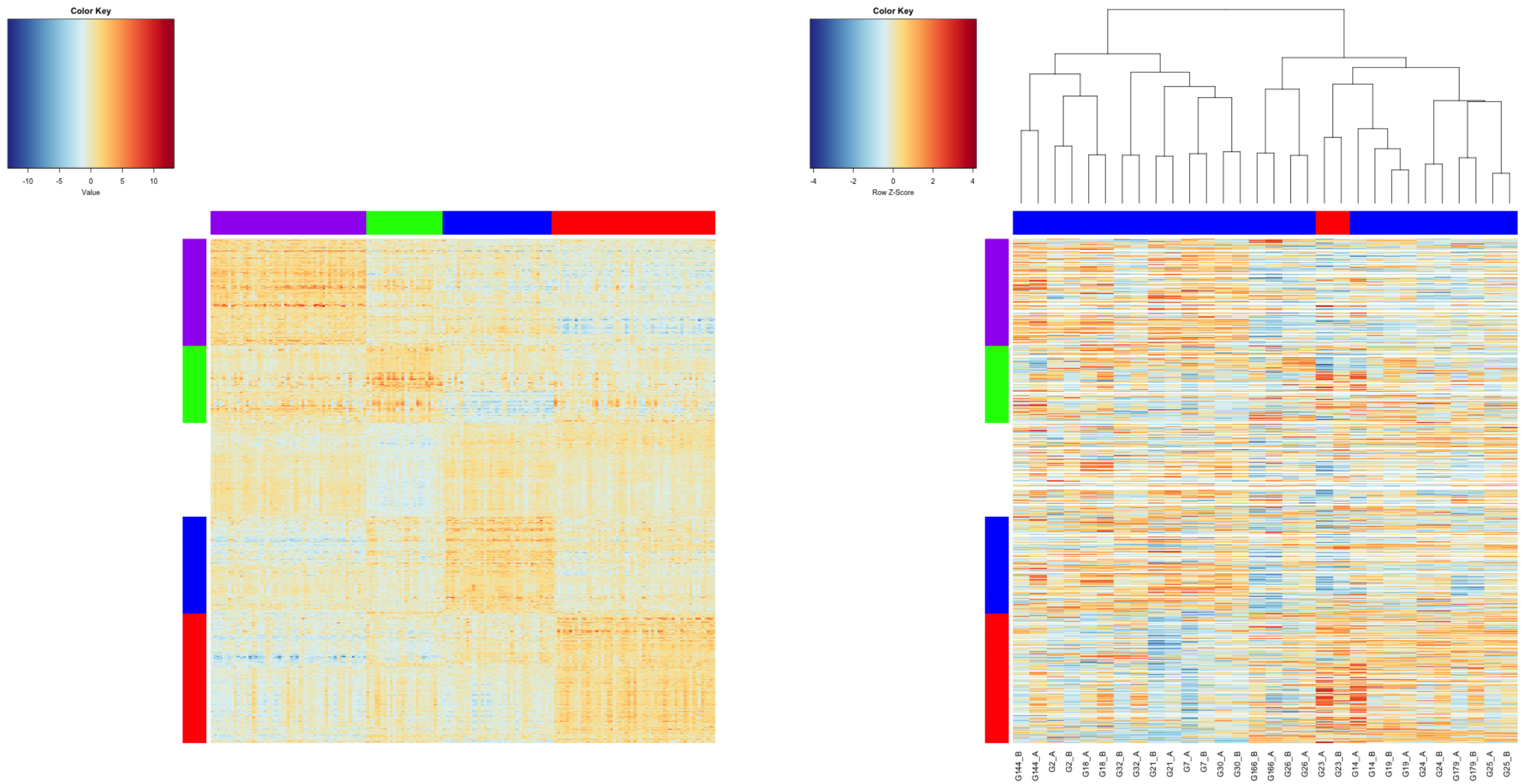


Fig. 4.4 Coexpression of GNS versus NS differentially expressed genes in glioma. Z-score expression for GNS and NS differentially expressed genes are anti-correlated (Pearson's r = -0.78, $p < 2.2 \times 10-16$) to each other with molecular subtypes dispersed across the full range of expression (Left panel). Violin plots showing correlation of each gene to the module from which it was derived (Right Panel). GNS (red violin) and NS (blue violin) differentially expressed genes are more correlated to each other than a random gene set (grey violin) indicating true coexpression. Furthermore restricting the gene sets to those with a high fold change increases this correlation. Horizontal coloured lines indicate the mean values for each gene set.

### 4.2.2 Verhaak *et al.* subtype centroid classification of GNS lines

Glioma tumours have been described as consisting of four distinct subtypes based on a clustering of 202 glioblastoma multiforme samples [270]. All GNS lines investigated were determined to be IDH wild type through targeted sequencing [55] suggesting these lines reflect the IDH wildtype molecular subtype of glioma [53]. Tumour samples can be divided into these subtypes using a centroid based classifier [82]. This classifier is based on calculating the sample correlation to 840 gene centroids for each subtype (Figure 4.5). Classification of GNS lines using this tumour data derived classifier labels a majority of GNS lines as classical with the exception of G23 which was more highly correlated to the mesenchymal centroids (Figure 4.5). This is in contrast to tumour subtype analysis set out in Chapter 3 which indicates no evidence of a classical subtype. Although this is the strictly correct way to apply this classifier, differences in expression profile and cell type content between glioma tumours and the tumour derived GNS cells leads to niche specific correlation biases therefore affecting classification.

While many genes in the Verhaak centroids are also highly expressed in GNS cells the correlation of mean expression between GNS and glioma data is low (Pearson's r = 0.53, Figure 4.6). With variation in average gene expression between GNS and glioma data correlations to classifier centroids does not produce equivalent results. This may be due to differences in cellular composition where tumours are composed of a mixed population of cells such as macrophages, fibroblasts and the glioma cells themselves that each contribute proportionally to the subtype signature. This resulting mixed cell signature is then compared to a homogenous clonal cell line leading to distinct differences in expression. Other components of the tumour niche like hypoxia and rate of proliferation may also have an effect on the comparason between cell line and tumour sample. As such the correlation for each GNS line to its closest subtype centroid is relatively low compared to glioma data classification (Mean Pearson's r, GNS = 0.07, glioma = 0.54) suggesting limited utility in the centroid correlations's predictive value.



(a)

(b)

Fig. 4.5 Verhaak *et al.* centroid classification of glioma and GNS data. Replication of centroid classification of glioma data described by Verhaak *et al.* [270] (A) alongside classification of GNS lines using the same method (B). Here GNS lines are mainly classified as classical with the exception of G23 which are closer to the mesenchymal centroids. GNS lines show poor correlation to their respective closest subtype centroids compared to glioma data (Mean Pearson's r, GNS = 0.07, glioma = 0.54). Units are row mean normalised log2 FPKM (a) and row mean normalised log2 intensity units (b).

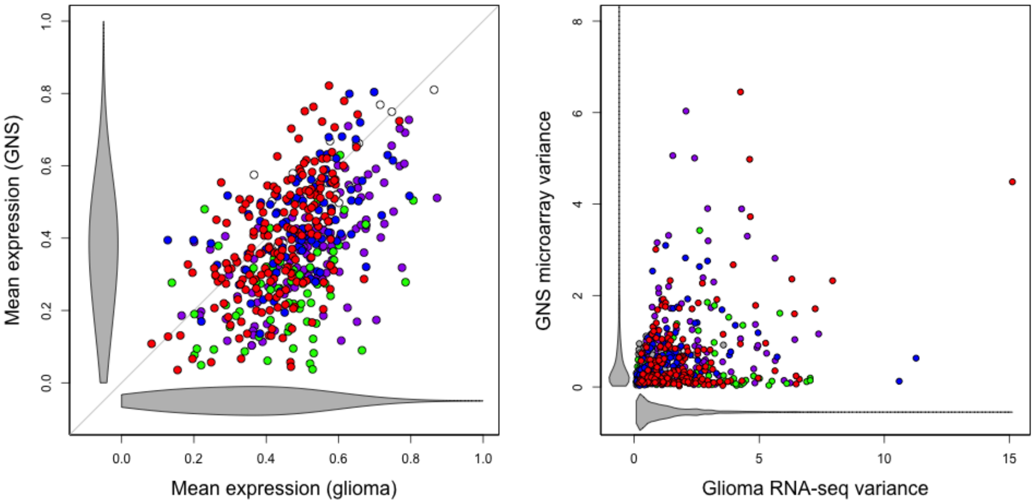
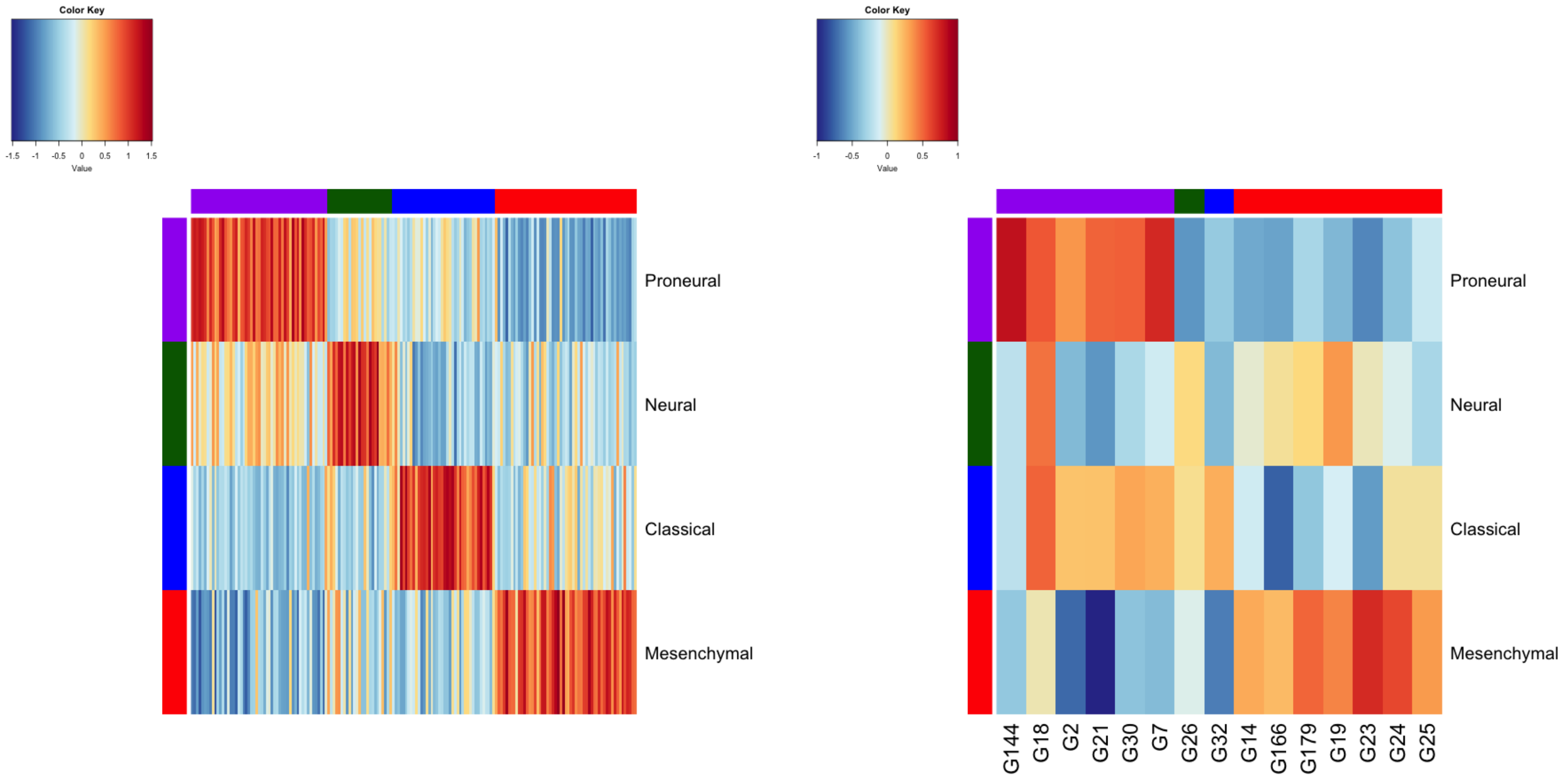


Fig. 4.6 Verhaak *et al.* centroid genes in glioma (Log2 FPKM) and GNS data (Log2 intensity) showing both mean expression (Left panel) and variation in expression across each dataset (Right panel). Differences in cellular composition and environmental factors between glioma tumour and GNS expression produces an inconsistent correlation between conditions. This implies that classification methods that rely on relative expression ranking, like centroid-based methods, will produce non-representative results when applied outside of a tumour context. Colored dots reflects the subtype to which each gene belongs and grey 'violin' shapes illustrate the average density of all genes in each dataset. Many subtype marker genes with high variance in glioma have low variance in GNS lines and vice versa implying these low variance markers are of reduced predictive value in GNS.

An alternative to centroid classification would be to compare the relative expression of centroid genes and determine the subtype by high expression of marker genes compared to other GNS lines. Samples that are representative of a distinct subtype should present the highest levels of a marker gene compared to non-subtype samples. Dependent on the assumption that all subtypes are present in this panel of GNS lines, subtypes could be distinguished based on summarised relative gene expression of subtype marker genes. This allows for the differences in within-group variation and average gene expression between glioma and GNS data. The centroid genes identified by Verhaak *et al.* overexpressed in each subtype were converted into z-scores and the highest mean z-score of each subtype's marker genes was used to classify samples as the nearest subtype. Using this method on the data used by Verhaak *et al.* to construct their centroid classifier resulted in 96.6% reclassification accuracy (Figure 4.7a). Applying this method to the GNS expression data classifies seven lines as mesenchymal, six as proneural, alongside single classical and neural lines (Figure4.7b, Total fifteen lines). In contrast to the centroid classifier this relative expression based method identifies relatively few GNS lines as being closest to the classical subtype. In summary, classification of GNS lines into the four subtypes defined by Verhaak *et al.* is dependent on method used and the different assumptions they depend on. As GNS cells and tumour samples present considerable differences in average expression and variation, a tumour based classifier could be considered a poor method of characterising these lines and alternative methods should be identified.



(a) (b)

Fig. 4.7 Relative expression based classification of GNS data to Verhaak *et al.* subtypes in glioma (A) and GNS data (B). Glioma samples used to define the Verhaak *et al.* subtypes are reclassified using relative expression. Application of this method to GNS data classifies samples as either mesenchymal (G14, G166, G179, G19, G23, G24 and G25) or proneural (G144, G18, G30, G2 and G21), a single classical like line (G32) and one neural line (G26). Column marker colours indicate the subtype into which samples has been classified (highest overall marker signature) and the row colour markers indicate subtype marker gene expression. Units displayed in the heatmaps are the sample average of row mean normalised expression values for subtype marker genes, generated from log2 FPKM (a) and log2 intensity units (b).

### 4.2.3 Expression modules identified in GNS lines correspond to glioma derived proneural and mesenchymal expression modules

To characterise gene expression variation across these GNS lines, CMC coexpression clustering was employed as previously described in Chapter 2 for examining tumour expression datasets. A total of 96 CMC modules were found and from this set modules of interest were prioritised based on a selection of different criteria. Associations between these GNS modules and glioma proneural and mesenchymal CMC modules described in Chapter 2 were found by identifying GNS modules with intersecting gene sets and a positive correlation to the respective glioma module ($\geq$ 3 shared genes). For the glioma proneural module the largest intersection to a GNS module was 20 genes (GNS module 2, Table 4.3). For the glioma mesenchymal module a GNS module was identified that shares 15 genes (GNS module 1, Table 4.3). These GNS proneural and mesenchymal associated modules are amongst the largest and most variable found in the GNS panel (Figure 4.8).

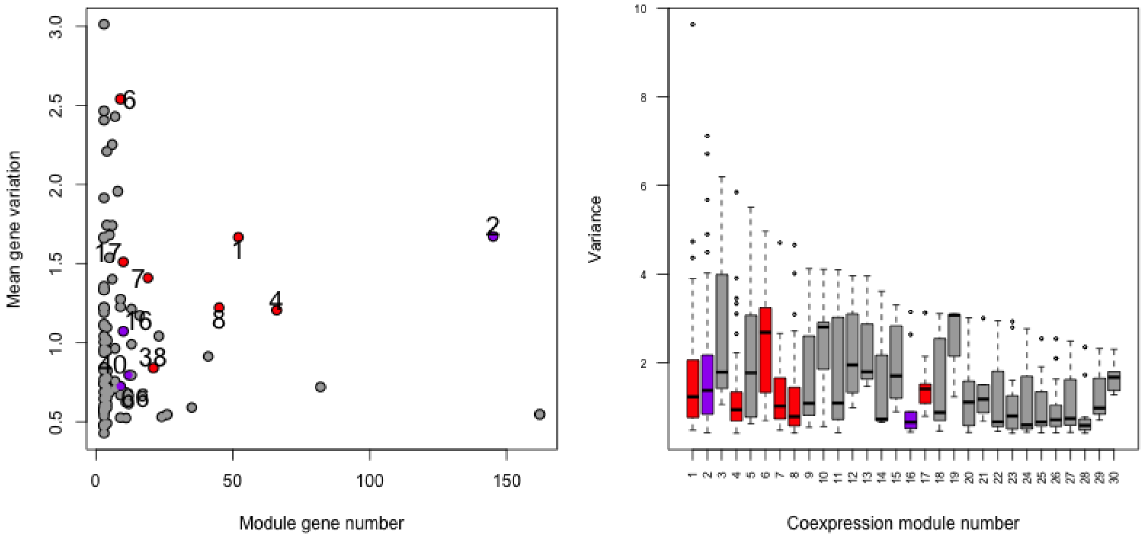


Fig. 4.8 Variance and gene number of coexpression modules. Modules with large mean variation and many component genes, tend to have intersecting gene sets with glioma proneural (Purple) or mesenchymal modules (Red). The highest variance genes are found within the first two coexpression modules that also intersect the highest number of glioma proneural or mesenchymal modules. Module variance is based on log2 microarray intensity units.

These two proneural and mesenchymal associated, highly variable modules separate the GNS samples into two distinct clusters (GNS modules 1 and 2, Figure 4.9). This separation of GNS lines is also mirrored in the first three principal components of the data (Figure 4.10). The biological replicates for each line also cluster closely to each other indicating these signatures are a true characteristic of each independently derived line rather than experimental variation. As these clusters become apparent in the GNS modules most representative of the glioma proneural and mesenchymal modules, these distinct clusters were labeled as proneural and mesenchymal GNS subtypes. The separation of proneural and mesenchymal GNS lines is somewhat mirrored by the centroid-based subtype classification with the exceptions of G32 (CMC: proneural, Centroid: classical) and G26 (CMC: mesenchymal, Centroid: neural) (Figure 4.7b).

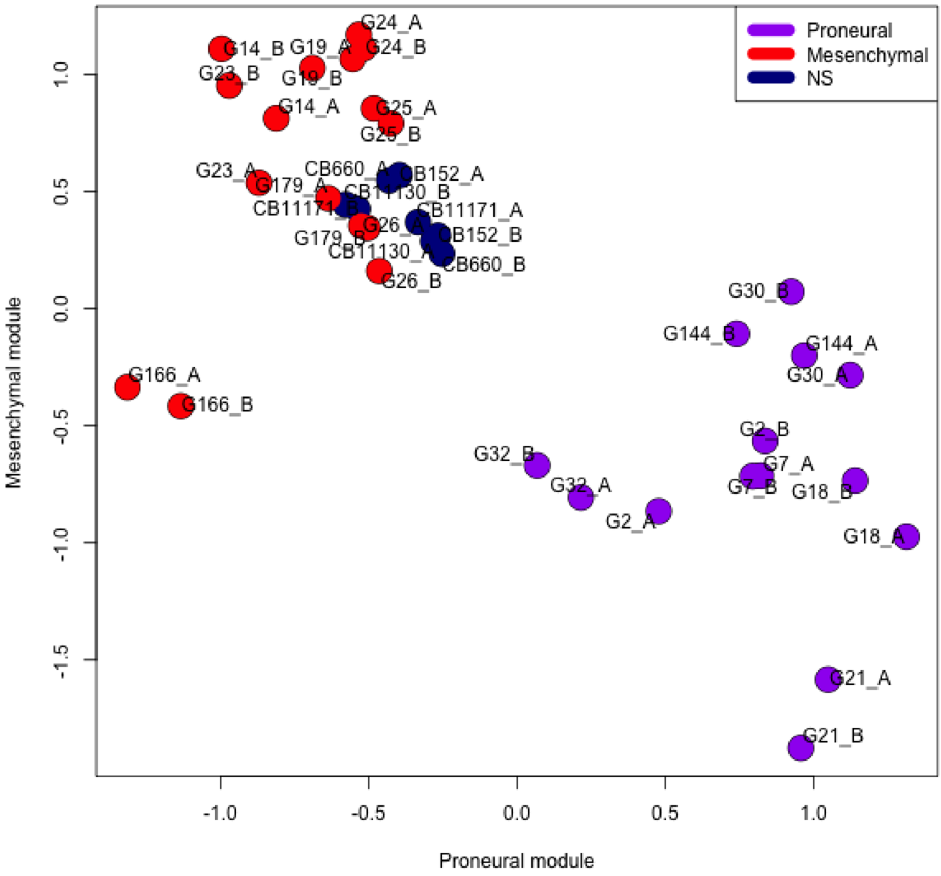


Fig. 4.9 Proneural and mesenchymal CMC modules separate GNS cells into two clusters. NS lines tend to cluster more closely with mesenchymal GNS lines with average expression of these genes. Units are z-score normalised log2 intensity units.

Beyond the primary GNS proneural and mesenchymal modules several other GNS modules are enriched for glioma proneural and mesenchymal genes. Visualisation of these modules reveals some consistency between independent cell lines yet also clear variation in

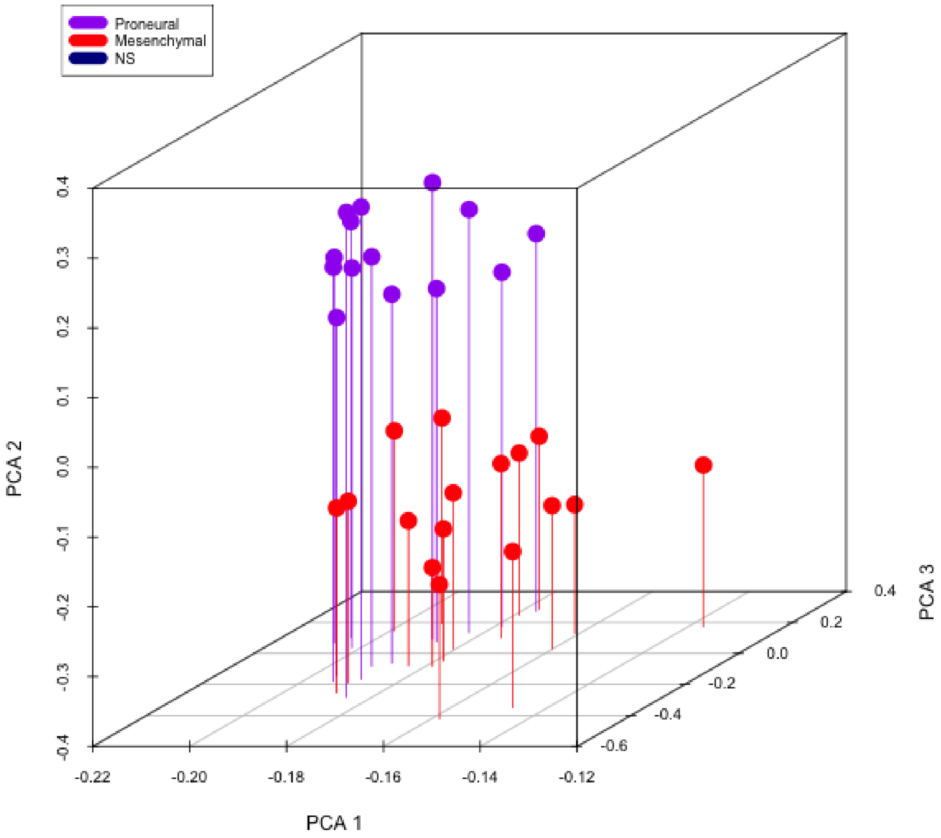


Fig. 4.10 Principal component analysis of GNS array data (log2 intensity units) showing distinctively separated proneural (Purple) and mesenchymal (Red) clusters.

other modules (Figure 4.11). Three additional proneural modules reveal cell line specific proneural gene expression with G144 showing high expression of all 4 proneural associated modules. For mesenchymal associated modules there is no GNS line that shows consistently high mesenchymal signature in all mesenchymal like modules in comparison to the consistent high expression of the GNS proneural modules in G144. Interestingly NS lines tend to cluster more closely to the mesenchymal GNS lines than to the proneural GNS lines. While expression of these genes in NS lines is inconsistently correlated, these karyotypically normal cell lines tend to express mesenchymal associated genes more highly than proneural associated genes. Examining the consensus genes shared between the glioma and major GNS proneural module (GNS module 1) reveals many genes consistently associated with the established proneural phenotype in glioma (Table 4.3). The well characterised *OLIG2* and *OLIG1* transcription factors along with *SOX6* are found within this proneural module and may be considered the best candidates for the root of a proneural transcriptional network. The Notch ligands *DLL3*, *DLL1* and *BCAN* are also established cell surface markers for the proneural phenotype. The consensus shared mesenchymal genes shared between glioma and GNS mesenchymal modules (GNS module 1) include *CCL2*, *LOXL1*, *PTRF*, *SERPINE1* and *THBS1*.

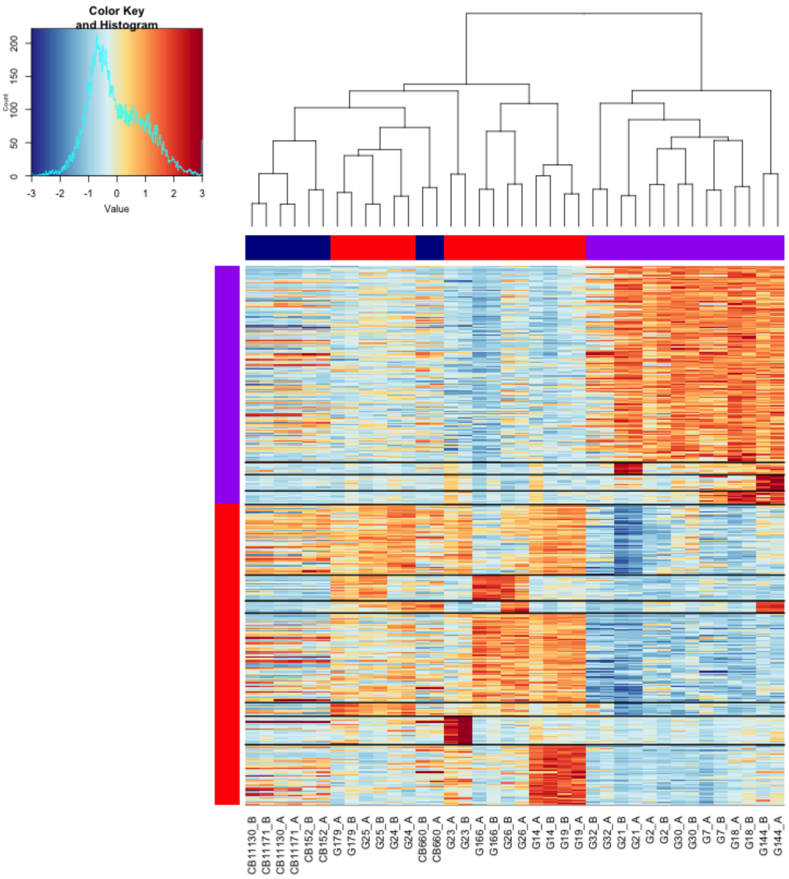


Fig. 4.11 Heatmap showing expression of proneural and mesenchymal modules in GNS and NS data. NS lines tend to cluster with mesenchymal GNS lines with inconsistant expression of GNS derived modules. Proneural GNS lines cluster separately from GNS mesenchymal and NS cell lines. Colours within the heatmap are representations of row mean normalised, log2 intensity units. Column colour makers represent the cell type including NS (blue) and GNS subtypes (Discussed in Figure 4.9, red = mesenchymal, purple = proneural). Row colour markers indicate the inclusion of Verhaak *et al.* proneural and mesenchymal marker genes within each CMC module.

| Mesenchymal lines | | | Proneural lines | | |
|---|---|---|---|---|---|
| Gene name | Log2 fold change | Adj. p-value | Gene name | Log2 fold change | Adj. p-value |
| LAYN | -3.29 | 3.43e-10 | GRIK3 | 4.16 | 2.16e-11 |
| MET | -3.46 | 5.98e-10 | SEZ6L | 3.92 | 2.41e-11 |
| DKK1 | -4.62 | 5.19e-09 | ASCL1 | 4.84 | 8.84e-11 |
| CASP4 | -2.47 | 3.22e-08 | NLGN3 | 2.89 | 2.44e-10 |
| BAG2 | -1.70 | 1.66e-07 | ATCAY | 2.93 | 2.61e-10 |
| CARD16 | -3.48 | 2.00e-07 | ADCYAP1R1 | 3.66 | 1.07e-09 |
| ALPK1 | -1.72 | 3.15e-07 | DLL3 | 3.23 | 4.51e-09 |
| FAM65B | -2.09 | 3.24e-07 | LRRTM3 | 2.17 | 4.72e-09 |
| ARHGAP29 | -3.13 | 9.09e-07 | STMN4 | 2.16 | 5.19e-09 |
| RAB27A | -1.68 | 9.58e-07 | SCG3 | 3.41 | 8.64e-09 |
| CALCRL | -3.68 | 5.81e-06 | WSCD1 | 3.36 | 8.73e-09 |
| LATS1 | -1.21 | 5.81e-06 | SEPT3 | 2.16 | 8.96e-09 |
| NABP1 | -1.65 | 6.01e-06 | OLIG2 | 1.78 | 1.74e-08 |
| ECE1 | -1.54 | 6.59e-06 | LHFPL3 | 3.17 | 2.04e-08 |
| TLR4 | -2.31 | 6.60e-06 | C1orf61 | 4.23 | 3.22e-08 |
| MDK | -1.69 | 8.61e-06 | ELAVL3 | 2.79 | 4.13e-08 |
| B3GALNT1 | -1.73 | 8.76e-06 | NKAIN4 | 2.09 | 6.89e-08 |
| GBP1 | -2.14 | 8.90e-06 | CSPG5 | 1.30 | 7.39e-08 |
| DCBLD2 | -1.87 | 9.17e-06 | SMOC1 | 3.17 | 9.23e-08 |
| APOBEC3F | -1.03 | 9.50e-06 | MEGF10 | 2.15 | 1.06e-07 |
| AMIGO2 | -1.64 | 9.62e-06 | BMP7 | 2.75 | 1.12e-07 |
| FAM188B | -0.97 | 1.03e-05 | ZDHHC22 | 3.24 | 1.83e-07 |
| FBLN1 | -2.30 | 1.06e-05 | ADGRB1 | 1.42 | 2.47e-07 |
| PDCD1LG2 | -1.85 | 1.09e-05 | DCX | 3.33 | 2.92e-07 |
| PAK1 | -1.82 | 1.15e-05 | PREX1 | 2.23 | 3.15e-07 |
| NT5DC3 | -2.27 | 1.25e-05 | FAM131B | 2.09 | 3.15e-07 |
| TRIM34 | -0.71 | 1.26e-05 | MOB3B | 2.12 | 3.16e-07 |
| CAPG | -1.11 | 1.27e-05 | HEPN1 | 2.79 | 3.41e-07 |
| ARHGAP18 | -1.92 | 1.45e-05 | TRIM9 | 2.32 | 5.40e-07 |
| CCL2 | -3.59 | 1.45e-05 | HEY2 | 2.32 | 7.06e-07 |
| MID2 | -2.18 | 1.48e-05 | MTSS1 | 2.26 | 8.12e-07 |
| PALLD | -1.70 | 1.53e-05 | BCAN | 3.57 | 8.54e-07 |
| GPRASP2 | -1.19 | 1.59e-05 | GAD1 | 2.61 | 8.54e-07 |
| CAV1 | -2.62 | 1.96e-05 | KCNA2 | 1.95 | 9.02e-07 |
| NRP1 | -2.27 | 1.96e-05 | CACNG7 | 1.95 | 9.49e-07 |
| COPZ2 | -1.60 | 1.96e-05 | GPR19 | 2.94 | 1.01e-06 |
| GULP1 | -1.77 | 1.96e-05 | EPHB1 | 2.78 | 1.27e-06 |
| DDO | -0.91 | 1.98e-05 | SOX6 | 2.47 | 1.66e-06 |
| UBA7 | -1.19 | 2.16e-05 | SHC3 | 2.33 | 1.92e-06 |
| TRIM21 | -0.82 | 2.16e-05 | PTP4A3 | 1.48 | 2.39e-06 |

Table 4.2 Table of top 40 differentially expressed genes between proneural and mesenchymal GNS lines ordered by p-value (Full tables 7.6 and 7.7)

| Consensus proneural | Consensus mesenchymal |
|---|---|
| ACSL6 | ABCC3 |
| ATCAY | ANXA2 |
| BCAN | C1R |
| DLL1 | C1RL |
| DLL3 | CCL2 |
| GNAO1 | CFI |
| GRIA2 | COL6A2 |
| MAP2 | GLIPR1 |
| MYT1 | HFE |
| NCAM1 | LOXL1 |
| OLIG1 | MYOF |
| OLIG2 | PTRF |
| PHYHIPL | SERPINE1 |
| RUNDC3A | THBS1 |
| SCG3 | TMBIM1 |
| SEPT3 | |
| SEZ6L | |
| SHD | |
| SOX6 | |
| ZDHHC22 | |

Table 4.3 Consensus proneural and mesenchymal genes found in both Verhaak *et al.* glioma and GNS CMC modules.

### 4.2.4 GNS proneural associated modules

The modules that are associated with the proneural subset of GNS lines show variation between lines and a general enrichment for genes associated with neural and glial development. Comprehensive investigation of the genes that differentiate subsets may help inform on the cellular origin or distinctive epigenetic state of glioma stem cells. Here I examine these genes and modules in detail extracted from the literature (Figure 4.12).

#### The primary GNS proneural module

Examining the core GNS proneural module (GNS module 2) many canonically proneural associated genes are presented. Of these Olig2 is perhaps the most commonly described proneural gene. Olig2 was initially identified as an early expressed transcription factor in the differentiation of oligodendrocyte precursor cells [294]. Following this discovery Olig2 was found to be relevant to glioma with high expression in oligodendrogliomas and low expression in astrocytomas [149, 161]. With the arrival of cancer subtype methods gliomas were divided into different subsets with proneural Olig2 being a critical component of these signatures [198, 270]. Olig2 is also described as a key transcription factor for the propagation of glioma and furthermore was able to reprogram differentiated cells towards a tumour propagating cell [252]. Another highly variable transcription factor is *ASCL1* (also known as MASH1), a complex master regulator of neurogenesis promoting proliferation and differentiation of neural progenitor cells [268]. The role of *ASCL1* in glioma is not well understood. Forced expression of *ASCL1*, *BRN1* and *NGN2* enabled the differentiation of GSCs into functional neurones [291]. RAS/ERK signalling was found to modulate ASCL1 induced differentiation of NS cells towards either neuronal (RAS/ERK low) or glial progenitors (RAS/ERK high) [148]. Mouse studies found that *C1orf61*, also known as CROC-4 was found to participate in brain specific c-FOS signalling [111]. The glutamate receptors including GRIA2 and GRIK3 have been identified as potential regulators of migration proliferation and differentiation of NS cells [109]. *CHRDL1* was found to antagonise BMP-4 induced astrocyte differentiation of NS cells and instead steer the cells towards a neural lineage [72].

The most variable gene is the well characterised cell cycle regulator cyclin D2 (*CCND2*) and along with the commonly amplified *CDK4* (not included in a CMC proneural module) is a critical regulator of the G1/S cell cycle transition. RNA interference mediate suppression of cyclin D2 caused G1 arrest of GSCs [128]. Cyclin D2 expression was also reduced on serum differentiation. Asymmetrically dividing neural stem cells were shown to transport *CCND2* mRNA to the basal process biasing apical proximal daughter cells towards terminal differentiation during cell division indicating a critical role in NS and GNS self renewal [265].

Brevican (*BCAN*) is a brain specific proteoglycan that forms a component of the neural extra cellular matrix though interactions with tenacin-C, tenacin-R and hyaluronic acid [66]. Importantly brevican expression is limited to histologically glial-like tumours and brevican status has been shown to correlate to infiltrative capability [110]. Xenograft cultures of *in vivo* brevican negative GSCs failed to infiltrate the host brain and produced tumours similar to carcinoma metastasises while some lines while presenting as brevican negative in culture, expressed brevican *in vivo* and grew as infiltrative tumours suggesting that brain niche specific factors enabled this transformation. While these CSCs were cultured in serum containing media, GSC growth in FGF-2 and EGF enables the expansion of brevican positive cells that are capable of infiltrative growth. Brevican shRNA-mediated knockdown did not lead to the reduction of proneural regulators like *ASCL1* or *OLIG2* but did reduce the infiltrative capability of GSCs [52]. Neurocan (*NCAN*) is a proteoglycan, with similar functions to brevican, that forms a component of perineuronal nets (PNNs) [1, 66, 105]. It was suggested that neurocan inhibits neural adhesion [64]. *PTPRZ1* (phosphacan) is a proteoglycan similar to brevican, is a receptor for Midkine (*MDK*) and the expression of both these genes varies thougout the brain [68]. Analysis of *PTPRZ1* knock-out mice suggests this receptor has a negative regulatory role in oligodendrocyte development [131]. Further work suggests that both soluble and membrane bound PTPRZ1 interact with contactin-1 (*CNTN1*), amongst other extracllular matrix proteins, to regulate the proliferation and differentiation of oligodendrocyte progenitor cells [133]. Gene fusions of *PTPRZ1* and the HGF receptor *MET* are also common in glioma [35]. The pituitary adenylyl cyclase-activating peptide receptor *ADCYAP1R1* (Also known as PAC1-R) has a complex relationship to both CNS and systematic inflammation, differentiation and repair [279]. In a reactive astrocyte model *ADCYAP1R1* and *GFAP* were found to be unregulated on injury [180]. Hypoxia acting via HIF-1$\alpha$ activates a PACAP38-PAC1-R signalling cascade to facilitate bone marrow-derived immune cells resulting in reduced injury [152]. Similarly *GAP43* expression was found to be highly upregulated one day after ischemic lesion [81]. Doublecortin (*DCX*) is a microtubule assocciated protein that is crucial for cell movement [188] and has been proposed as a marker for adult migrating neuroblasts [273]. The Shc-like adaptor protein Rai (*SHC3*) was also shown to regulate Doublecortin dependent migration and infiltration [187].

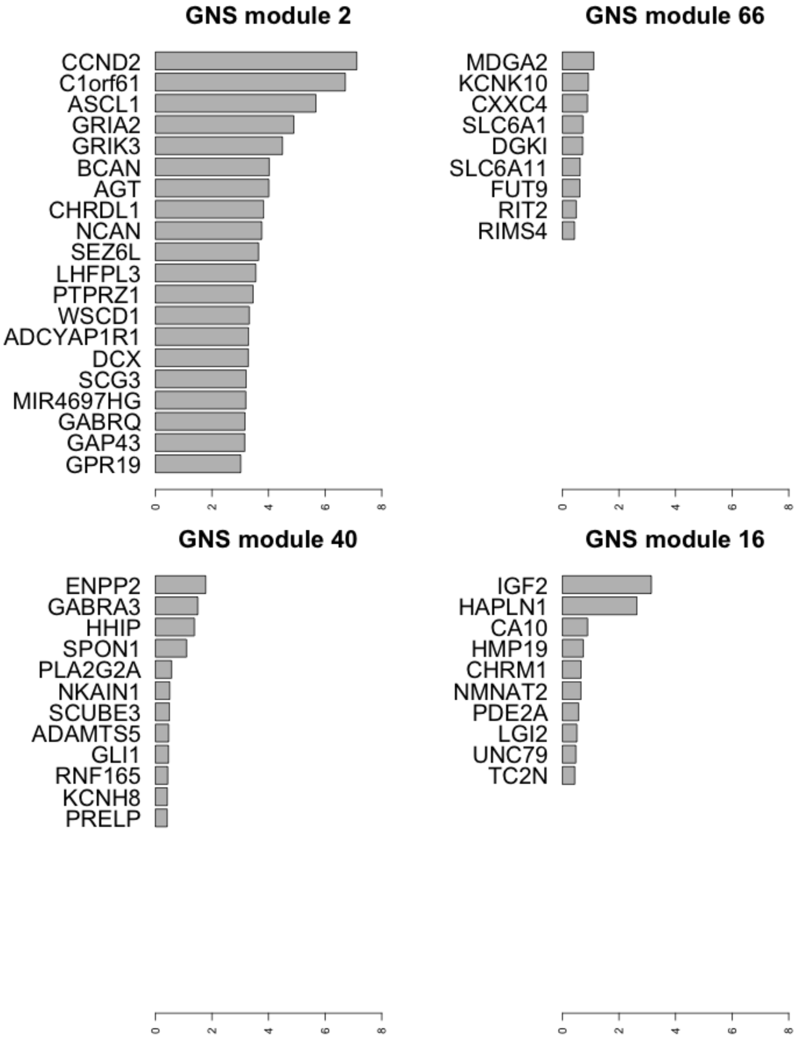


Fig. 4.12 Variation of genes across GNS samples within proneural modules. The primary proneural module (GNS module 2) is shown alongside secondary proneural modules (GNS modules 66, 40 and 16). Variance based on log2 microarray intensity values.

### Secondary GNS proneural modules

The secondary proneural modules are relatively small and contain few highly variable genes. Differences in *IGF2* expression was found to identify neurogenic NS cells expressing *SOX2* and *DCX* from the dentate gyrus compared to comparable NS cells from the subventricular zone [18]. Proliferation of NS cells was regulated by IGF2 in a primarily autocrine manner via AKT-dependent signalling. IGF2 was also found to be critical to the 'Shh' subtype of medulloblastoma [135, 213]. Another component of the PNN, *HAPLN1* (Crtl1), encodes an adaptor protein that binds to both hyaluronan and neural proteoglycans [274]. The formation of PNNs is induced by Crtl1 production and HAPLN1 knockout mice show aberrant PNN formation [32]. Hedgehog-interacting protein (*HHIP*) encodes a Shh inhibitor that is also most commonly associated with Medulloblastoma [132, 162].

## 4.2.5 GNS mesenchymal associated modules

GNS derived modules with mesenchymal associated genes show some of the largest variation in these GNS cells (Figure 4.13). Mesenchymal gene expression in glioma tumours is associated with immune cell infiltration, hypoxia, tumour grade and progression (Chapter 3). A detailed review of the literature for these module genes may help identify the functional or causative origins of this variation.

### The primary GNS mesenchymal module

The primary mesenchymal module has the greatest number of consensus genes shared with the glioma mesenchymal module and also show some of the greatest variation between GNS cell lines (Figure 4.13). The gene showing the highest variation in this module is *EDIL3*, also known as Del-1, encodes an anti-adhesive factor that inhibits integrin binding limiting inflammatory promoting leukocytes [38]. IL17 expression and subsequent neutrophil recruitment was higher in *EDIL3*$^{-/-}$ mice suggesting Del-1 acts as a locally induced suppressor of inflammation reducing subsequent tissue damage [57]. Del-1 also suppresses macrophage activation further limiting inflammatory responses [144]. Inflammation in the central nervous system is associated with many conditions including multiple sclerosis. While expression of *EDIL3* in endothelial cells is long established, expression within neuronal cells was recently detected [39]. The blood brain barrier in *EDIL3*$^{-/-}$ mice was significantly disrupted yet double knockout *EDIL3*$^{-/-}$ *IL17RA*$^{-/-}$ mice were less disrupted.

*TMEFF2*, expressed in both the brain and prostate, enhances the survival of midbrain and hippocampal neurones [102] and regulate RhoA activation and Integrin expression in prostate cancer cells [36]. Thrombospondin-1 (*THBS1*) has a role in both neural development and response to injury [153, 229]. Astrocytes have been shown to express *CCL2* induced my mechanical injury [79] and CCL2 induced migration of microglia, NS and oligodendrocytes has been observed [100, 178]. Overexpression of *GLIPR1* induces the production of reactive oxygen species leading to apoptosis [147] and recombinant GLIPR1 protein has been considered to have therapeutic potential [118]. Fibulin-3 (*EFEMP1*) expression inhibits Notch signalling to promote growth in glioma [103].

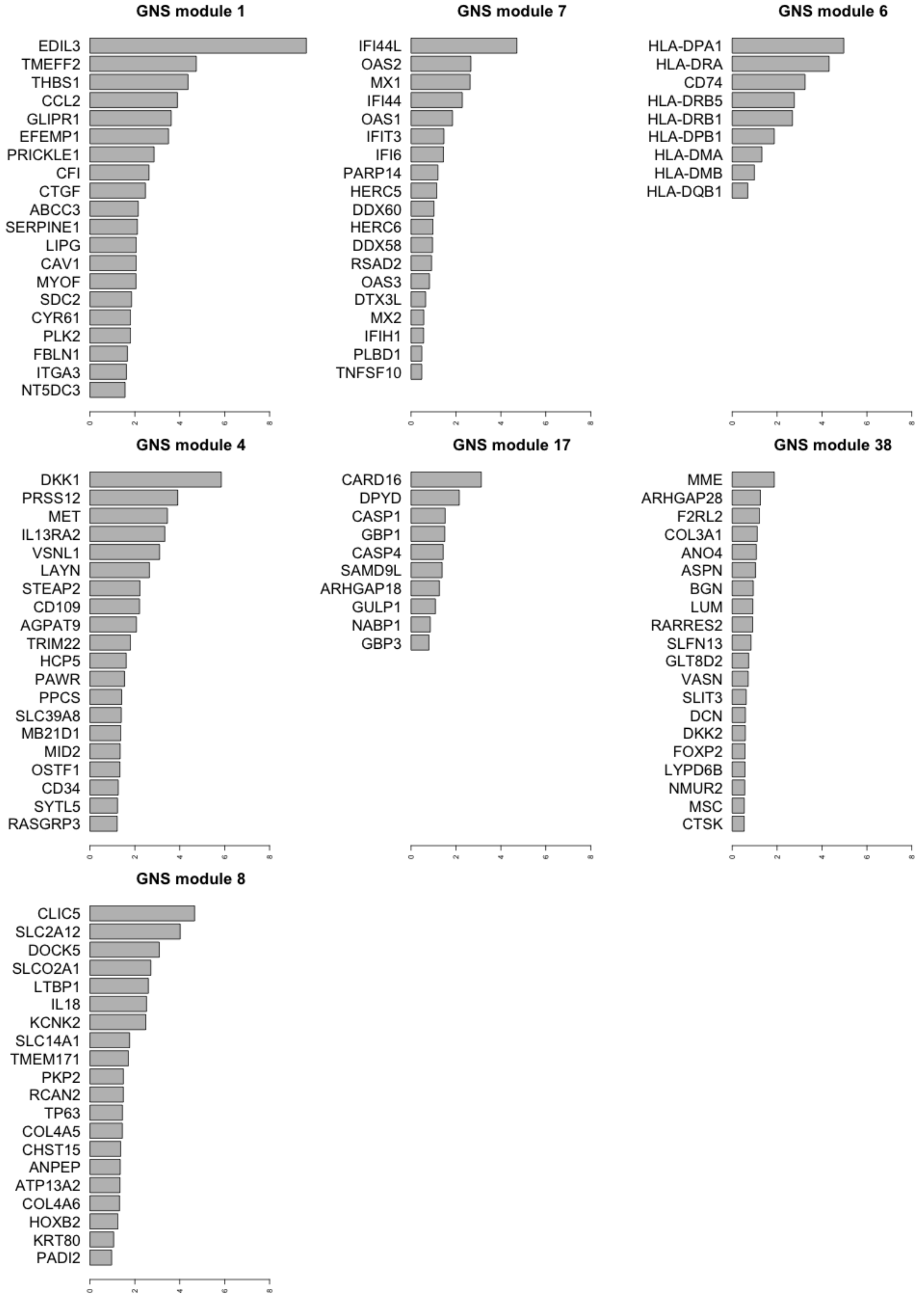


Fig. 4.13 Variation of genes across GNS samples within mesenchymal modules. Primary mesenchymal (GNS module 1), interferon type II module (GNS module 7) and HLA class II module (GNS module 6) are shown alongside secondary mesenchymal modules (GNS modules 4, 17, 38 and 8). Variance based on log2 microarray intensity values.

### Secondary GNS mesenchymal modules

Analysis of secondary mesenchymal modules may help further contextualise the mesenchymal phenotype. Additional mesenchymal modules include two modules enriched for interferon type II response and major histocompatibility complex class II genes. Other mesenchymal modules are enriched for hypoxia and mesenchymal phenotype genes. *CARD16* is a component of the inflammasome and promotes IL-1$\beta$ processing [119]. Inflammasome functioning in non-myeloid cells is not well understood, however inflammasome activity has been identified in neurons and astrocytes [271]. Dihydropyrimidine dehydrogenase (*DPYD*) has recently been described as critical for inducing the epithelial to mesenchymal transition most likely via pyrimidine degradation metabolites [233]. *DPYD* is also implicated in the degradation of chemotheraputic agents [140]. Mesenchymal associated GNS module 4 includes the Wnt inhibitor *DKK1* that is induced by hypoxia in glioma [89]. Another notable gene found within this module is the tyrosine kinase receptor *MET* which also is associated with hypoxia, necrosis and high grade [196] and has been found to be amplified in in individual glioma cells in a mutually exclusive mosaic fashion [240].

### GNS Interferon type II and MHC II modules

Expression of interferon type II response genes has previously been described in cancer [50] and was identified as a consensus coexpression module across multiple cancer types in Chapter 1 of this thesis. The role of interferon induced genes in cancer is poorly understood with most work focused on their antiviral functions. Many of the genes listed in GNS module 7 including IFI44L, MX1 and MX2 were identified as interferon stimulated genes (ISGs) that inhibit viral replication [228]. Different combinations of ISGs were found to be effective against different viruses suggesting a broad range of effectors evolved to counter the diversity of viral threats. The expression of ISGs is induced by multiple pathways including NF-$\kappa$B, RIG-1 (*DDX58*) and STAT1,2 [282]. The RNA helicase RIG-1, gene name *DDX58* and member of the ISG-like GNS module 7, is expressied in response to interferon and initiates antiviral pathways, activated by the detection of intracellular viral components [282]. Members of the OAS gene family synthesise $2'$-$5'$ oligoadenylates which bind and activate RNAse L to degrade cellular and viral RNAs [282]. Interferon treatment has been used against various cancer types [15]. Interferon treatment has been shown to suppress angiogenesis, inhibit proliferation, promote cell death and modulate immune response [231]. Variation in interferon response has been associated with multiple diseases including viral infections, multiple sclerosis as well as cancer [16]. Whether ISG expression in cancer is related to viral infection or rather as a consequence of other interferon pathway functions is in question. Viral infections have been found in cancers, however it has been suggested that increased chance of infection is a consequence of tumour development rather than a tumour initiating factor [117, 211, 243, 244]. Interferon-$\gamma$ signalling induces the expression of major histocompatability complex II (MHC II) genes via the MHC II transactivatior CIITA [16]. Expression of MHC II genes is mostly associated with bone marrow-derived antigen presenting cells like dendritic cells, however some non-professional antigen presenting cells have been shown to possess MHC II antigen presentation including cancer cells [200]. Examining the induction, expression and function of ISG and MHC II responses in GNS cells may further illuminate the complex relationship between tumour cells and immune response.

### 4.2.6 Analysis of GNS proneural and mesenchymal modules in GNS-like cells

Having established coexpression modules in our GNS panel I explored if these observations provided any insight in three previously published GSC datasets alongside GNS cells assayed on the Affymetrix U133 Plus 2.0 array [87, 143, 204] and one dataset on the Affymetrix U133 version 2 array [11]. Lee *et al* examined how GSCs respond when transferred from growth factor media to serum containing media. Two separate GSC lines established from in growth factor media all present a proneural expression signature clustering along with G144, G144ED, GliNS2, the oligoastrocytoma derived G174 and normal NS cells (Figure 4.14). Remarkably, transfer of proneural like GSC lines to serum containing media led to a shift towards a mesenchymal phenotype clustering with mesenchymal GNS lines G166 and G179 as well as traditional glioma cell lines which have been maintained long term in serum. Expression of the mesenchymal module increases with greater passage numbers in the 1228-GSC line. Murine xenograft tumours from growth factor derived 1228-GSC cells present the lowest mesenchymal signature of the 1228-GSC samples and demonstrated extensive infiltration and migration along white matter tracts [143] (Figure 4.15). Similarly 308-GSC xenograft tumours show low mesenchymal expression yet serum cultured 308-GSC cells produced non-infiltrative tumours similar to the long term serum maintained glioma cell lines and expresses high levels of the mesenchymal module [143].

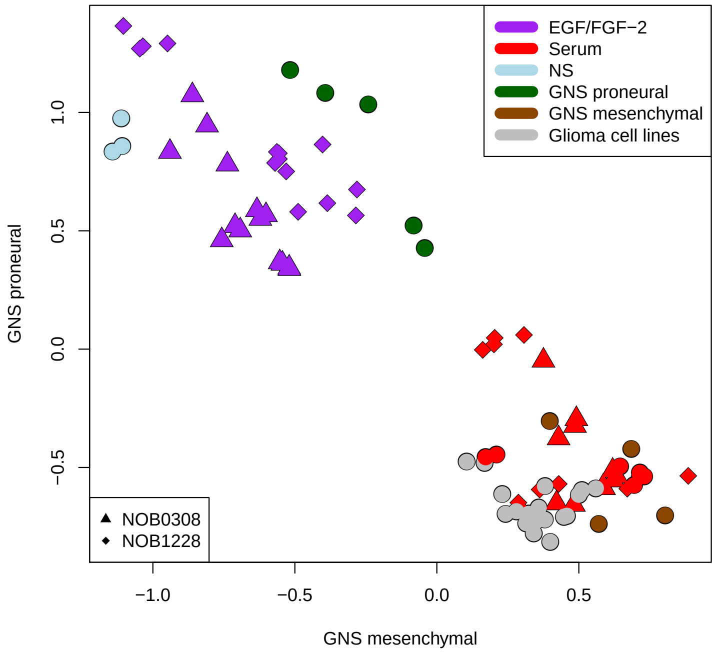


Fig. 4.14 GNS proneural and mesenchymal modules in GNS-like GSCs (data from Lee *et al* [143]). GSC lines cultured initially in EGF/FGF (Purple) and then transferred to serum based culture conditions (Red) are represented by either triangle (NOB0308) or diamond shaped symbols (NOB1228). Neural stem cells cluster with the proneural GSCs here (Light blue circles). GNS lines assayed on the same microarray platform are dispersed along the proneural to mesenchymal axis or EGF/FGF to serum axis without serum culture. Traditional serum cultured glioma cell lines present high mesenchymal signature and cluster with the other serum cultured cells. Units are z-score normalised log2 microarray intensity values.

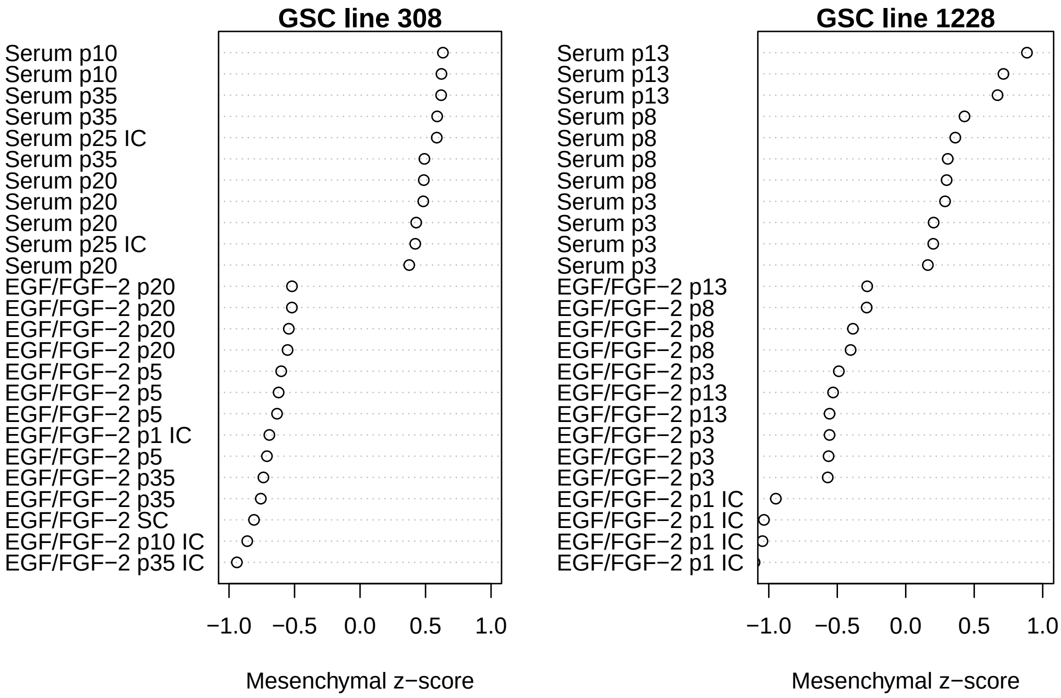


Fig. 4.15 Sorted GNS mesenchymal module expression in data from Lee *et al* [143] showing serum associated mesenchymal transition through increasing passage number for two cell lines. Cell lines initiallly established in EGF/FGF when transferred to serum containing media transition to a high mesenchymal gene expression state. Passage number is denoted by p[passage number], see Lee *et al.* for the details of each sample displayed. Units are z-score normalised log2 microarray intensity values.

Günther *et al* reported two distinct phenotypes of GSCs distinguished by morphology and CD133 expression [87]. Examining proneural and mesenchymal modules in GSCs isolated and classified by Günther *et al* finds the same distinct proneural (GSf) and mesenchymal (GSr) cell line clusters as found in GNS and cell lines generated by Lee *et al* (Figure 4.16). Bhatt *et al* also found two major clusters of GSCs [11] and when mapped to GNS co-expression modules produce a proneural to mesenchymal axis with less distinct separation between the clusters. In their paper, Bhatt *et al* state that TNF-$\alpha$/NF-$\kappa$B activation induced a mesenchymal shift, however TNF-$\alpha$ treated lines did not present particularly high levels of the mesenchymal module in comparison to other mesenchymal lines suggesting either incomplete differentiation or alternate definitions of what a mesenchymal signature describes (Figure 4.16). Generally the examination of proneural and mesenchymal modules in other datasets reinforces the dominant impact of these signatures that have been previously identified before in separate studies without unification into a comprehensive picture.

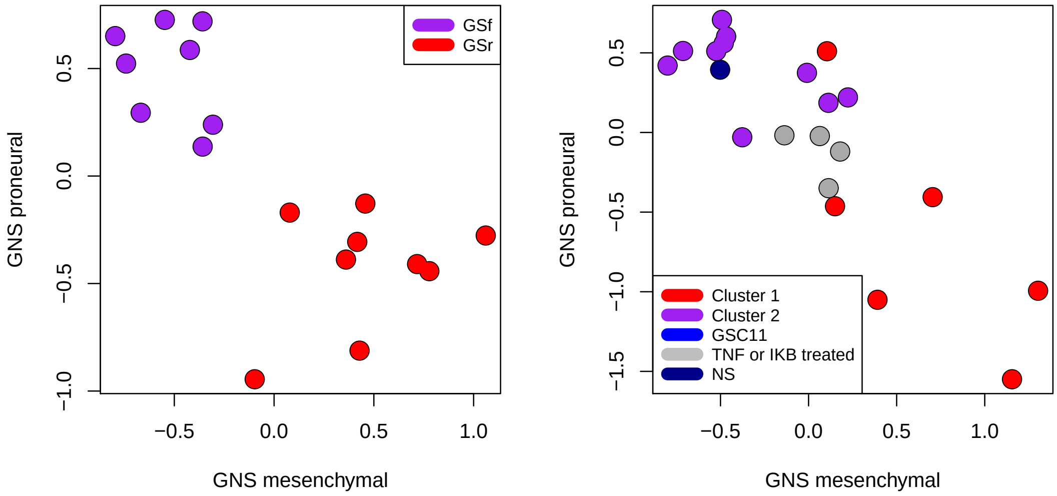


Fig. 4.16 GNS proneural and mesenchymal coexpression modules in data from Günther *et al* [87] and Bhatt *et al* [11]. The clustering of GNS-like GSCs reveals a comparable proneural to mesenchymal axis as derived in our GNS panel matching independently proposed GSf/GSr and cluster 1/2 classifications from their respective papers (Figure 4.9). Units are z-score normalised log2 microarray intensity values.

## 4.3 Discussion

In this chapter the transcriptomic characterisation of glioma derived cancer stem cells (GNS) is set out by comparisons to karyotypically normal fetal neural stem cells, classification by glioma subtype signatures and coexpression clustering. Extending the analysis by Engström *et al*. expression profiling of 12 GNS and 4 NS lines reveals an more detailed profile of the transcriptomic differences between these phenotypically similar yet ontologically distinct cell types. The consistency of NS expression contrasts sharply with the diversity of expression found between GNS lines. Many of the over expressed in GNS compared to NS are transcription factors implicated in neural development like *FOXG1*, *TFAP2A*, *NKX2-2*, and *LMO4*. Regulators of proliferation and cell division like *RNF114*, *OGFR* and *TGFA* are likewise highly expressed in GNS cells. Comparatively overexpressed in NS cells are members of the Ran activated importin family including *RANBP17*, *KPNA3* (IPOA4) and *IPO5* (RANBP5). Differential expression of these genes may imply differences in the transport of molecular cargo into the nucleus. Further studies may help reveal the role of these importins on GNS and NS biology. Another class of genes that are highly expressed in NS cells are established tumour suppressors like *RB1*, *PTEN EPB41L3*, *NEDD4* and *PTPRD*. While the degree to which these GNS-like cell lines are capable of recapitulating the cellular state of tumour cells *in vivo* I show that genes differentially expressed between GNS and NS cells show a anticorrelated and coexpressed relationship in glioma tumours. While these results do not ensure GNS cells are a perfect reflection of the functional reality present within tumours it could be suggested that these cells are the most tractable model for high throughput characterisation of live cells. Care should be take to avoid long term growth of cell lines in *in vitro* to avoid selection for a phenotype that is readily proliferative in culture and far from the cell types from which they were originally derived.

Examination of the subtypes established by Verhaak *et al.* in GNS lines reveal the difficulty in comparing between relatively homogeneous cell lines and tumour samples that are composed of multiple cell types in varying microenvironment. A modified approach to Verhaak *et al.*'s centroid classification was able to identify subtype diversity between GNS lines with a bias towards the proneural and mesenchymal subtypes. The application of coexpression clustering to GNS expression data reveals many modules that contain proneural or mesenchymal associated genes. The two GNS modules that intersected the largest number of glioma proneural and mesenchymal genes respectively have the some of the most variance across GNS lines. Expression of these modules and principal component analysis replicates the separation of GNS lines into two distinct proneural or mesenchymal clusters indicating the dominance of glioma subtype line expression. NS lines were found to cluster more closely with the GNS mesenchymal cluster lines than GNS proneural. A total of 4 modules were associated with glioma proneural genes and 7 modules were associated with glioma mesenchymal expression. The observation that proneural GNS-like cells shifted towards a mesenchymal phenotype when cultured in serum containing media suggests that these cells are not locked into a single subtype expression program. Moreover it raises the possibility that proneural cancer stem cells may transition towards a mesenchymal phenotype *in vivo* as a response to environmental cues. The identification of proneural or mesenchymal-like subtypes in GSCs is not novel as many of these distinguishing features have been described before [11, 45, 87]. The analysis set out in this chapter is distinguished in reference to the novel identification of a proneural to mesenchymal axis between and within glioma tumours. Exclusion of other non-identifiable subtypes suggests these GSCs subtypes are able to represent the major forms of GSC variance in glioma tumours. Mapping research that describes these distinct GSC features, previously considered to be independent, back to an intratumoural glioma context may provide substantial insight into the development, progression and treatment of this complex disorder.

A survey of the literature describing the highly variable genes within the proneural and mesenchymal modules reveals a functional insight that deserves further experimental work. Classically proneural genes like *OLIG2*, *OLIG1*, *BCAN* and *DLL3* are represented in the primary proneural module. Many other genes found within the proneural modules are transcription factors with established roles in neural development and fate specification including *ASCL1*, *NFIA* and *NEUROD1*. Proneural expression of CyclinD2/Cdk4 cell cycle control genes may relate to cell type specific regulation of proliferation and asymetric cell division that is promoted by alternative mechanisms in mesenchymal GNS cells.

Another proneural associated group of genes are those encoding extracellular matrix proteoglycans like *BCAN*, *NCAN* and *PTPRZ1*. These PNN related proteins have roles adhesion and migration of CNS resident cells and may play a role in the distinct morphology of proneural GNS cells. The neural developmental skew of these genes suggests the proneural phenotype of glioma represents the disregulated proliferation of glial progenitors like oligodendrocyte precursor cells.

GNS mesenchymal modules contain little gene expression directly relating to immune cell infiltration as seen in glioma coexpression modules due to their *in vitro* culture separately from immune cells. Mesenchymal associated modules in GNS are enriched for genes involved in inflammation, immune modulation and response to injury including *EDIL3* and *CCL2*. These associated processes suggest that the mesenchymal phenotype relates to CNS immune responses and inflammatory processes. The association of many of these genes with CNS injury responses like glial scar formation draws the suggestion that the mesenchymal signature may represent a co-opted transcriptional response to CNS injury that is exploited by glioma cells in response to changes in their microenvironment. This co-opted transcriptional program could be compared to the epithelial to mesenchymal transition observed in carcinomas. The enrichment of these microenvironmental response processes in the mesenchymal phenotype compared to the proneural phenotype suggests the mesenchymal signature is a response to the shifting tumour environment increasingly burdened by the tumour bulk with gradually worsening grade. The identification of interferon response and MHC class II genes highlights what is a poorly understood mechanism in cancer. The function of induced interferon response genes in GNS cells may represent a pre-emptive programmed response to potential viral infection or may play a unique role in a different aspect of GNS biology. Understanding these processes may help improve the application of Interferon-$\gamma$ as a therapeutic strategy. Similarly an understanding of MHC class II expression in GNS cells may assist in the search for targeted therapeutics. In a recent paper by Meyer *et al.*, multiple clonal GSC lines derived from within single tumours were shown to show glioma subtype expression patterns [174]. In this study separate clones from within the same tumour were found to present differential response to temozolomide therapy. The genes found to be differentially expressed between resistant and non-resistant lines include many described in this chapter as members of proneural or mesenchymal modules including *OLIG2*, *GRIA2*, *BCAN*, *GRIK3*, *NCAN*, *MAPT* for the non-resistant clone and *NMEFF2*, *CALCRL*, *RNF217*, *LOX* and *CD44* for resistant clones. This may suggest that resistance to temozolomide and other therapeutics is higher in mesenchymal phenotype cells.

The analysis set out in this chapter presents many avenues of potential further work. Identifying critical differences between GNS and NS cells could help identify exploitable targets for therapeutic intervention. Likewise understanding the origins and plasticity of proneural and mesenchymal phenotypes in GNS cells may help to understand the complex origins and evolution of gliomas or more significantly identify the mechanisms exploited by these cells.

# Chapter 5

# ATAC-seq analysis and its application to GNS and NS cells

## 5.1 Introduction

Examination of cancer gene expression, like the analysis described in the preceding chapters, is critical to understanding the regulatory gap between genotype and phenotype. However transcriptome analysis alone reveals little information describing how gene expression is controled beyond expression of known regulators. In order to investigate other contributions to the regulation of expression, various methods utilising next generation sequencing have been developed. Most methods for interrogating genomic features are based on the relative enrichment of DNA sequences at different loci. For example ChIP-seq uses antibody specificity to a protein of interest to selectively isolate target protein bound regions of genomic DNA. These methods rely on intentionally enriching the DNA library with genomic DNA sequences proximal to features of interest. Using the relative density of alignments allows for examining features genome wide. As such, a great variety of methods have been developed to exploit the signal of relative density of aligned reads to identify various features of interest. One genomic feature with a regulatory role describes the physical accessibility and compaction of DNA. Accurate identification of open chromatin can identify regulatory features like gene promotors, enhancers or insulators and from these features, regulated genes can be inferred and motifs describing the DNA sequences at which transcription factors (TFs) may bind can be identified.

### Methods for detecting chromatin accessibility

Various methods have been applied to identify open chromatin. The most commonly used methods are MNase-seq, DNase-seq and the comparatively new method ATAC-seq [288]. MNase-seq exploits the single-stranded nucleic acid digesting MNase enzyme to identify nucleosome-bound digestion protected DNA fragments [97]. The MNase enzyme has a preference towards digestion of AT-rich sequences which may bias downstream analysis [173]. Similarly MNase-seq relies upon a DNA fragment size selection step to intentionally extract DNA associated with a particular class of proteins. For the analysis of nucleosome positioning, fragments of ~150bp are selected with the knowledge that MNase digests up to the edge of nucleosome protected DNA. Smaller fragments have been used to identify protected TF binding sites and associated open chromatin [97]. By contrast DNase-seq identifies accessible chromatin by DNaseq I mediated fragmentation [99]. These fragments are then ligated to polymerase chain reaction (PCR) adaptors for library preparation and sequencing. The DNase enzyme does not possess exonuclease activity, in comparison to MNase, and fragment size selection is commonly restricted to ≤100bp to avoid the inclusion of overwhelming numbers of ≥150bp nucleosome bound fragments. DNase also has a significant strand, sequence and CpG methylation binding site bias that influence the frequency and location of DNA cleavage events [139]. A variant of DNAse-seq called FAIRE-seq uses formaldehyde assisted crosslinking of DNA to regulatory proteins followed by selective isolation of these protein/DNA complexes in an aqueous phase extraction [78]. This method has the advantage of restricting the library to protein bound DNA.

Other considerations for the analysis of sequencing based methods include sequencing depth, PCR amplification bias and PCR duplicates. The depth to which each library is sequenced influences how much information can be extracted with a degree of confidence. Methods like MNase-seq for nucleosome bound fragments will typically require comparably deeper sequencing due to the abundance of nucleosome bound DNA in comparison to methods intended to identify the open chromatin regions which make up a smaller fragment of the epigenome. The amount of PCR amplification will also influence the quality of library preparation as a restricted number of over amplified sequences can overwhelm a sequencing library with duplicates, reducing the pool of other more unique fragments. The effect of PCR bias may also introduce amplification variation with over or under estimation of particular loci.

### ATAC-seq

A recently described method for profiling accessible chromatin and nucleosome positioning exploits a hyperactive derivative of the bacterial transposase Tn5. Wild-type Tn5 acts as a dimer to insert its transposon which is flanked by two IS50-element mosaic ends. Loading the transposase with single mosaic ends, not linked to the dimer partner's mosaic end DNA, leads to fragmentation of the bound double stranded genomic DNA, as opposed to the transposon insertion generated in wild-type Tn5 transposition. This random fragmentation of double strand DNA was used to efficiently prepare whole genome sequencing libraries with the added benefit of utilising the inserted mosaic end DNA as PCR primers for library amplification [2]. This fragmentation of DNA by sequencing adaptor insertion was applied to *in vitro* native chromatin structure where open chromatin is identified by the ability of the transposase to bind accessible loci [24]. Heavily condensed heterochromatin wrapped around nucleosomes is protected from transposition therefore ATAC-seq fragments are largely restricted to regions of open chromatin (Figure 5.1). Genome alignment of Tn5 fragmented DNA allows for the length of the DNA fragment to be evaluated. As nucleosome bound DNA is protected from transposase access, transposition of sequencing adaptors either side of a nucleosome produces $\geq$150bp length fragments. This protection and enrichment of $\sim$150bp fragments is observed in fragment size distribution with $\sim$180bp periodicity with two nucleosome and three nucleosome bound fragments enriched at $\sim$260bp and $\sim$340bp respectively (Figure 5.1). This enrichment for likely nucleosome-bound fragments enables the identification of nucleosome bound sites using ATAC-seq.

Another advantage to Tn5 fragmentation is that sequencing adaptors are directly integrated into the native chromatin which leads to reduced fragment loss in subsequent library preparation steps. This means that fewer cells are required for library preparation facilitating the analysis of samples that can only be isolated in small populations. Indeed, these reduced input requirements ATAC-seq has even been applied to analyse chromatin accessibility in single cells [25]. ATAC-seq libraries can also be prepared quickly suggesting that this method could be applied to clinical samples providing swift analysis and extraction of prognostic features on clinical time scales. It was also suggested that the relative protection TF binding sites or footprints could be used to reconstruct regulatory networks [24], however this methodology has not been comprehensively investigated. Another factor affecting ATAC-seq analysis is the Tn5 transposase's preference for binding to GC nucleotides [2]. This bias towards GC rich loci will influence both the identification and quantification of accessible loci.

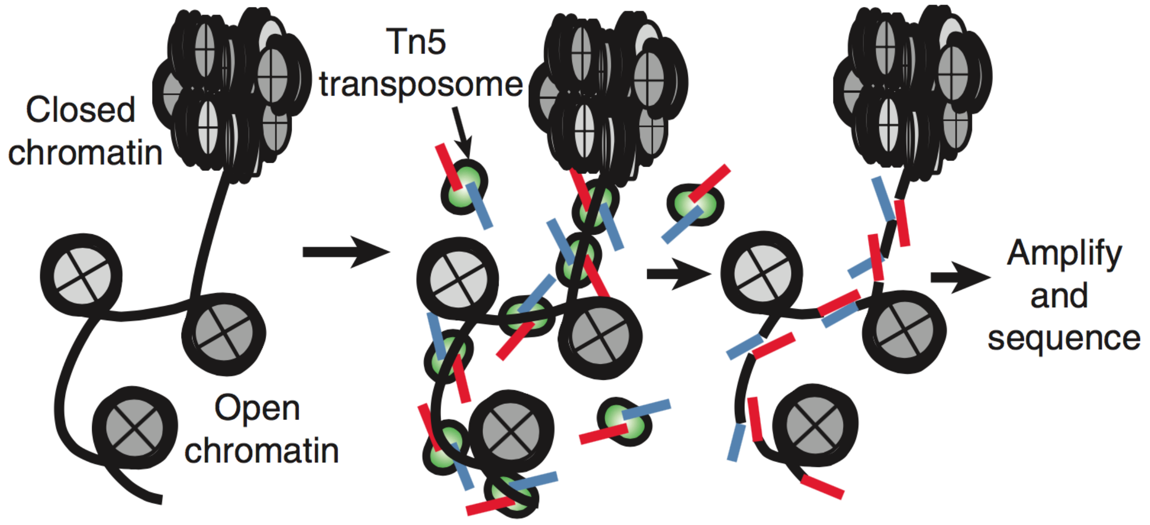


Fig. 5.1 Sequencing adaptor transposition into accessible chromatin via ATAC-seq. Tightly condensed heterochromatin and open euchromatin are differentially accessible for DNA binding by the Tn5 transposasome. Sequencing adaptors shown in red and blue are inserted into accessible chromatin consequently producing an enrichement for short nucleosome free fragments or long ∼150bp nucleosome bound fragments. Adapted from Buenrostro *et al.* [24].

## 5.2 Results: Methods for ATAC-seq analysis

Methods and tools for preparing ATAC-seq data for robust downstream analysis are not yet well established as of the time of writing. To enable downstream analysis and eventual conclusions novel methods and analysis pipelines had to be developed. In this section I set out some of the unique attributes of ATAC-seq data preprocessing and analysis. Sources of potential bias in the ATAC-seq method are also identified and evaluated in this chapter.

### 5.2.1 An analysis of the Tn5 transposase DNA binding bias

As may be expected for a DNA binding enzyme, the Tn5 transposase used for ATAC-seq library preparation has been described as having A GC base bias and physically covers an ∼28bp footprint of double strand DNA [2]. Sites of native Tn5 transposition are characterised by a 9bp replicated region flanking the donor inserted or transposed DNA. Tn5 transposition in ATAC-seq analysis attaches sequencing adaptors to both $5'$ ends of the target DNA duplex and duplicates a 9bp replicated region in both strands through the action of host cell DNA repair factors. After sequencing and adaptor trimming the sequence read will therefore start with this 9bp replicated region continuing through to the non-replicated sequence downstream. The binding bias of the Tn5 transposase is however not comprehensively understood having unknown effects on the downstream analysis of ATAC-seq data. In order to characterise the transposase binding site (TBS), 10,000 random paired end reads were selected from two seperate human ATAC-seq libraries and the nucleotide frequencies for 60bp upstream and downstream of the read ends were extracted. Examining the GC content of these sites finds variation within the 9bp replicated region as well as ∼7bp on either flank of the replicated region (Figure 5.2). The GC content extending into the read is slightly higher than the flank of genome not sequenced which may be explained by PCR bias in the amplification of the sequencing library [129].

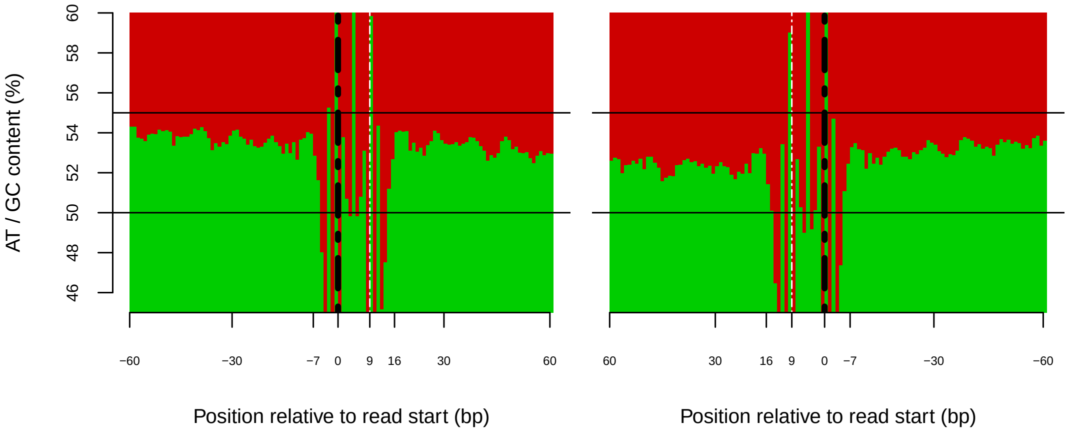


Fig. 5.2 GC nucleotide bias in transposase insertion site. Variation in the nucleotide bias of the Tn5 transposase extends 7bp beyond the edge of the read and 16bp within the read incorporating the 9bp replicated region and a symmetrical 7bp overhang. The frequency of GC nucleotides are shown in red and AT nucleotides in green. GC content is enriched relative to the genome background within the read, likely due to PCR bias. Boundary edges of the 9bp replicated region are indicated by thick and thin dashed lines. The black thicker dotted line indicates the boundary edge to which the sequencing adaptor was ligated for this fragment.

Examination of the 23bp region within the TBS reveals reflectional and rotational symmetry in AT/CG and purine/pyrimidine content respectively (Figure 5.3). The 9bp replicated region has a general enrichment for GC content with the exception of the central nucleotide which has an 11% higher frequency of AT bases. On the boundary of the 9bp replicated region there is a preference for GC on the inside boundary and AT on the outside of the boundary replicated in both read pairs. This boundary bias is highest on the sequencing adaptor bound edge compared to the within-read boundary edge. This subtle asymmetry in GC content extends beyond the 9bp replicated region with a preference for two alternating GCs with ATs followed by a string of ∼3 GC bases. This boundary difference may in part relate to PCR bias however the observation of AT enrichment beyond the edge of the read, and thus a sequence replaced by a sequencing adaptor for PCR amplification, suggests this bias may originate with the transposase itself. One potential explanation is that the AT to GC boundary is biochemically favourable for the sequencing adaptor's ligation to the target DNA. In this way the within read 9bp boundary would still be favoured for the transposase dimer binding, yet fragments sitting on a AT to GC boundary are slightly more likely to be represented in the sequencing library. Alongside the GC bias there is also a bias in purine/pyrimidine content showing rotational symmetry through the middle of the 9bp replicated region. This rotational symmetry fits with the proposed transposase binding as a dimer with each unit orientated with opposing DNA strands. These results describing a GC bias and the footprint of the transposase on accessible chromatin are crucial for further downstream analysis.

### 5.2.2 Considerations for ATAC-seq pre-processing

With the establishment of new methods, the processing steps necessary for downstream analysis are evolved to adapt to the unique biases and experimental factors that effect the utility of the method. For ATAC-seq the transposase's preference for binding GC rich loci will affect the estimation of accessibility at varying loci. Likewise the effect of DNA fragment insert size variation between libraries may introduce another form of bias to the analysis. Experimental factors like variation in chromatin structure, disruption or disassociation during exposure to the transposase may introduce other types of bias or experimental variation to consider.

#### Adaptor trimming

ATAC-seq libraries are amplified using the transposed sequencing adaptors as PCR primers. As the transposase occupies $\sim$28bp the minimal fragment insert size is $\sim$38bp including two 9bp replicated regions and two $\sim$9.5bp flanking regions [2]. Adaptors from ATAC-seq library reads are found at both $5'$ and $3'$ ends of each strand so when the sequencing extension length is longer than the insert size, the sequencing adaptor will be included in the $3'$ end of the sequence. For this reason the identical adaptor sequence must be trimmed from both $5'$ and $3'$ ends of each read before alignment.

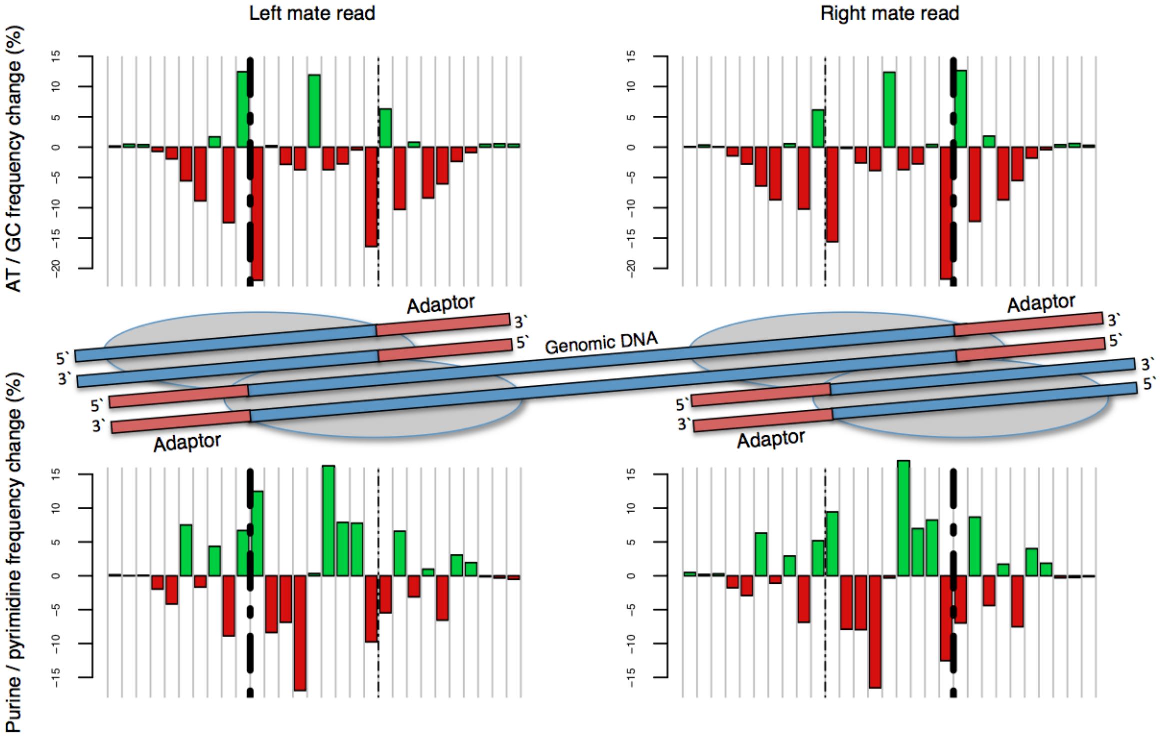


Fig. 5.3 GC/AT and purine/pyrimidine bias within the transposase binding site (TBS). AT or Purine bias is represented by green positive bars and GC or Pyrimidine bias is represented by red negative bars (% change from genome average frequency). A model of the Tn5 transposase binding as a dimer and transposition of sequencing adaptors is illustrated, approximately scaled to its predicted footprint. Boundary edges of the 9bp replicated region are indicated by thick and thin dotted black lines. The thicker dotted line indicates the boundary edge to which the sequencing adaptor was ligated for this fragment. Replicates of these plots derived from an independent ATAC-seq library can be found in the appendix (Figure 7.9).

### Detecting accessible chromatin

Having established the nucleotide biases and physical footprint of the Tn5 transposase the next step in the analysis of ATAC-seq data is determining how to identify open chromatin. The process by which ATAC-seq finds open chromatin is by identifying regions of the genome where the transposase is able to bind and successfully insert sequencing adaptors. Extending this logic, the 28bp transposase footprint represents the known locus of open chromatin stretching 9.5bp beyond the edge of the read and 18.5bp into the read inclusive of the 9bp replicated region and the symmetrical 9.5bp flanking region. For practical purposes to annotating these loci the extension flanking the 9bp replicated region were reduced to 9bp for a conservative total footprint of 27bp. These transposase binding site (TBS) footprints across the genome can then be used to look for enrichment with peak calling or used for differential accessibility using loci counts. An additional factor to consider is that a TBS is only detected if a comparable second transposition event happens proximal to the first. Stochastic events where a single transposase finds rare access to a region of compacted chromatin should be less likely to be presented in an ATAC-seq library compared to insertion events that take place in larger accessible loci.

**Paired-end insert size analysis**

As ATAC-seq tags chromatin fragments with sequencing adaptors, paired-end sequencing and subsequent alignment returns an additional feature of interest, the fragment insert size. Analysis of the fragment insert size density revealed an ~200bp periodicity which was proposed to represent nucleosome bound loops of transposase protected chromatin [24]. Using these ~200bp multiple insert size fragments, nucleosome binding sites could be inferred. Fragments with an insert size of less than 100bp were considered to likely originate from nucleosome free loci. Comparing the insert size distribution of 30 ATAC-seq libraries reveals substantial variation between libraries (Figure 5.4). Eight libraries were purified using a slightly different AMPure bead protocol at the insistence of the sequencing facility resulting in a reduced capture of fragments less than ~120bp compensated by a higher representation of fragments larger than ~150bp. These libraries with greater average insert size also seem to have a more pronounced 200bp "nucleosome bound" periodicity likely representing an enrichment for these protected fragments.

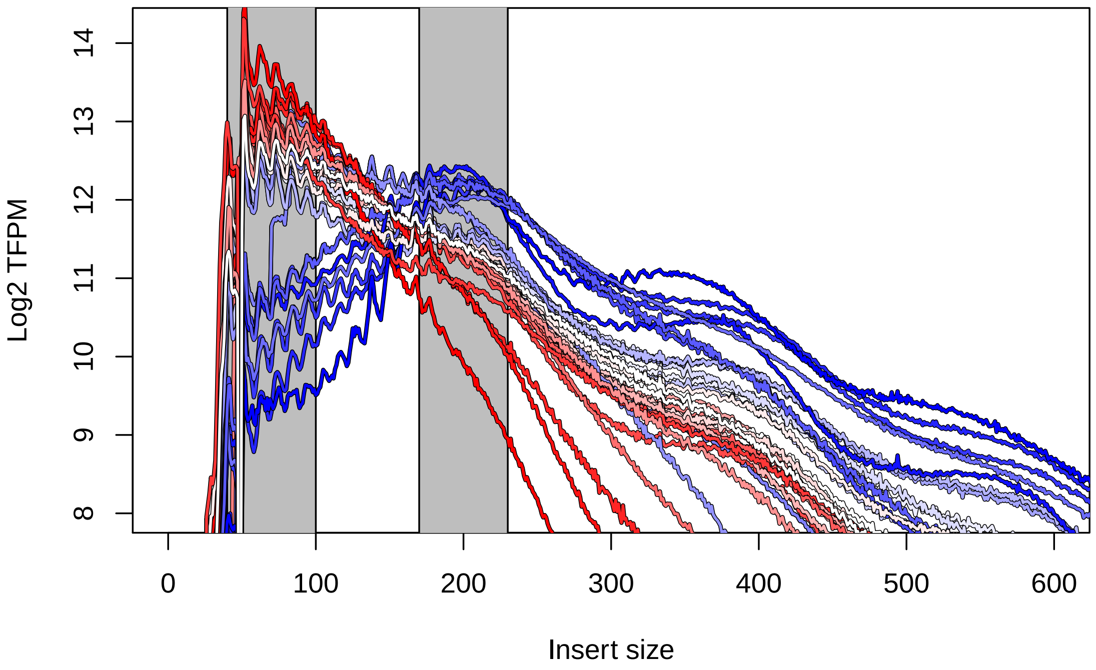


Fig. 5.4 Variation in insert/fragment size between libraries. The red to blue gradient of lines represents the ratio of average insert sizes in windows between 40-100bp and 170-230bp (represented by grey boxes). The eight libraries prepared using the slightly different protocol are dark blue and have a distinct enrichment at around ∼200bp insert size. Y axis shows the Log2 transposed fragments per million.

Theoretically variation in the ratio of nucleosome free and nucleosome bound regions at different loci could lead to over or under estimation of specific loci in insert size skewed libraries. While experimental variation in insert size distribution can be mitigated by improved and more consistent library preparation, variatiation of insert size between libraries is somewhat inescapable therefor examination of this potential bias may be of importance. Using the karyotypically normal foetal forebrain U3, U4 and U7 NS libraries to mitigate the effect of genomic copy number variation, counts of TBSs for fragments between 40-100bp or 170-230bp at 266,060 loci (See "Loci selection" below) were compared (Figure 5.5 and 5.6). As expected the potentially nucleosome bound 170-230bp insert size counts are more frequent in large fragment biased libraries compared to short 40-100bp insert counts. Clustering and principal component analysis of the sample and insert size counts reveals that library insert size counts tend to cluster together indicating that the variation between insert

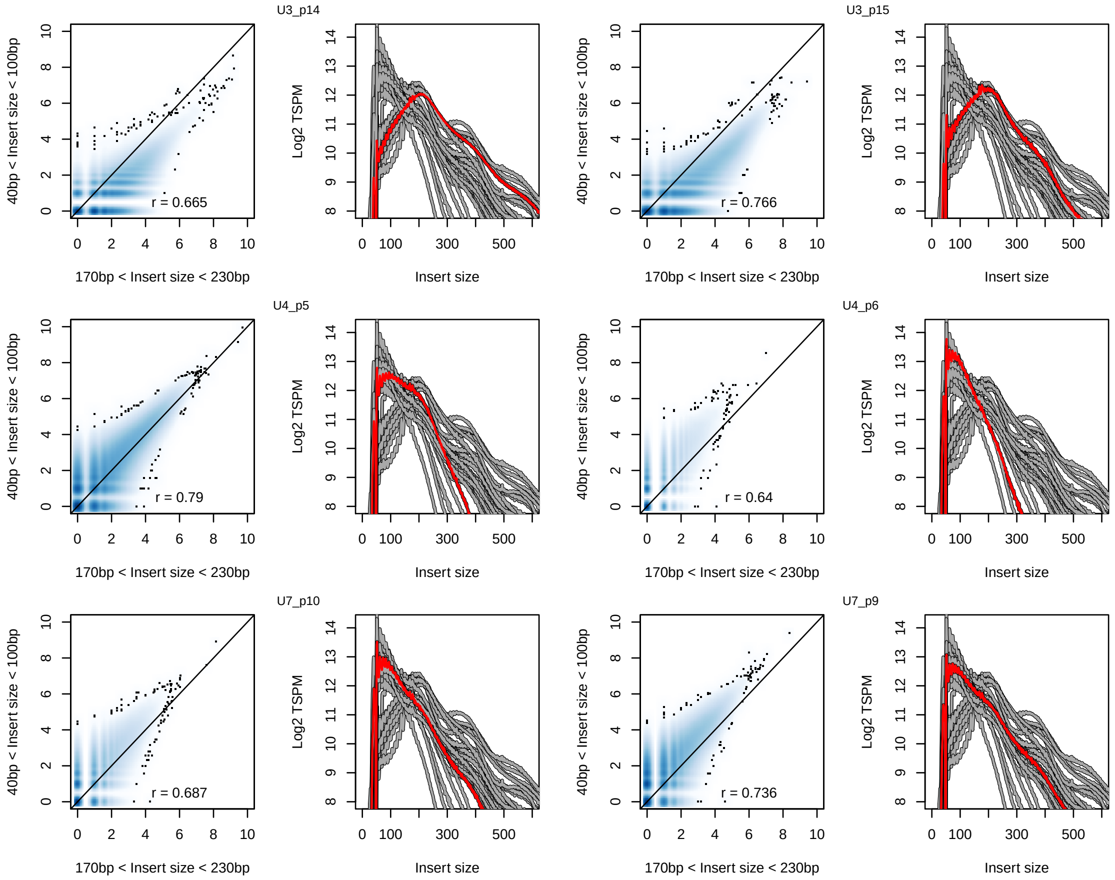


Fig. 5.5 Analysis of library insert/fragment size variation on loci accessibility estimates. Counts for transposase binding sites extracted from reads with either 40-100bp or 170-230bp insert sizes alongside insert size distribution plots for 6 karyotypically normal NS lines. Counts from libraries with an enrichment for ~200bp insert size fragments show a relative over abundance of counts in the 170-230bp insert size set but remain broadly correlated.

size counts is lower than between cell lines and passages. These findings indicate that while library insert size distribution differences do introduce some additional variation, this variation is low for even the most divergent libraries biased by experimental library preparation differences.

### Detecting nucleosome positioning

The observed periodicity of ~200bp multiple fragments suggests that nucleosome bound DNA, which covers ~150bp extended by 18bp on each flank of the nucleosome to allow for transposase binding, is protected from transposition. Enrichment of these ~200bp multiple

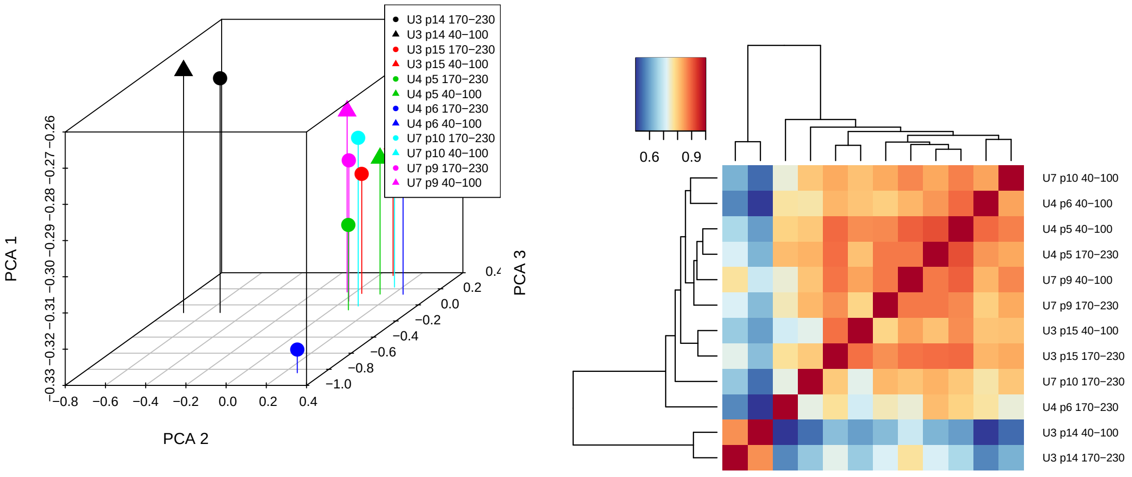


Fig. 5.6 Clustering analysis of library insert/fragment size variation on loci accessibility estimates. Principal component analysis and hierarchical clustering of 40-100bp or 170-230bp insert size count sets reveals a preference for inset size counts to cluster together rather than separating into separate 40-100bp or 170-230bp insert size clusters irrespective of original line. This preference for within cell line clustering suggests that insert size variation has less effect on accessibility estimates than other forms of variation.

fragments was used to infer nucleosome enrichment at genomic loci [24]. As the insert size distribution varies between libraries the power to consistently identify nucleosome bound regions will differ between libraries. Splitting aligned reads into ≤100bp nucleosome free reads and ≥150bp nucleosome bound reads can reveal the organisation of genomic features like transcription start sites [24].

### 5.2.3 Application of ATAC-seq for differential analysis of chromatin accessibility

#### Loci selection

In ATAC-seq analysis open chromatin is identified by finding loci enriched with TBSs compared to the genome background. As set out above the Tn5 transposase has a GC base bias which will cause more TBSs to be identified in GC rich regions. Likewise the likelihood of transposition at individual loci with be partially influenced by similarity to the preferred Tn5 transposase AT/GC or purine/pyrimidine sequence. At the time of writing no peak-calling algorithm is described that can integrate binding bias variables to adjust the influence of individual access sites on enrichment. The flexible peak-calling software package F-seq identifies sequence tag enriched loci compared to the background using local kernel density estimates for specified feature sized windows [17]. Applying F-seq to TBS loci identifies the loci more accessible compared to the condensed heterochromatin background. The F-seq model enables the enrichment of different expected feature sizes. To enable the identification of both broad and relatively small regions of accessible chromatin F-seq was applied using two separate sets of parameters for broad (2kb window) and fine peaks (600bp). Enriched loci for each feature size were called using TBS loci for each ATAC-seq library and were subsequently merged into a collection of broad or fine loci identified in any individual ATAC-seq library. Combining the broad and fine loci into a single loci set enables counts of TBSs to be compared across different libraries for the same sites. Loci that overlapped regions blacklisted for functional genomics analysis by the ENCODE consortium were removed from further analysis [259]. Any loci uniquely more accessible in a single library or phenotype should be represented in the final sample/loci count matrix. Loci for TBS count extraction can also be selected based on annotated features like transcription start sites or known enhancers.

#### Conditional quantile normalisation and differential analysis

The established GC bias of the Tn5 transposase leads to over estimation of accessibility of GC rich loci. Loci varying in GC content are therefore not directly comparable without adjusting for GC bias. Similarly differences in the distribution of counts must be accounted for and proportionally adjusted before comparing between samples. The conditional quantile normalization (CQN) algorithm was implemented to robustly normalise RNA-seq data taking in to account both the proportion of GC content and transcript length [94]. These loci accessibility estimates normalised by CQN taking into account GC content and loci length provides both a general use matrix for the visualisation of chromatin accessibility. The offsets calculated by CQN normalisation can also be passed to differential count enrichment tools like DEseq and edgeR as loci specific normalisation factors [173]. The application of differential count enrichment tools allows for the identification of loci differentially accessible between different groups or conditions.

## 5.3 Results: Application of ATAC-seq to the characterisation of GNS and NS cells

The above discussed pre-processing, biases and methods used for the analysis of ATAC-seq datasets were then applied to examine chromatin accessibility in a panel of seven GNS and five NS cell lines with biological replicates. The analysis of these ATAC-seq libraries broadly mirrors the expression analysis from Chapter 3 by comparing GNS versus NS cell chromatin accessibility as well as identifying variation within GNS lines that is representative of GNS proneural and mesenchymal chromatin organisation.

### 5.3.1 ATAC-seq GNS to NS cell line comparison

A total of 266,060 enriched loci were found across twenty three ATAC-seq libraries. Principal component analysis of CQN normalised counts finds the NS cells cluster together with one outlying NS library (U3 p14) in principal component 1 (Figure 5.7). Meanwhile the GNS libraries are dispersed across the first three principal components. This within NS consistency and significant diversity between GNS samples mirrors the equivalent gene expression profiles described in Chapter 3 (Figure 4.3). Differential analysis of CQN normalised counts using DESeq2 [159] between fourteen GNS and nine NS libraries finds 40,579 loci enriched in GNS cells covering 50.6Mb and 30,157 loci enriched in NS cells covering 22.9Mb ($p \leq 0.05$). While GNS cell line genomes are highly aneuploid with frequent amplification of chromosome 7 and loss of heterozygosity of chromosome 10, differential accessibility is not strictly dependent on copy number variation. Significantly enriched loci on the frequently amplified chromosome 7 are biased towards the GNS lines with 12,042 loci in GNS compared to 500 in NS. For chromosome 10, which is regularly deleted in glioma, the relationship is inverted with 407 GNS enriched loci compared to 3,541 in NS. While this strictly represents the reality that genes and regulatory regions residing on chromosome 7 tend to be more accessible in GNS due to the increased copy number, loci that are a selectively made relatively inaccessible may be difficult to identify without compensatory larger fold changes. The effects of copy number variation may be compensated by attempts to accurately quantify copy number or simply by estimating the library size for differential analysis in sub-chromosome windows. Attempts to adjust for copy number variation will however depend on the accuracy of genomic copy number estimates.

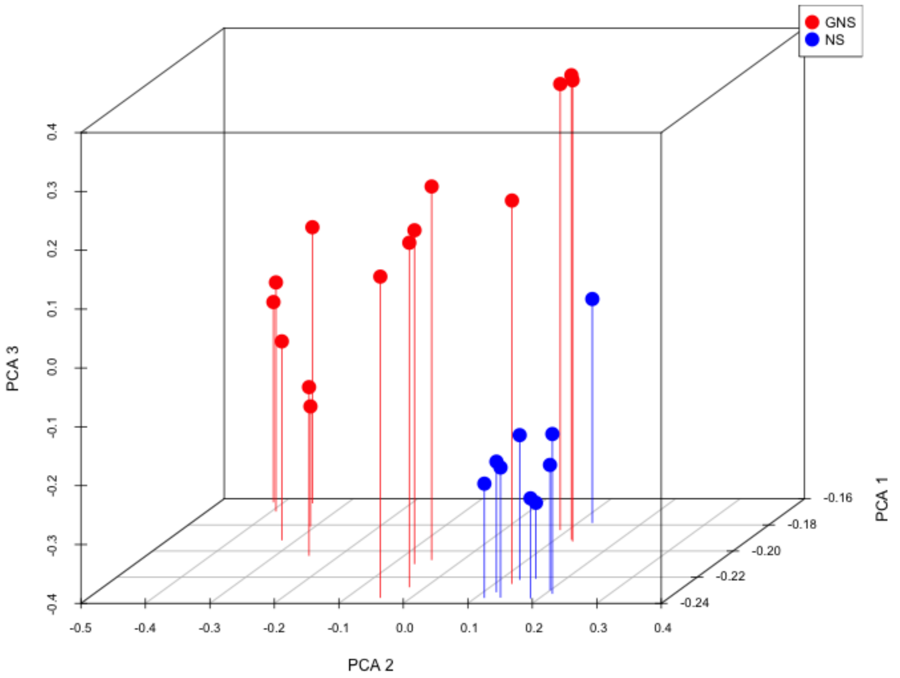


Fig. 5.7 Principal component analysis of GNS and NS CQN normalised log2 TBS counts. NS lines cluster together as a tight group with the exception of the U3 p14 library. GNS libraries appear to be responsible for a majority of the variation across this dataset as may be expected from their transcriptional and genomic diversity.

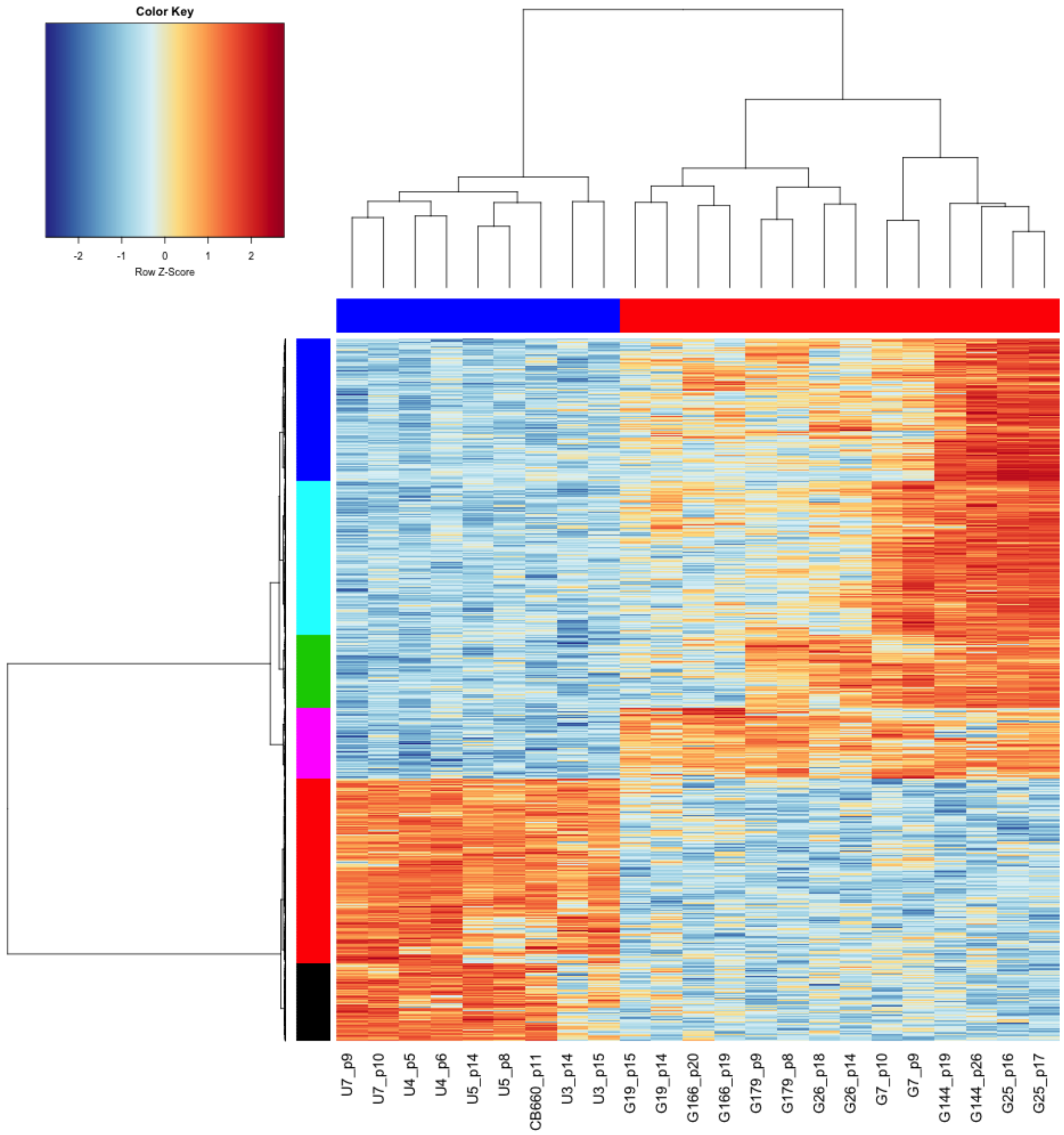


Fig. 5.8 Heatmap of the top 1000 high fold-change normalised loci counts differentially accessible between GNS and NS ($p \leq 0.05$). NS enriched loci show relative consistency in comparison to the diversity seen in GNS enriched loci. This pattern of NS consistency and GNS diversity matches the expression profiles of these lines (Figures 4.3 and 7.8). Units used in the heatmap represent row mean subtracted, CQN normalised log2 TBS counts.

Visualisation of high fold change differentially accessible loci between GNS and NS cells again echoes the expression analysis from Chapter 3 with generally consistent variation between the NS lines and comparably more diversity within the GNS lines (Figure 5.8). Examining the loci of large fold change genes differentially expressed between GNS and NS lines (Chapter 4, Figure 4.1) reveals differentially accessibility of these genes transcription start sites (Table 5.1). Those genes with larger expression fold changes are more likely to have a significantly differentially accessible TSSs. These results reinforce the importance of these gene sets in differentiating between GNS and NS cells but also reveal a consistency of results originating from both expression analysis and ATAC-seq.

| Overexpressed in GNS | Differential TSS in GNS | Overexpressed in NS | Differential TSS in NS |
|---|---|---|---|
| RNF114 | ✓ | ANO4 | ✗ |
| FOXG1 | ✓ | EPB41L3 | ✓ |
| TFAP2A | ✓ | CELSR1 | ✗ |
| GET4 | ✓ | REC8 | ✓ |
| MR1 | ✓ | IPO5 | ✓ |
| HOXD10 | ✗ | NEDD4 | ✗ |
| FAM102A | ✓ | MGST1 | ✓ |
| NKX2-2 | ✓ | SORBS2 | ✗ |
| TGFA | ✗ | ACTA2 | ✓ |
| APCDD1 | ✗ | TAGLN | ✓ |
| HPSE | ✗ | AP3M1 | ✗ |
| CD82 | ✗ | FAM204A | ✗ |
| NUDCD3 | ✓ | SYT1 | ✓ |
| NMNAT3 | ✓ | NHLRC2 | ✗ |
| MITF | ✓ | GFRA1 | ✓ |
| SHOX2 | ✓ | GNG12 | ✗ |
| PMS2P3 | ✓ | PDLIM1 | ✗ |
| NR1D1 | ✗ | BASP1 | ✓ |
| KATNAL2 | ✗ | ME1 | ✓ |
| LMO4 | ✗ | RGMB | ✗ |
| SNAI2 | ✗ | FAM160B1 | ✗ |
| HIST2H2BF | ✗ | PPP3CB | ✗ |
| TBC1D8 | ✗ | RNF182 | ✗ |
| NID2 | ✗ | KPNA3 | ✗ |
| DNAH9 | ✗ | TUSC3 | ✓ |
| MRM1 | ✗ | MYO1B | ✓ |
| SKAP2 | ✗ | H2AFY2 | ✓ |
| FAM220A | ✓ | ZFAND4 | ✓ |
| GLUL | ✗ | TLE4 | ✓ |
| CLDN15 | ✓ | PHKB | ✗ |
| WDR91 | ✓ | FARP1 | ✗ |
| SIX1 | ✗ | WDFY3 | ✗ |
| HOXC6 | ✗ | ATE1 | ✗ |
| TRRAP | ✓ | SUPT3H | ✗ |
| PCDHB3 | ✗ | RB1 | ✗ |
| SLC47A1 | ✗ | LOX | ✗ |
| ANAPC2 | ✗ | GLUD1 | ✗ |
| C10orf90 | ✗ | INPP5F | ✗ |
| GRB10 | ✓ | SHOC2 | ✗ |
| MOCS1 | ✗ | KIF1BP | ✗ |

Table 5.1 Transcription start site (TSS) accessibility in genes differentially expressed between GNS and NS. Blue cell checkmarks indicate that a differentially more accessible loci for that cell type, overlaps with the respective gene's TSS. More significantly differentially expressed, and high fold change, genes are more likely to also have a significantly differentially accessible TSSs.

In order to directly compare ATAC-seq and gene expression datasets, TBS counts from 1kb upstream and 200bp downstream of transcription start sites for genes identified as differentially expressed in the microarray dataset. Comparing log fold change intensity and normalised ATAC-seq transposase accessibility ratio between GNS and NS samples reveals a strong relationship between expression and chromatin accessibility (Figure 5.9, F-test $p < 2.22 \times 10^{-16}$). This relationship is far from determinative with a large proportion of differentially expressed genes showing little change in accessibility between cell types. Moreover many genes overexpressed in one condition are more accessible in the opposite condition (i.e. More expressed in GNS yet more accessible in NS). This reflects what we would expect from the complex process of gene regulation alongside cancer associated copy number changes.

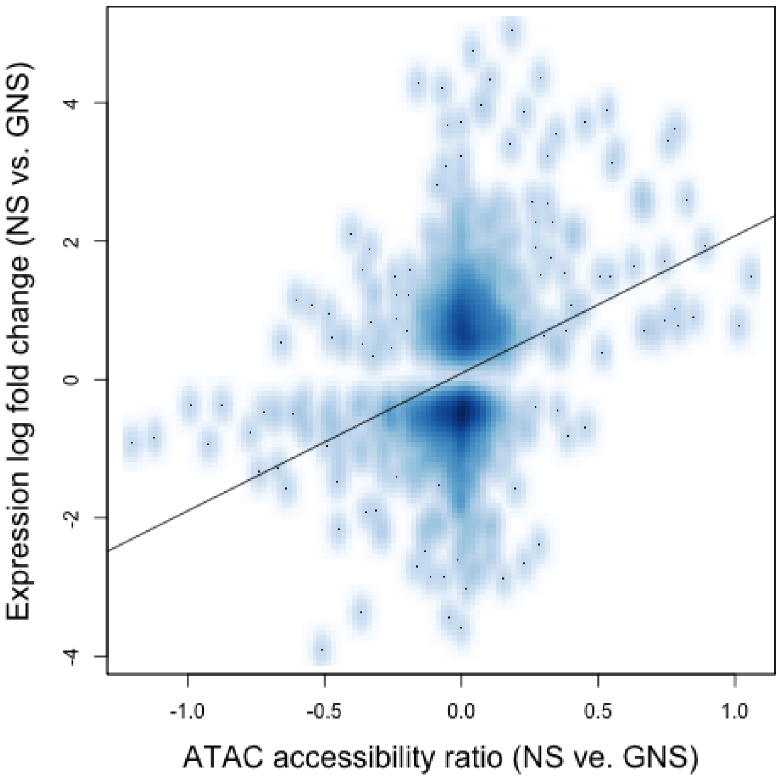


Fig. 5.9 Relationship between expression fold change and ATAC-seq TSS accessibility between NS and GNS samples. Higher values on both axes indicate higher expression or accessibility in NS samples. Regression fit is illustrated by a black line (F-test p < $2.22 \times 10^{-16}$, $R^2$ = 0.09). Expression units are the log fold change of microarray log intensity and ATAC-seq units represent log normalised TBS count ratio between mean NS and GNS samples. Units used in the for accessibility are DESeq2 derived, CQN normalised log2 fold change TBSs (X axis) and log2 fold change intensity units for the expression values (Y axis).

### 5.3.2 ATAC-seq GNS subtype analysis reveals proneural and mesenchymal associated differences in chromatin accessibility

As with gene expression profiles of GNS cells, extensive variation within the ATAC-seq accessibility profiles for different GNS cell lines is apparent. Clustering both with NS cells (Figure 5.8) and in GNS libraries alone Figure 5.10 reveals that GNS lines can be separated into two major groups. The sample to group assignment of these lines largely mirrors the expression based clustering into proneural and mesenchymal GNS lines (Chapter 3) with the exception of G25. In the expression analysis G25 clusters with the mesenchymal lines however in the ATAC-seq profile this line clusters with the proneural lines G7 and G144.

A subset of loci used in the GNS clustering are highly enriched within G25 compared to the other GNS lines. These loci are spaced across 17.2Mb of chromosome 12 including the frequently amplified *CDK4* and appears to represent significant amplification of this region beyond the level of any amplification found in other GNS lines making these loci some of the most variable in this dataset.

Comparing between the proneural and mesenchymal-like lines, the loci differentially accessible between these two subtypes were identified. This analysis identified 21,273 loci more accessible in proneural-like lines and 18,586 loci more accessible in the mesenchymal-like lines. Examining the transcription start sites for consensus proneural or mesenchymal genes (Table 5.2) finds subtype specific differential accessibility. Further examination of GNS proneural and mesenchymal coexpression module genes described in Chapter 3 reveals a clear enrichement for differential accessibility of proneural loci at proneural gene transcription start sites 5.3. The association of mesenchymal accessible chromatin with mesenchymal module genes is comparatively poor which could be explained by the relative diversity within the mesenchymal subtype (Figure 5.7). Comparing between expression and ATAC-accessibility for the GNS proneural and mesenchymal subtypes reveals a similar significant positive relationship (Figure 5.11, F-test $p < 2.22 \times 10^{-16}$) to that between GNS and NS samples (Figure 5.9).

### 5.3.3 Transcription factor motif enrichment in differentially accessible chromatin

Maintenance and specification of gene expression programs and developmental lineages are at least in part controlled by the binding of sequence specific transcription factors (TFs) to suitable loci regulating gene expression. Identification of loci readily accessible to TF binding may be enriched for their associated TF DNA motifs inferring a regulatory role. The MEME-suite tool AME [169] was applied to sequences extracted from loci differentially accessible between GNS and NS to identify known TF motifs relatively enriched within each cell type compared to all accessible loci. Comparing motif enrichment in GNS and NS finds 1,003 motifs for GNS and 11 for NS (Tables 7.10 and 7.11). All motifs found to be enriched in NS cells are minor variants of the AP-1 transcription factor motif. Motifs enriched in GNS loci include a majority of large developmental TF families including, in alphabetical order, CEBP, E2F, ELF, ELK, ETS, FOX, GATA, HMGA, HOX, MEF, NKX, PAX, POU, SOX, and STAT. Many of the highest ranked GNS enriched motifs are variants of the forkhead motif with 13 of the top 20 motifs associated with various FOX/forkhead

| Consensus proneural | Differential TSS | Consensus mesenchymal | Differential TSS |
|---|---|---|---|
| ACSL6 | ✗ | ABCC3 | ✓ |
| ATCAY | ✓ | ANXA2 | ✗ |
| BCAN | ✓ | C1R | ✗ |
| DLL1 | ✗ | C1RL | ✗ |
| DLL3 | ✓ | CCL2 | ✓ |
| GNAO1 | ✗ | CFI | ✓ |
| GRIA2 | ✓ | COL6A2 | ✗ |
| MAP2 | ✓ | GLIPR1 | ✓ |
| MYT1 | ✗ | HFE | ✗ |
| NCAM1 | ✗ | LOXL1 | ✓ |
| OLIG1 | ✓ | MYOF | ✗ |
| OLIG2 | ✓ | PTRF | ✗ |
| PHYHIPL | ✓ | SERPINE1 | ✗ |
| RUNDC3A | ✓ | THBS1 | ✗ |
| SCG3 | ✓ | TMBIM1 | ✗ |
| SEPT3 | ✗ | | |
| SEZ6L | ✓ | | |
| SHD | ✗ | | |
| SOX6 | ✗ | | |
| ZDHHC22 | ✓ | | |

Table 5.2 Consensus proneural and mesenchymal genes and TSS accessibility status. Consensus proneural genes are highly likely to have a more accessible TSS in proneural lines with 13 out of 16 genes identified. Consensus mesenchymal genes and TSS accessibility is comparatively poorly associated with only 7 out of 23 identified. The influence of glioma copy number variation may explain this reduced association with expression.

transcription factors. Transcription factor motifs enriched in loci differentially accessible between the proneural and mesenchymal-like lines were also identified. A total of 214 TF motifs were found to be significantly enriched within the proneural-like accessible loci and for the mesenchymal counterpart loci, a total of 492 motifs were identified. Proneural enriched motifs include members of the forkhead family Foxl1, Foxo1, Foxo3, Foxp2 and Foxq1, amongst other developmentally significant TF familes such as Sox, Hox and Pou to highlight a few (Table 7.12). Meanwhile the mesenchymal enriched TFs include the TF families CEBP, AP-1, Fox, Hox, Irf, Pou and Sox (Table 7.13). One of the few motifs sets enriched within the NS lines attributed to the AP-1 TF associated proteins are also enriched within the mesenchymal-like GNS lines suggesting low AP-1 motif frequency may be a feature of proneural GNS lines instead.

The identification of motifs that are found in differentially accessible loci does not quantitatively describe the comparative presence of TF motifs but rather helps to inform whether a motif could be identified as having a potential role in that cell type. Instead the locations of motif instances were identified within regions of open chromatin and the frequency of mo-

| Primary proneural GNS module | Differential TSS | Primary mesenchymal GNS module | Differential TSS |
|---|---|---|---|
| CCND2 | ✗ | EDIL3 | ✗ |
| C1orf61 | ✗ | TMEFF2 | ✗ |
| ASCL1 | 🟪 ✓ | THBS1 | ✗ |
| GRIA2 | 🟪 ✓ | CCL2 | 🟥 ✓ |
| GRIK3 | 🟪 ✓ | GLIPR1 | 🟥 ✓ |
| BCAN | 🟪 ✓ | EFEMP1 | ✗ |
| AGT | ✗ | PRICKLE1 | 🟪 ✓ |
| CHRDL1 | ✗ | CFI | 🟥 ✓ |
| NCAN | ✗ | CTGF | ✗ |
| SEZ6L | 🟪 ✓ | ABCC3 | 🟥 ✓ |
| LHFPL3 | 🟪 ✓ | SERPINE1 | ✗ |
| PTPRZ1 | 🟪 ✓ | LIPG | ✗ |
| WSCD1 | ✗ | CAV1 | ✗ |
| ADCYAP1R1 | 🟪 ✓ | MYOF | ✗ |
| DCX | ✗ | SDC2 | ✗ |
| SCG3 | 🟪 ✓ | CYR61 | ✗ |
| MIR4697HG | ✗ | PLK2 | ✗ |
| GABRQ | 🟪 ✓ | FBLN1 | ✗ |
| GAP43 | ✗ | ITGA3 | 🟥 ✓ |
| GPR19 | ✗ | NT5DC3 | ✗ |

Table 5.3 Primary proneural and mesenchymal GNS module genes and TSS accessibility status. Similarly to the consensus proneural and mesenchymal genes, the primary proneural module genes are highly likely to have a significantly more accessible TSS where primary mesenchymal module genes are inconsistently accessible. PRICLE1 has proneural and mesenchymal differential accessibility at alternative TSSs.

tif instances per megabase of differentially accessible DNA found in GNS, NS, proneural and mesenchymal lines were used as a metric to identify motifs associated with each condition. Comparing motif frequencies for each condition versus to the motif frequency found in global accessible DNA reveals that loci differentially accessible between conditions tends to be enriched compared to the background accessible chromatin (Figure 5.12). This general enrichment for motifs may be explained by a relative abundance of promotor proximal and enhancer-like regions within differentially accessible chromatin in comparison to generally accessible loci.

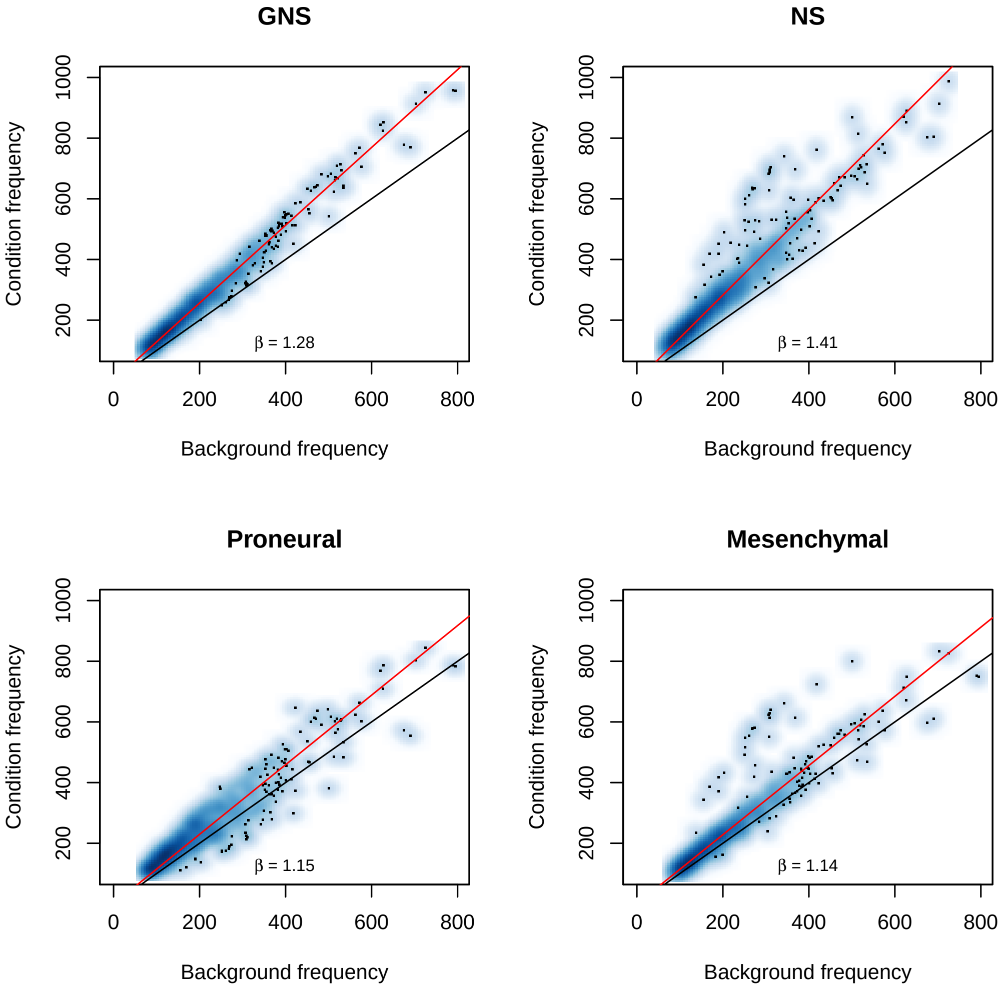


Fig. 5.12 Quantification of motif instances in differentially accessible chromatin versus global accessible chromatin. Differentially accessible loci tend to contain more motif instances compared to the background loci indicating that regulatory regions may be enriched within these differential sets. Linear regression indicates that loci enriched in NS compared to GNS have the highest relative frequency of motif instances which may be explained by an enrichment of genomic amplifications present in the GNS differentially accessible loci. The red line indicates a linear regression fit and the black line represents equal motif frequency between background and the different conditions.

Comparing motif frequency between GNS and NS differentially accessible loci reveals that there are slightly more motif instances per megabase in NS differentially accessible loci compared to the GNS differentially accessible loci (Linear regression, $\beta$ = 1.28 versus 1.41, Figure 5.13). This may be explained by the calling of large amplified regions of DNA as GNS enriched which may contain fewer motif instances than more defined regulatory regions. As the relative frequencies of motifs will be biased by the unique features of each differential accessibility set, motifs associated with each condition were extracted in a ranked fashion in comparison to its control (GNS versus NS and proneural versus mesenchymal). The top 50 motifs for each condition are presented in tables 7.14, 7.15, 7.16 and 7.17. The top 19 motifs enriched within NS loci are variants of the AP-1 motif and form the outlying NS motifs shown in figure 5.13. Remarkably the AP-1 motif also represents the top 28 motifs in mesenchymal loci compared to proneural loci. The motifs with the highest relative frequency in GNS loci compared to NS loci are variants of ZNF238 motif. Other motifs enriched within GNS loci include Tal1, IKZF1, CEBPB and the MEF2 family. In NS loci the TEAD, SOX and Runx family of motifs as well as the abundant AP-1 like motifs are more accessible compared to GNS loci. Motifs that have a higher frequency in mesenchymal enriched loci compared to proneural loci again include the AP-1 motifs as well as OTX2 and PITX2 alongside TEAD, RAR and Runx family motifs reflecting a similarity to NS enriched motifs. The top proneural motifs include ARIA3A and NHP6B motifs alongside Fox and Pou family members.

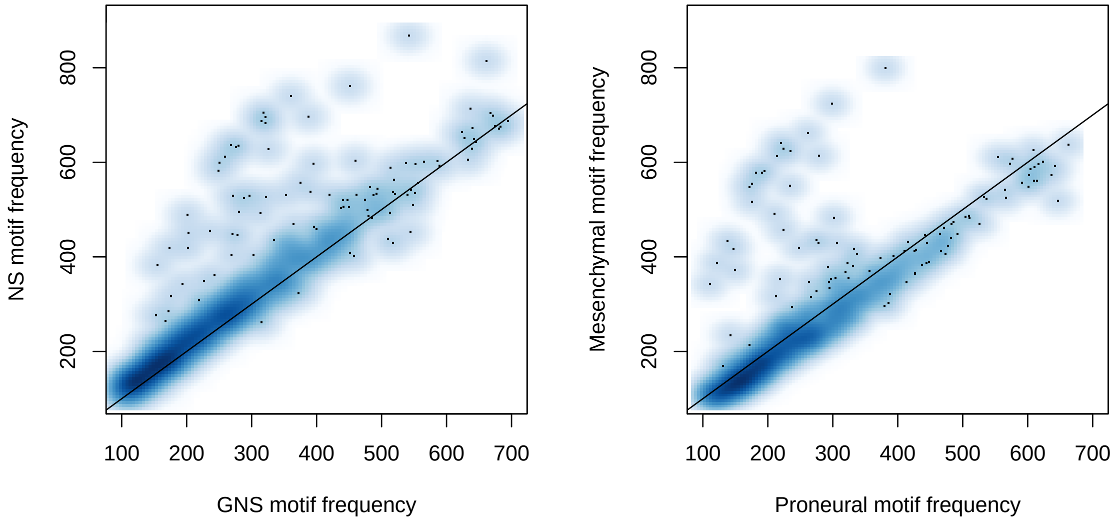


Fig. 5.13 Comparing motif instance frequency in differentially accessible loci identifies motifs enriched in each condition. A subset AP-1 transcription factor motifs are enriched in both tests between NS versus GNS as well as mesenchymal versus proneural. These AP-1 motifs are displayed as the outlying points dispersed above the black line which indicates equality of motif frequency between conditions. Non AP-1 associated motifs with less drastic differences in frequency are described in detail in tables 7.14, 7.15, 7.16 and 7.17.

### 5.3.4 An examination of transcription factor motif footprint profiles

Examination of the profile of accessible chromatin proximal to instances of a transcription factor's motif may reveal functional characteristics or steric footprints. Aggregate accessibility at motif instances of the top 50 ranked motifs enriched in each condition were extracted covering 3kb in either direction from the center of the motif. The average transcription factor accessibility profile (TFAP) across GNS and NS libraries is presented in figure 5.14. TFAPs tend to follow a trend from high accessibility at the motif steric footprint edge and low accessibility ∼1.1kb distal from the motif to low proximal and high distal accessibility. A similar distal/proximal trend is seen in the predicted nucleosome density plots (Figure 5.14, right column). A more detailed view on the proximal accessibility reveals the profile of steric hinderance within ∼30bp of the motif center alongside a reflective symmetry in predicted nucleosome signal (Figure 5.14, bottom row). The TFAPs with high proximal and low distal accessibility have a pronounced predicted nucleosome signal ∼120bp from the center of the motif which may represent an enriched positioning of nucleosomes either side of the transcription factor motif. DNA bound by nucleosomes cover ∼150bp therefore the edge of the nucleosome would sit at ∼45bp sitting close to the ∼30bp steric footprint allowing space for transposes binding. TFAPs with low proximal and high distal accessibility also present a transposition protected steric footprint, however in these profiles the accessibility increases from the motif center to peak at ∼1.1kb distal. The predicted nucleosome signal for these TFAPs suggests nucleosomes may be frequently positioned directly over these motif instances however it is also possible that other DNA binding factors at these sites lead to an enrichment of ∼150kb fragments that are inappropriately inferred to represent nucleosomes. Motifs for known transcription factor families tend to cluster this trend for variable proximal and distal accessibility (Figure 5.15). TFAPs with the highest proximal and lowest distal accessibility are variants of the symetrical AP-1 transcription factor motif. These AP-1 like motifs compose the majority of this class of TFAPs followed by the RUNX and SOX families. The TFAPs with the lowest proximal and highest distal accessibility were attributed to the FOX and MEF2 families of transcription factors. Curiously this proximal/distal trend of AP-1 and RUNX versus FOX and MEF2 factors matches the assocciation of these motifs between GNS and NS enriched loci as well as between proneural and mesenchymal enriched loci (Tables 7.14,7.15,7.16,7.17). This bias for the type of transcription factor motifs that are found in differentially accessible loci may relate to differences in the functional roles of the loci like enhancers, promotors, insulators or silencers.

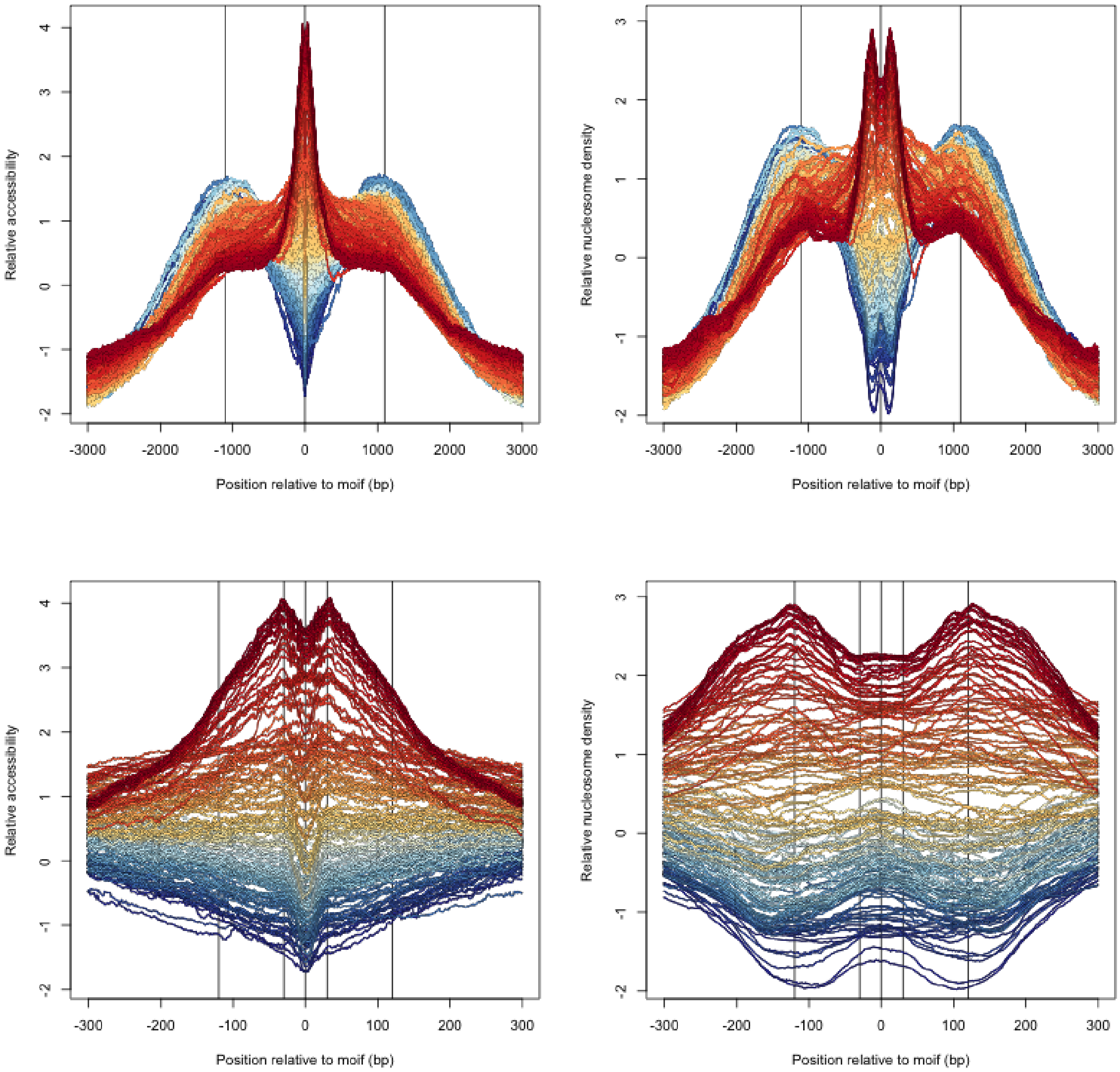


Fig. 5.14 Transcription factor accessibility profiles (TFAPs) follow a trend between proximal and distal enrichment. Examining TFAPs either 3kb (Top row) or 300bp (Bottom row) either side of the motif in both accessibility (Left column) and predicted nucleosome density (Right column). The symmetrical AP-1 motifs make up a majority of the highest proximal and lowest distal accessibility TFAPs (Red). Steric hinderance of each transcription factor protecting its motif site (footprinting) can be seen in the bottom left plot as a less accessible groove in the profile. Nucleosome density estimates suggest high proximal and low distal accessibility TFAPs tend to have nucleosomes directly flanking the motif and low proximal and high distal accessibility TFAPs (Blue) present the inverse. Units used are z-score normalised average TBS density.

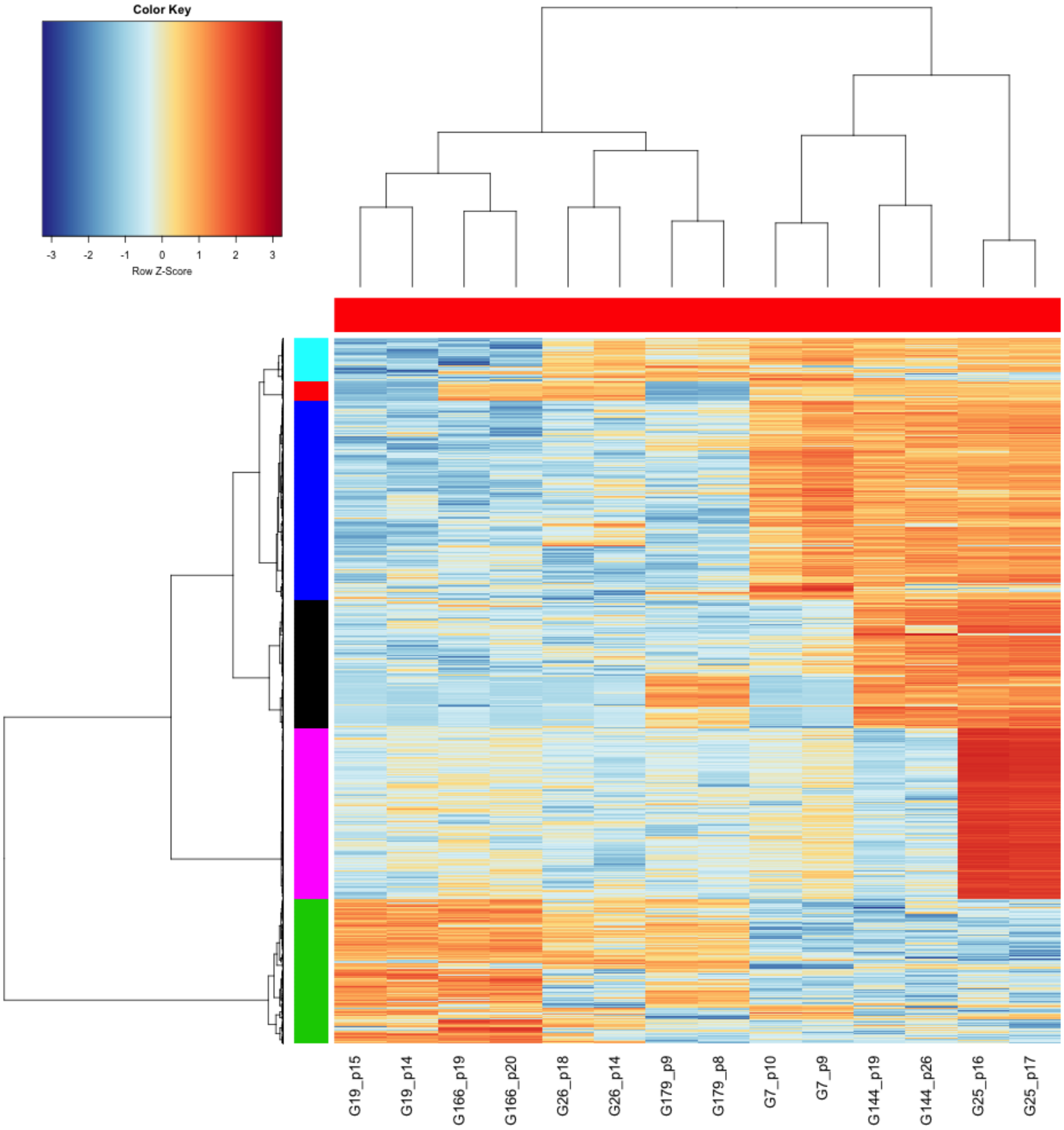


Fig. 5.10 Heatmap clustering of 2,000 highly variable normalised loci counts in GNS cell lines. The proneural like libraries G7, G144 and G25 cluster separately from the more mesenchymal like libraries. A large proportion of the most variable loci map back to a highly enriched 17.2mb section of chromosome 12 in G25. Units within the heatmap are row mean normalised log2 TBS counts normalised with CQN.

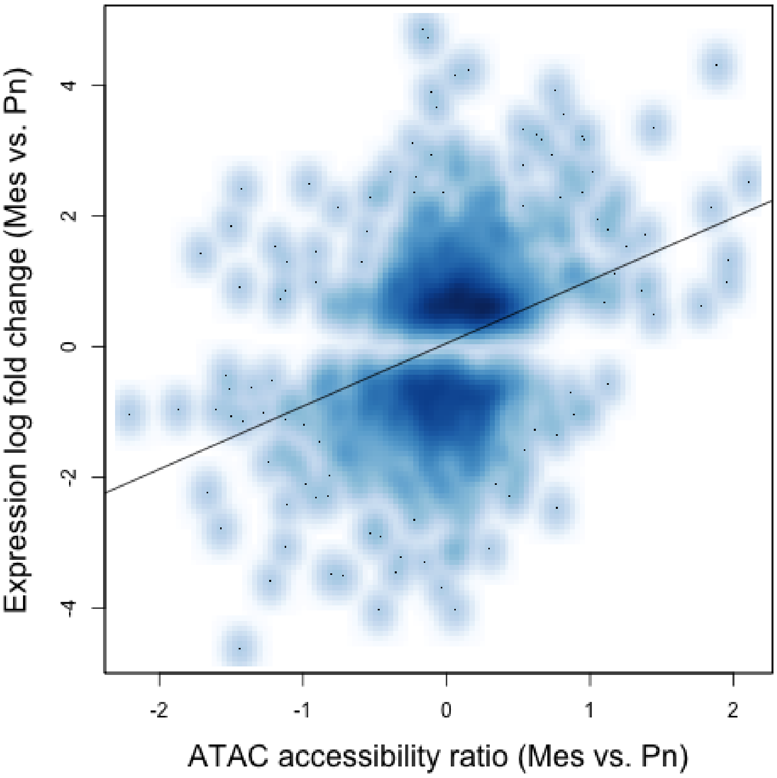
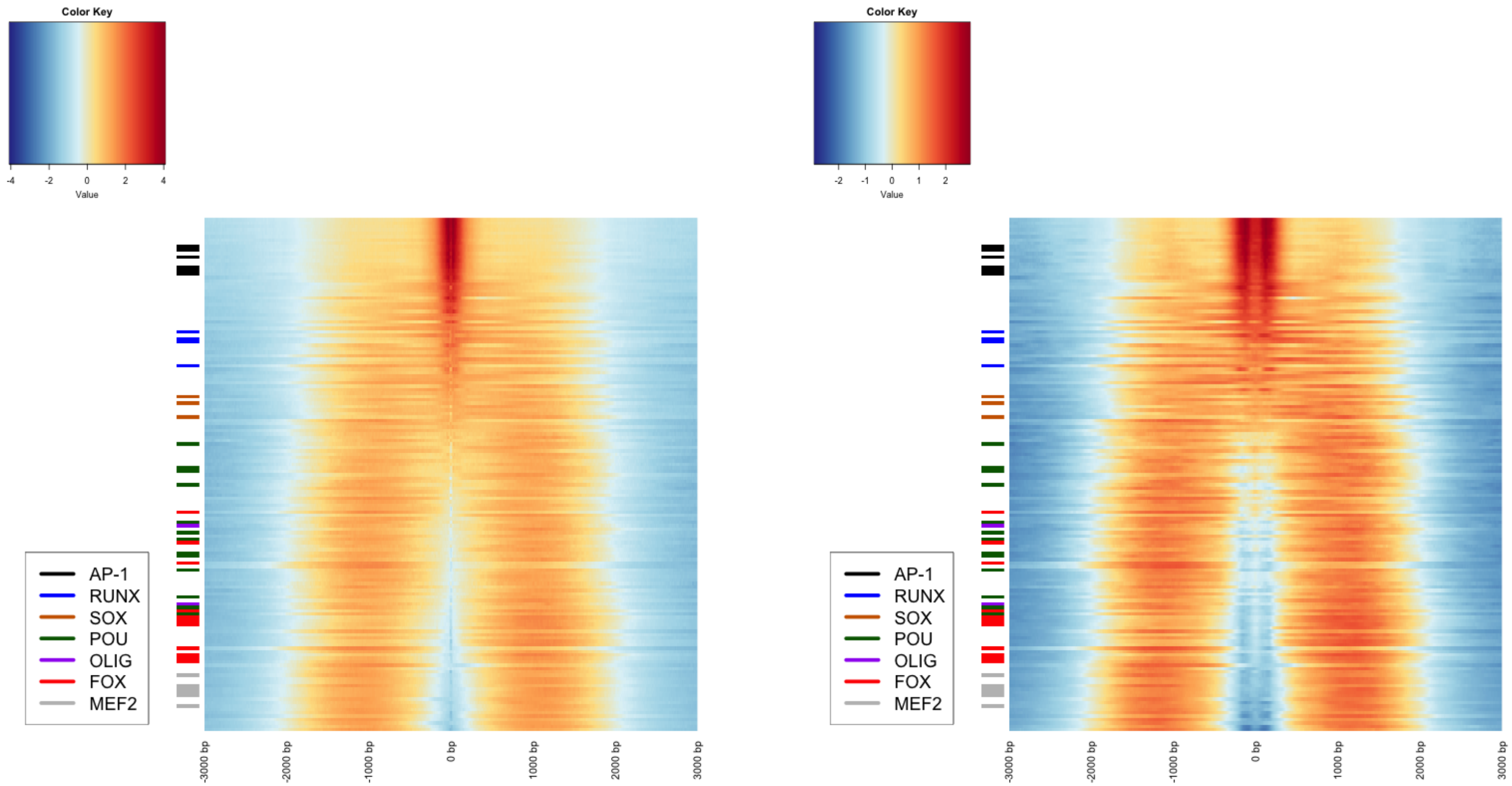


Fig. 5.11 Relationship between expression fold change and ATAC-seq TSS accessibility between GNS subtypes. Higher values on both axes indicate higher expression or accessibility in mesenchymal samples. Regression fit is illustrated by a black line (F-test $p < 2.22 \times 10^{-16}$, $R^2 = 0.11$). Units used in the for accessibility are DESeq2 derived, CQN normalised log2 fold change TBSs (x axis) and log2 fold change intensity units for the expression values (y axis).



(a) Relative accessibility

(b) Relative nucleosome density

Fig. 5.15 Ranked TFAPs reveals consistency between transcription factor families. AP-1 associated motifs group together with high proximal accessibility and MEF2 associated motifs grouping with low proximal accessibility. Units used are z-score normalised average TBS density. Only motifs directly annotated with TF family names are marked by row colours while many motifs that have a sequence similarity to these families are still included without row colour markers.

## 5.4 Discussion

Set out in this chapter is an overview of the methodological challenges apparent in the analysis of ATAC-seq data. Application of these methods to the characterisation of GNS and NS cell lines finds that chromatin accessibility mirrors and further supports the identification of differences between GNS and NS cells as well as the separation of GNS lines into proneural and mesenchymal subtypes. The characterisation of the Tn5 transposase binding site reflects the rotational and reflectional symmetry of the transposase which binds DNA as a dimer. As the transposase has a preferred binding site sequence, methods that integrate this sequence preference may enable more accurate accessibility estimates and improve the identification of enriched open chromatin loci. With the limitation of current tools the GC base bias of the transposase is accounted for by including loci GC estimates as a cofactor in the normalisation of loci accessibility estimates. Differences in paired-end library insert size were shown to have little overall effect on the intra-sample variation. Application of ATAC-seq to the characterisation of GNS and NS cells finds a general consistency of NS line accessibility profiles. GNS lines on the other hand display a diverse range of accessibility variation with GNS lines separating into proneural and mesenchymal profiles. Genes differentially expressed between GNS and NS cells are also shown to have differential TSS accessibility. Extending the analysis to clustering within GNS lines finds comparable proneural and mesenchymal clusters. Genes from proneural and mesenchymal modules are likewise differentially accessible between GNS subtype lines. Analysis of transcription factor motif accessibility finds enrichment for numerous neural development related transcription factors. The most significant differences observed relates to variation in AP-1 transcription factor accessibility. Loci more accessible in both NS versus GNS as well as loci more accessible in mesenchymal versus proneural lines are enriched for AP-1 motifs. This may suggest that proneural accessible loci are depleted for AP-1 motifs rather than the inverse.

# Chapter 6

# Discussion and Outlook

## 6.1 Discussion

Identifying sources of variation both within cancer tissues or between normal and neoplastic cells is a critical process in the understanding of the disease. Variation between tumour samples and by extension, individual patient's tumours have generally been characterised as members of discrete subtype categories. While this discrete subtypes hypothesis has been successful for breast cancer tumours, other tumour types, including glioma, have been difficult to divide into reproducible and distinct categories [166]. This inconsistency of subtype distinctions extends into the analysis of GSCs, where cells removed from the tumour niche unsurprisingly present an independent expression profile to their derived tumour tissue samples. Many have questioned whether GSC cultures can accurately represent the diversity and functional potential of the neoplastic cells found *in vivo* [71, 87, 106, 143, 237, 263]. Set out within this thesis I apply a novel coexpression method to identify features of subtype like expression in tumour samples including a proneural to mesenchymal axis in glioma. I then continue to show how this proneural to mesenchymal axis can be comparably identified in GSCs using both microarray expression and ATAC-seq data. Along side this core finding other results have included an investigation of breast cancer subtypes, ATAC-seq analysis and a comparison of the differences between GNS and NS cell lines.

### Analysis of tumour transcriptomes and cancer subtypes

The analysis of tumour transcriptomes has focused either on discrete categories of cancer subtypes or broad networks of coexpressed genes largely used to associate genes with biological processes. Discrete subtypes are intended to reduce the diversity of tumour expression into a set of profiles that can provide clinical and therapeutic benefits to the patient. In contrast coexpression studies have largely focused on identifying relationships among genes. The observation that common biological processes allow for coexpressed gene networks implies a large component of tumour expression variation can be ascribed to these common features. These coexpression networks represented by independent gradients of expression would therfore be expected consequences for the establishment of discrete subtypes. Application of discrete clustering methods to a smooth continuous gradient will still identify discrete separations of the data regardless of whether the data can be better represented as a gradient. Set out in this thesis I show that these two distinct methodologies, subtype discovery and coexpression, can be combined to reveal reproducible subtypes and infer gene expression that characterises them. Application of the novel coexpression method, correlation marker clustering, was applied to two types of tumours that have previously been the focus of frequent subtype analysis. In breast invasive ductal carcinoma (BRCA) I find coexpression modules representative of luminal and basal like expression in agreement with previous subtype study signatures. Similarly in glioma the primary coexpression modules are representative of the established proneural and mesenchymal subtypes. Where low expression of the luminal module distinguishes basal like samples from the luminal-like samples, expression of the glioma proneural and mesenchymal modules presents a continuous gradient representative of a subtype axis. As all BRCA samples and subtypes present basal module expression (Figure 3.4b), clustering within the basal module reveals that *FOXC1* and *SFRP1* are more highly expressed in basal subtype samples than in the luminal subtype samples that present high basal module expression. This process of clustering within each independent coexpression module enables the identification of independent subtype classifications that relate to established subtypes. The correlation marker subclustering methodology enables the reclassification of established BRCA basal, luminal, claudin-low and Her2-enriched subtypes within independent modules. Basal subtype samples are best distinguished by high *FOXC1* expression relative to other basal module genes. Similarly the claudin-low are distinguished by stromal gene variation and the Her2-enriched subtype by loss of *ESR1* expression and retention of *FOXA1* and *SPDEF* expression.

In comparison clustering within glioma proneural and mesenchymal modules reveals a poor intersection between established subtypes. No discrete clusters of glioma samples are identified mirroring other studies that have suggested the classical and neural subtypes to be absent. Instead glioma variation presents a dominant proneural to mesenchymal axis that is present intratumourally, is associated with immune infiltration, progression and grade. This lack of subtype identification may indicate that within module variation presents a more complex substructure that cannot be summarised by a simple dominant sample clustering. Some established glioma subtypes may be replicated through analysis of coexpression modules at a range of different cut off heights and a diverse range of within module subcluster k divisions. Future work will seek to examine if robust glioma subtype can be extracted using further developed CMC-like methods.

### Analysis of GNS and NS expression

Having investigated the variation found between glioma samples I move forward to examine the variation found with glioma stem cell lines and between GSCs (or GNS cells) and neural stem cells (NS). Comparing between GNS and NS lines I replicate and reinforce analysis by Engström *et al* [55] highlighting overexpression of *FOXG1* and *TFAP2A* in GNS cells alongside Importins and numerous tumour suppressors including *RB1 PTPRB* and *NEDD4* in NS cells. This characterisation of genes associated with either GNS or NS cells may help identify glioma specific features that can be used to develop targeted therapeutics and direct future GSC research alongside reassurance that these cells are a good *in vitro* representation of the disease they are proposed to model. Application of the Verhaak *et al.* classifier [270] to GNS expression profiles demonstrates that tumour derived signatures must be adapted to enable *in vitro* characterisation due to differences in cellular environment. Application of the CMC coexpression method to GNS expression data finds modules enriched for proneural and mesenchymal genes. These GNS proneural and mesenchymal modules separate GNS lines into proneural and mesenchymal clusters. Proneural GNS lines cluster more closely together than the more divergent mesenchymal lines. Mesenchymal modules are enriched for genes associated with immune cell infiltration and inflammation suggests this phenotype is may be induced, or enable, the increased presence of macrophages and other immune cells into the tumour microenvironment. The association of neural developmental transcription factors, perineuronal net and asymmetric division related genes with proneural modules may suggest this phenotype is closest to the glial progenitor cell type in comparison to the mesenchymal GNS cells. Through reanalysis of public data in the context of GNS coexpression modules finds that cells transferred from EGF/FGF growth factor media to serum media shifted towards a mesenchymal phenotype. The factors that lead to this transformation are not well understood. Furthermore its not clear if GNS cells can transition from a mesenchymal to proneural phenotype or if this process can occur during tumour expansion. As proneural GSC have been able to transition to a mesenchymal type, these cells for fill the cancer stem cell ability to propagate tumours with differentiated progeny representative of the parental tumour [136]. If however mesenchymal GSCs are unable to transition to a proneural subtype, mesenchymal GSCs may be better described as cancer propagating cells.

### Application of ATAC-seq for the characterisation of GNS and NS cells

As a method for interrogating chromatin accessibility, ATAC-seq is an exciting and flexible new method [24]. In order to apply this method I discuss technical biases and practical aspects of ATAC-seq analysis. Subsequent application of these methods on a panel of GNS and NS ATAC-seq libraries reveals chromatin organisation that mirrors the expression differences between GNS and NS cell lines. Similarly comparative ATAC-seq analysis reveals proneural and mesenchymal subtype specific chromatin organisation. Examination of transcription factor motif enrichment finds an over representation of AP-1 motifs in both NS lines compared to GNS and in mesenchymal lines versus proneural lines. Combination of ATAC-seq data with other high-throughput 'omics methods like ChIP-seq, RNA-seq and 4C-seq will improve efforts to comprehensively integrate different components of transcriptional regulation.

## 6.2 Outlook and potential future work

### Tumour subtype analysis

The application of CMC to simultaneously profile both samples and genes in independent modules to other tumour types and other large expression datasets may provide further valuable insight. Provision of CMC sample, gene and module information as a publicly accessible database could become a useful resource for both research and clinical usage. Association of clinical factors like drug efficacy to coexpression modules and sample subclusters may extend the reach of personalised medicine. For research the association of genes in CMC modules could be used in a similar way to gene ontology terms, annotating genes with coexpression features. Itterative improvement of CMC-like methods may assist in clarifying generally inconsistant tumour expression subtypes like GBM.

### GNS analysis

For GNS and GSC biology a number of important questions become apparent following the results set out in this thesis. The proneural and mesenchymal phenotypes of GNS lines are of unknown importance to glioma biology. While these profiles clearly relate to glioma proneural and mesenchymal subtypes, further work is required to profile their functional and clinical significance. While the proneural to mesenchymal axis is present within individual tumours, it has yet to be shown that both phenotypes of GNS cell could be extracted from each tumour. Having established that proneural like cells can be forced towards a mesenchymal phenotype, the factors present in serum that enable this transition are unknown. Similarly its unknown whether mesenchymal like cells can be converted into a proneural phenotype by any viable experimental methods. If proneural transition can be achieved, what processes are required to force NS cells into proneural or mesenchymal phenotypes. Another aspect to follow up is subtype specific differences in asymmetric division with proneural expression of *CCND2* and noted quiescent GSC populations and clonogenic capacity mediated by cell-cell contact [28]. It may be possible to identify both proneural and mesenchymal GSCs within individual tumours. The derivation and characterisation of these GSCs within individual tumours would reveal a switch in GSC type as reflective of the tumour environment and lineage potential.

### ATAC-seq analysis

As a relatively new method, many of the unique attributes and opportunities to exploit ATAC-seq data have yet to be explored. A significant improvement in the accuracy of ATAC-seq accessibility estimates could be achieved by including the Tn5 transposase's binding site sequence preferences as a cofactor. Many stages of ATAC-seq analysis could benefit from this including peak-calling, transcription factor footprinting, accessibility estimation and normalisation. Another area that could be improved is the provision of software tools for ATAC-seq analysis. Toolsets like Bedtools [209] and Samtools [146] are essential components of many analysis pipelines and a comparable ATACtools package would improve the speed and consistency of ATAC-seq data analysis for the bioinformatic community at large. The relationship between transcription factor motifs and transposase accessibility is another area of significant potential. Deeper sequencing of individual libraries and normalisation for transposase bias may enable the estimation of transcription factor binding at individual loci. This could work similarly to ChIP-seq estimates for transcription factor binding with the weakness of a greater reliance on motif accuracy but also a lesser dependence on antibody quality. The other advantage is that many different motifs could be estimated from one library preparation as opposed to a single transcription factor per library in CHIP-seq. ATAC-seq also has a reduced dependence on the quantity of genomic DNA required for each library.

# References

[1] Abaskharoun, M., Bellemare, M., Lau, E., and Margolis, R. U. (2010). Expression of hyaluronan and the hyaluronan-binding proteoglycans neurocan, aggrecan, and versican by neural stem cells and neural cells derived from embryonic stem cells. *Brain research*, 1327:6–15.

[2] Adey, A., Morrison, H. G., Asan, Xun, X., Kitzman, J. O., Turner, E. H., Stackhouse, B., MacKenzie, A. P., Caruccio, N. C., Zhang, X., and Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome biology*, 11(12):R119.

[3] Anido, J., Sáez-Borderías, A., Gonzàlez-Juncà, A., Rodón, L., Folch, G., Carmona, M. A., Prieto-SAnchez, R. M., Barba, I., Martínez-Sáez, E., Prudkin, L., Cuartas, I., Raventós, C., Martínez-Ricarte, F., Poca, M. A., GarcIa-Dorado, D., Lahn, M. M., Yingling, J. M., Rodón, J., Sahuquillo, J., Baselga, J., and Seoane, J. (2010). TGF-$\beta$ Receptor Inhibitors Target the CD44(high)/Id1(high) Glioma-Initiating Cell Population in Human Glioblastoma. *Cancer Cell*, 18(6):655–668.

[4] Assanah, M., Lochhead, R., Ogden, A., Bruce, J., Goldman, J., and Canoll, P. (2006). Glial progenitors in adult white matter are driven to form malignant gliomas by platelet-derived growth factor-expressing retroviruses. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(25):6781–6790.

[5] Baird, A. (1994). Fibroblast growth factors: activities and significance of non-neurotrophin neurotrophic growth factors. *Current opinion in neurobiology*, 4(1):78–86.

[6] Bao, S., Wu, Q., Li, Z., Sathornsumetee, S., Wang, H., McLendon, R. E., Hjelmeland, A. B., and Rich, J. N. (2008). Targeting cancer stem cells through L1CAM suppresses glioma growth. *Cancer Research*, 68(15):6043–6048.

[7] Beier, C. P., Kumar, P., Meyer, K., Leukel, P., Bruttel, V., Aschenbrenner, I., Riemenschneider, M., Fragoulis, A., Rümmele, P., Lamszus, K., Schulz, J. B., Weis, J., Bogdahn, U., Wischhusen, J., Hau, P., Spang, R., and Beier, D. (2012). The Cancer Stem Cell Subtype Determines Immune Infiltration of Glioblastoma. *Stem cells and development*.

[8] Beier, D., Hau, P., Proescholdt, M., Lohmeier, A., Wischhusen, J., Oefner, P. J., Aigner, L., Brawanski, A., Bogdahn, U., and Beier, C. P. (2007). CD133(+) and CD133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles. *Cancer Research*, 67(9):4010–4015.

[9] Ben-Porath, I., Thomson, M. W., Carey, V. J., Ge, R., Bell, G. W., Regev, A., and Weinberg, R. A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature genetics*, 40(5):499–507.

[10] Bernardo, G. M., Bebek, G., Ginther, C. L., Sizemore, S. T., Lozada, K. L., Miedler, J. D., Anderson, L. A., Godwin, A. K., Abdul-Karim, F. W., Slamon, D. J., and Keri, R. A. (2013). FOXA1 represses the molecular phenotype of basal breast cancer cells. *Oncogene*, 32(5):554–563.

[11] Bhat, K. P. L., Balasubramaniyan, V., Vaillant, B., Ezhilarasan, R., Hummelink, K., Hollingsworth, F., Wani, K., Heathcock, L., James, J. D., Goodman, L. D., Conroy, S., Long, L., Lelic, N., Wang, S., Gumin, J., Raj, D., Kodama, Y., Raghunathan, A., Olar, A., Joshi, K., Pelloski, C. E., Heimberger, A., Kim, S. H., Cahill, D. P., Rao, G., Den Dunnen, W. F. A., Boddeke, H. W. G. M., Phillips, H. S., Nakano, I., Lang, F. F., Colman, H., Sulman, E. P., and Aldape, K. (2013). Mesenchymal Differentiation Mediated by NF-$\kappa$B Promotes Radiation Resistance in Glioblastoma. *Cancer Cell*.

[12] Bijlmakers, M.-J., Kanneganti, S. K., Barker, J. N., Trembath, R. C., and Capon, F. (2011). Functional analysis of the RNF114 psoriasis susceptibility gene implicates innate immune responses to double-stranded RNA in disease pathogenesis. *Human molecular genetics*, 20(16):3129–3137.

[13] Boase, N. A. and Kumar, S. (2015). NEDD4: The founding member of a family of ubiquitin-protein ligases. *Gene*, 557(2):113–122.

[14] Bohman, L.-E., Swanson, K. R., Moore, J. L., Rockne, R., Mandigo, C., Hankinson, T., Assanah, M., Canoll, P., and Bruce, J. N. (2010). Magnetic resonance imaging characteristics of glioblastoma multiforme: implications for understanding glioma ontogeny. *Neurosurgery*, 67(5):1319–27– discussion 1327–8.

[15] Borden, E. C., Lindner, D., Dreicer, R., Hussein, M., and Peereboom, D. (2000). Second-generation interferons for cancer: clinical targets. *Seminars in Cancer Biology*, 10(2):125–144.

[16] Borden, E. C., Sen, G. C., Uze, G., Silverman, R. H., Ransohoff, R. M., Foster, G. R., and Stark, G. R. (2007). Interferons at age 50: past, current and future impact on biomedicine. *Nature Reviews Drug Discovery*, 6(12):975–990.

[17] Boyle, A. P., Guinney, J., Crawford, G. E., and Furey, T. S. (2008). F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics*, 24(21):2537–2538.

[18] Bracko, O., Singer, T., Aigner, S., Knobloch, M., Winner, B., Ray, J., Clemenson, G. D., Suh, H., Couillard-Despres, S., Aigner, L., Gage, F. H., and Jessberger, S. (2012). Gene expression profiling of neural stem cells and their neuronal progeny reveals IGF2 as a regulator of adult hippocampal neurogenesis. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(10):3376–3387.

[19] Brennan, C. W., Verhaak, R. G. W., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., Zheng, S., Chakravarty, D., Sanborn, J. Z., Berman, S. H., Beroukhim, R., Bernard, B., Wu, C.-J., Genovese, G., Shmulevich, I., Barnholtz-Sloan, J., Zou, L., Vegesna, R., Shukla, S. A., Ciriello, G., Yung, W. K., Zhang, W., Sougnez, C., Mikkelsen,

T., Aldape, K., Bigner, D. D., Van Meir, E. G., Prados, M., Sloan, A., Black, K. L., Eschbacher, J., Finocchiaro, G., Friedman, W., Andrews, D. W., Guha, A., Iacocca, M., O'Neill, B. P., Foltz, G., Myers, J., Weisenberger, D. J., Penny, R., Kucherlapati, R., Perou, C. M., Hayes, D. N., Gibbs, R., Marra, M., Mills, G. B., Lander, E., Spellman, P., Wilson, R., Sander, C., Weinstein, J., Meyerson, M., Gabriel, S., Laird, P. W., Haussler, D., Getz, G., Chin, L., and TCGA Research Network (2013). The somatic genomic landscape of glioblastoma. *Cell*, 155(2):462–477.

[20] Britto, R., Umesh, S., Hegde, A. S., Hegde, S., Santosh, V., Chandramouli, B. A., and Somasundaram, K. (2007). Shift in AP-2alpha localization characterizes astrocytoma progression. *Cancer biology & therapy*, 6(3):413–418.

[21] Broad Institute of MIT (2015). Broad institute of MIT Genome Data Analysis Center.

[22] Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America*, 101(12):4164–4169.

[23] Buchwalter, G., Hickey, M. M., Cromer, A., Selfors, L. M., Gunawardane, R. N., Frishman, J., Jeselsohn, R., Lim, E., Chi, D., Fu, X., Schiff, R., Brown, M., and Brugge, J. S. (2013). PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. *Cancer Cell*, 23(6):753–767.

[24] Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, 10(12):1213–1218.

[25] Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y., and Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561):486–490.

[26] Busse-Wicher, M., Wicher, K. B., and Kusche-Gullberg, M. (2014). The exostosin family: proteins with many functions. *Matrix biology : journal of the International Society for Matrix Biology*, 35:25–33.

[27] Campbell, A. M., Zagon, I. S., and McLaughlin, P. J. (2013). Astrocyte proliferation is regulated by the OGF-OGFr axis in vitro and in experimental autoimmune encephalomyelitis. *Brain research bulletin*, 90:43–51.

[28] Campos, B., Gal, Z., Baader, A., Schneider, T., Sliwinski, C., Gassel, K., Bageritz, J., Grabe, N., von Deimling, A., Beckhove, P., Mogler, C., Goidts, V., Unterberg, A., Eckstein, V., and Herold-Mende, C. (2014). Aberrant self-renewal and quiescence contribute to the aggressiveness of glioblastoma. *The Journal of pathology*, 234(1):23–33.

[29] Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61–70.

[30] Cancer Genome Atlas Research Network, Brat, D. J., Verhaak, R. G. W., Aldape, K. D., Yung, W. K. A., Salama, S. R., Cooper, L. A. D., Rheinbay, E., Miller, C. R., Vitucci, M., Morozova, O., Robertson, A. G., Noushmehr, H., Laird, P. W., Cherniack,

A. D., Akbani, R., Huse, J. T., Ciriello, G., Poisson, L. M., Barnholtz-Sloan, J. S., Berger, M. S., Brennan, C., Colen, R. R., Colman, H., Flanders, A. E., Giannini, C., Grifford, M., Iavarone, A., Jain, R., Joseph, I., Kim, J., Kasaian, K., Mikkelsen, T., Murray, B. A., O'Neill, B. P., Pachter, L., Parsons, D. W., Sougnez, C., Sulman, E. P., Vandenberg, S. R., Van Meir, E. G., von Deimling, A., Zhang, H., Crain, D., Lau, K., Mallery, D., Morris, S., Paulauskis, J., Penny, R., Shelton, T., Sherman, M., Yena, P., Black, A., Bowen, J., Dicostanzo, K., Gastier-Foster, J., Leraas, K. M., Lichtenberg, T. M., Pierson, C. R., Ramirez, N. C., Taylor, C., Weaver, S., Wise, L., Zmuda, E., Davidsen, T., Demchok, J. A., Eley, G., Ferguson, M. L., Hutter, C. M., Mills Shaw, K. R., Ozenberger, B. A., Sheth, M., Sofia, H. J., Tarnuzzer, R., Wang, Z., Yang, L., Zenklusen, J. C., Ayala, B., Baboud, J., Chudamani, S., Jensen, M. A., Liu, J., Pihl, T., Raman, R., Wan, Y., Wu, Y., Ally, A., Auman, J. T., Balasundaram, M., Balu, S., Baylin, S. B., Beroukhim, R., Bootwalla, M. S., Bowlby, R., Bristow, C. A., Brooks, D., Butterfield, Y., Carlsen, R., Carter, S., Chin, L., Chu, A., Chuah, E., Cibulskis, K., Clarke, A., Coetzee, S. G., Dhalla, N., Fennell, T., Fisher, S., Gabriel, S., Getz, G., Gibbs, R., Guin, R., Hadjipanayis, A., Hayes, D. N., Hinoue, T., Hoadley, K., Holt, R. A., Hoyle, A. P., Jefferys, S. R., Jones, S., Jones, C. D., Kucherlapati, R., Lai, P. H., Lander, E., Lee, S., Lichtenstein, L., Ma, Y., Maglinte, D. T., Mahadeshwar, H. S., Marra, M. A., Mayo, M., Meng, S., Meyerson, M. L., Mieczkowski, P. A., Moore, R. A., Mose, L. E., Mungall, A. J., Pantazi, A., Parfenov, M., Park, P. J., Parker, J. S., Perou, C. M., Protopopov, A., Ren, X., Roach, J., Sabedot, T. S., Schein, J., Schumacher, S. E., Seidman, J. G., Seth, S., Shen, H., Simons, J. V., Sipahimalani, P., Soloway, M. G., Song, X., Sun, H., Tabak, B., Tam, A., Tan, D., Tang, J., Thiessen, N., Triche, T., Van Den Berg, D. J., Veluvolu, U., Waring, S., Weisenberger, D. J., Wilkerson, M. D., Wong, T., Wu, J., Xi, L., Xu, A. W., Yang, L., Zack, T. I., Zhang, J., Aksoy, B. A., Arachchi, H., Benz, C., Bernard, B., Carlin, D., Cho, J., DiCara, D., Frazer, S., Fuller, G. N., Gao, J., Gehlenborg, N., Haussler, D., Heiman, D. I., Iype, L., Jacobsen, A., Ju, Z., Katzman, S., Kim, H., Knijnenburg, T., Kreisberg, R. B., Lawrence, M. S., Lee, W., Leinonen, K., Lin, P., Ling, S., Liu, W., Liu, Y., Liu, Y., Lu, Y., Mills, G., Ng, S., Noble, M. S., Paull, E., Rao, A., Reynolds, S., Saksena, G., Sanborn, Z., Sander, C., Schultz, N., Senbabaoglu, Y., Shen, R., Shmulevich, I., Sinha, R., Stuart, J., Sumer, S. O., Sun, Y., Tasman, N., Taylor, B. S., Voet, D., Weinhold, N., Weinstein, J. N., Yang, D., Yoshihara, K., Zheng, S., Zhang, W., Zou, L., Abel, T., Sadeghi, S., Cohen, M. L., Eschbacher, J., Hattab, E. M., Raghunathan, A., Schniederjan, M. J., Aziz, D., Barnett, G., Barrett, W., Bigner, D. D., Boice, L., Brewer, C., Calatozzolo, C., Campos, B., Carlotti, C. G., Chan, T. A., Cuppini, L., Curley, E., Cuzzubbo, S., Devine, K., DiMeco, F., Duell, R., Elder, J. B., Fehrenbach, A., Finocchiaro, G., Friedman, W., Fulop, J., Gardner, J., Hermes, B., Herold-Mende, C., Jungk, C., Kendler, A., Lehman, N. L., Lipp, E., Liu, O., Mandt, R., McGraw, M., McLendon, R., McPherson, C., Neder, L., Nguyen, P., Noss, A., Nunziata, R., Ostrom, Q. T., Palmer, C., Perin, A., Pollo, B., Potapov, A., Potapova, O., Rathmell, W. K., Rotin, D., Scarpace, L., Schilero, C., Senecal, K., Shimmel, K., Shurkhay, V., Sifri, S., Singh, R., Sloan, A. E., Smolenski, K., Staugaitis, S. M., Steele, R., Thorne, L., Tirapelli, D. P. C., Unterberg, A., Vallurupalli, M., Wang, Y., Warnick, R., Williams, F., Wolinsky, Y., Bell, S., Rosenberg, M., Stewart, C., Huang, F., Grimsby, J. L., Radenbaugh, A. J., and Zhang, J. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *The New England journal of medicine*, 372(26):2481–2498.

[31] Canoll, P. and Goldman, J. E. (2008). The interface between glial progenitors and gliomas. *Acta neuropathologica*, 116(5):465–477.

[32] Carulli, D., Pizzorusso, T., Kwok, J. C. F., Putignano, E., Poli, A., Forostyak, S., Andrews, M. R., Deepa, S. S., Glant, T. T., and Fawcett, J. W. (2010). Animals lacking link protein have attenuated perineuronal nets and persistent plasticity. *Brain : a journal of neurology*, 133(Pt 8):2331–2347.

[33] Cautain, B., Hill, R., de Pedro, N., and Link, W. (2015). Components and regulation of nuclear transport processes. *The FEBS journal*, 282(3):445–462.

[34] Ceccarelli, M., Barthel, F. P., Malta, T. M., Sabedot, T. S., Salama, S. R., Murray, B. A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S. M., Anjum, S., Wang, J., Manyam, G., Zoppoli, P., Ling, S., Rao, A. A., Grifford, M., Cherniack, A. D., Zhang, H., Poisson, L., Carlotti, Jr, C. G., da Cunha Tirapelli, D. P., Rao, A., Mikkelsen, T., Lau, C. C., Yung, W. K. A., Rabadan, R., Huse, J., Brat, D. J., Lehman, N. L., Barnholtz-Sloan, J. S., Zheng, S., Hess, K., Rao, G., Meyerson, M., Beroukhim, R., Cooper, L., Akbani, R., Wrensch, M., Haussler, D., Aldape, K. D., Laird, P. W., Gutmann, D. H., Network, T. R., Anjum, S., Arachchi, H., Auman, J. T., Balasundaram, M., Balu, S., Barnett, G., Baylin, S., Bell, S., Benz, C., Bir, N., Black, K. L., Bodenheimer, T., Boice, L., Bootwalla, M. S., Bowen, J., Bristow, C. A., Butterfield, Y. S. N., Chen, Q.-R., Chin, L., Cho, J., Chuah, E., Chudamani, S., Coetzee, S. G., Cohen, M. L., Colman, H., Couce, M., D'Angelo, F., Davidsen, T., Davis, A., Demchok, J. A., Devine, K., Ding, L., Duell, R., Elder, J. B., Eschbacher, J. M., Fehrenbach, A., Ferguson, M., Frazer, S., Fuller, G., Fulop, J., Gabriel, S. B., Garofano, L., Gastier-Foster, J. M., Gehlenborg, N., Gerken, M., Getz, G., Giannini, C., Gibson, W. J., Hadjipanayis, A., Hayes, D. N., Heiman, D. I., Hermes, B., Hilty, J., Hoadley, K. A., Hoyle, A. P., Huang, M., Jefferys, S. R., Jones, C. D., Jones, S. J. M., Ju, Z., Kastl, A., Kendler, A., Kim, J., Kucherlapati, R., Lai, P. H., Lawrence, M. S., Lee, S., Leraas, K. M., Lichtenberg, T. M., Lin, P., Liu, Y., Liu, J., Ljubimova, J. Y., Lu, Y., Ma, Y., Maglinte, D. T., Mahadeshwar, H. S., Marra, M. A., McGraw, M., McPherson, C., Meng, S., Mieczkowski, P. A., Miller, C. R., Mills, G. B., Moore, R. A., Mose, L. E., Mungall, A. J., Naresh, R., Naska, T., Neder, L., Noble, M. S., Noss, A., O'Neill, B. P., Ostrom, Q. T., Palmer, C., Pantazi, A., Parfenov, M., Park, P. J., Parker, J. S., Perou, C. M., Pierson, C. R., Pihl, T., Protopopov, A., Radenbaugh, A., Ramirez, N. C., Rathmell, W. K., Ren, X., Roach, J., Robertson, A. G., Saksena, G., Schein, J. E., Schumacher, S. E., Seidman, J., Senecal, K., Seth, S., Shen, H., Shi, Y., Shih, J., Shimmel, K., Sicotte, H., Sifri, S., Silva, T., Simons, J. V., Singh, R., Skelly, T., Sloan, A. E., Sofia, H. J., Soloway, M. G., Song, X., Sougnez, C., Souza, C., Staugaitis, S. M., Sun, H., Sun, C., Tan, D., Tang, J., Tang, Y., Thorne, L., Trevisan, F. A., Triche, T., Van Den Berg, D. J., Veluvolu, U., Voet, D., Wan, Y., Wang, Z., Warnick, R., Weinstein, J. N., Weisenberger, D. J., Wilkerson, M. D., Williams, F., Wise, L., Wolinsky, Y., Wu, J., Xu, A. W., Yang, L., Yang, L., Zack, T. I., Zenklusen, J. C., Zhang, J., Zhang, W., Zhang, J., Zmuda, E., Noushmehr, H., Iavarone, A., and Verhaak, R. G. W. (2016). Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell*, 164(3):550–563.

[35] Chen, H.-M., Yu, K., Tang, X.-Y., Bao, Z.-S., Jiang, T., Fan, X.-L., Chen, X.-W., and Su, X.-D. (2015). Enhanced expression and phosphorylation of the MET oncoprotein by glioma-specific PTPRZ1-MET fusions. *FEBS letters*, 589(13):1437–1443.

[36] Chen, X., Corbin, J. M., Tipton, G. J., Yang, L. V., Asch, A. S., and Ruiz-Echevarría, M. J. (2014a). The TMEFF2 tumor suppressor modulates integrin expression, RhoA

activation and migration of prostate cancer cells. *Biochimica et biophysica acta*, 1843(6):1216–1224.

[37] Chen, X., Iliopoulos, D., Zhang, Q., Tang, Q., Greenblatt, M. B., Hatziapostolou, M., Lim, E., Tam, W. L., Ni, M., Chen, Y., Mai, J., Shen, H., Hu, D. Z., Adoro, S., Hu, B., Song, M., Tan, C., Landis, M. D., Ferrari, M., Shin, S. J., Brown, M., Chang, J. C., Liu, X. S., and Glimcher, L. H. (2014b). XBP1 promotes triple-negative breast cancer by controlling the HIF1$\alpha$ pathway. *Nature*, 508(7494):103–107.

[38] Choi, E. Y., Chavakis, E., Czabanka, M. A., Langer, H. F., Fraemohs, L., Economopoulou, M., Kundu, R. K., Orlandi, A., Zheng, Y. Y., Prieto, D. A., Ballantyne, C. M., Constant, S. L., Aird, W. C., Papayannopoulou, T., Gahmberg, C. G., Udey, M. C., Vajkoczy, P., Quertermous, T., Dimmeler, S., Weber, C., and Chavakis, T. (2008). Del-1, an endogenous leukocyte-endothelial adhesion inhibitor, limits inflammatory cell recruitment. *Science*, 322(5904):1101–1104.

[39] Choi, E. Y., Lim, J.-H., Neuwirth, A., Economopoulou, M., Chatzigeorgiou, A., Chung, K.-J., Bittner, S., Lee, S.-H., Langer, H., Samus, M., Kim, H., Cho, G.-S., Ziemssen, T., Bdeir, K., Chavakis, E., Koh, J.-Y., Boon, L., Hosur, K., Bornstein, S. R., Meuth, S. G., Hajishengallis, G., and Chavakis, T. (2015). Developmental endothelial locus-1 is a homeostatic factor in the central nervous system limiting neuroinflammation and demyelination. *Molecular psychiatry*, 20(7):880–888.

[40] Chunduru, S., Kawami, H., Gullick, R., Monacci, W. J., Dougherty, G., and Cutler, M. L. (2002). Identification of an alternatively spliced RNA for the Ras suppressor RSU-1 in human gliomas. *Journal of neuro-oncology*, 60(3):201–211.

[41] Clarke, C., Madden, S. F., Doolan, P., Aherne, S. T., Joyce, H., O'Driscoll, L., Gallagher, W. M., Hennessy, B. T., Moriarty, M., Crown, J., Kennedy, S., and Clynes, M. (2013). Correlating transcriptional networks to breast cancer survival: a large-scale co-expression analysis. *Carcinogenesis*, 34(10):2300–2308.

[42] Conti, L., Pollard, S. M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.-L., Cattaneo, E., and Smith, A. (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biology*, 3(9):e283.

[43] Cooper, L. A. D., Gutman, D. A., Chisolm, C., Appin, C., Kong, J., Rong, Y., Kurc, T., Van Meir, E. G., Saltz, J. H., Moreno, C. S., and Brat, D. J. (2012). The Tumor Microenvironment Strongly Impacts Master Transcriptional Regulators and Gene Expression Class of Glioblastoma. *The American journal of pathology*, 180(5):2108–2119.

[44] Crumley, S. M., Divatia, M., Truong, L., Shen, S., Ayala, A. G., and Ro, J. Y. (2013). Renal cell carcinoma: Evolving and emerging subtypes. *World journal of clinical cases*, 1(9):262–275.

[45] Cusulin, C., Chesnelong, C., Bose, P., Bilenky, M., Kopciuk, K., Chan, J. A., Cairncross, J. G., Jones, S. J., Marra, M. A., Luchman, H. A., and Weiss, S. (2015). Precursor States of Brain Tumor Initiating Cell Lines Are Predictive of Survival in Xenografts and Associated with Glioblastoma Subtypes. *Stem cell reports*.

[46] Cutler, M. L., Bassin, R. H., Zanoni, L., and Talbot, N. (1992). Isolation of rsp-1, a novel cDNA capable of suppressing v-Ras transformation. *Molecular and cellular biology*, 12(9):3750–3756.

[47] Danovi, D., Folarin, A., Gogolok, S., Ender, C., Elbatsh, A. M. O., Engstrom, P. G., Stricker, S. H., Gagrica, S., Georgian, A., Yu, D., U, K. P., Harvey, K. J., Ferretti, P., Paddison, P. J., Preston, J. E., Abbott, N. J., Bertone, P., Smith, A., and Pollard, S. M. (2013). A high-content small molecule screen identifies sensitivity of glioblastoma stem cells to inhibition of polo-like kinase 1. *PLoS ONE*, 8(10):e77053.

[48] Dawson, S.-J., Rueda, O. M., Aparicio, S., and Caldas, C. (2013). A new genome-driven integrated classification of breast cancer and its implications. *The EMBO journal*, 32(5):617–628.

[49] Di Fiore, R., D'Anneo, A., Tesoriere, G., and Vento, R. (2013). RB1 in cancer: different mechanisms of RB1 inactivation and alterations of pRb pathway in tumorigenesis. *Journal of cellular physiology*, 228(8):1676–1687.

[50] Doig, T. N., Hume, D. A., Theocharidis, T., Goodlad, J. R., Gregory, C. D., and Freeman, T. C. (2013). Coexpression analysis of large cancer datasets provides insight into the cellular phenotypes of the tumour microenvironment. *BMC Genomics*, 14:469.

[51] Donahue, R. N., McLaughlin, P. J., and Zagon, I. S. (2012). Under-expression of the opioid growth factor receptor promotes progression of human ovarian cancer. *Experimental biology and medicine (Maywood, N.J.)*, 237(2):167–177.

[52] Dwyer, C. A., Bi, W. L., Viapiano, M. S., and Matthews, R. T. (2014). Brevican knockdown reduces late-stage glioma tumor aggressiveness. *Journal of neuro-oncology*, 120(1):63–72.

[53] Eckel-Passow, J. E., Lachance, D. H., Molinaro, A. M., Walsh, K. M., Decker, P. A., Sicotte, H., Pekmezci, M., Rice, T., Kosel, M. L., Smirnov, I. V., Sarkar, G., Caron, A. A., Kollmeyer, T. M., Praska, C. E., Chada, A. R., Halder, C., Hansen, H. M., McCoy, L. S., Bracci, P. M., Marshall, R., Zheng, S., Reis, G. F., Pico, A. R., O'Neill, B. P., Buckner, J. C., Giannini, C., Huse, J. T., Perry, A., Tihan, T., Berger, M. S., Chang, S. M., Prados, M. D., Wiemels, J., Wiencke, J. K., Wrensch, M. R., and Jenkins, R. B. (2015). Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *The New England journal of medicine*, 372(26):2499–2508.

[54] Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25):14863–14868.

[55] Engstrom, P. G., Tommei, D., Stricker, S. H., Ender, C., Pollard, S. M., and Bertone, P. (2012). Digital transcriptome profiling of normal and glioblastoma-derived neural stem cells identifies genes associated with patient survival. *Genome medicine*, 4(10):76.

[56] Enwere, E., Shingo, T., Gregg, C., Fujikawa, H., Ohta, S., and Weiss, S. (2004). Aging results in reduced epidermal growth factor receptor signaling, diminished olfactory neurogenesis, and deficits in fine olfactory discrimination. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 24(38):8354–8365.

[57] Eskan, M. A., Jotwani, R., Abe, T., Chmelar, J., Lim, J.-H., Liang, S., Ciero, P. A., Krauss, J. L., Li, F., Rauner, M., Hofbauer, L. C., Choi, E. Y., Chung, K.-J., Hashim, A., Curtis, M. A., Chavakis, T., and Hajishengallis, G. (2012). The leukocyte integrin antagonist Del-1 inhibits IL-17-mediated inflammatory bone loss. *Nature immunology*, 13(5):465–473.

[58] Fan, X., Khaki, L., Zhu, T. S., Soules, M. E., Talsma, C. E., Gul, N., Koh, C., Zhang, J., Li, Y.-M., Maciaczyk, J., Nikkhah, G., DiMeco, F., Piccirillo, S., Vescovi, A. L., and Eberhart, C. G. (2010). NOTCH pathway blockade depletes CD133-positive glioblastoma cells and inhibits growth of tumor neurospheres and xenografts. *Stem cells (Dayton, Ohio)*, 28(1):5–16.

[59] Feng, J., Han, Q., and Zhou, L. (2012). Planar cell polarity genes, Celsr1-3, in neural development. *Neuroscience bulletin*, 28(3):309–315.

[60] Ferronha, T., Rabadán, M. A., Gil-Guiñon, E., Le Dréau, G., de Torres, C., and Martí, E. (2013). LMO4 is an essential cofactor in the Snail2-mediated epithelial-to-mesenchymal transition of neuroblastoma and neural crest cells. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(7):2773–2783.

[61] Fomchenko, E. I., Dougherty, J. D., Helmy, K. Y., Katz, A. M., Pietras, A., Brennan, C., Huse, J. T., Milosevic, A., and Holland, E. C. (2011). Recruited cells can become transformed and overtake PDGF-induced murine gliomas in vivo during tumor progression. *PLoS ONE*, 6(7):e20605.

[62] Forbes, D. J., Travesa, A., Nord, M. S., and Bernis, C. (2015a). Nuclear transport factors: global regulation of mitosis. *Current opinion in cell biology*, 35:78–90.

[63] Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C. Y., Jia, M., De, T., Teague, J. W., Stratton, M. R., McDermott, U., and Campbell, P. J. (2015b). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research*, 43(Database issue):D805–11.

[64] Friedlander, D. R., Milev, P., Karthikeyan, L., Margolis, R. K., Margolis, R. U., and Grumet, M. (1994). The neuronal chondroitin sulfate proteoglycan neurocan binds to the neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits neuronal adhesion and neurite outgrowth. *The Journal of cell biology*, 125(3):669–680.

[65] Friedmann-Morvinski, D. (2014). Glioblastoma heterogeneity and cancer cell plasticity. *Critical reviews in oncogenesis*, 19(5):327–336.

[66] Frischknecht, R. and Seidenbecher, C. I. (2012). Brevican: a key proteoglycan in the perisynaptic extracellular matrix of the brain. *The international journal of biochemistry & cell biology*, 44(7):1051–1054.

[67] Fu, H., Qi, Y., Tan, M., Cai, J., Takebayashi, H., Nakafuku, M., Richardson, W., and Qiu, M. (2002). Dual origin of spinal oligodendrocyte progenitors and evidence for the cooperative role of Olig2 and Nkx2.2 in the control of oligodendrocyte differentiation. *Development*, 129(3):681–693.

[68] Fujiwara, K., Horiguchi, K., Maliza, R., Tofrizal, A., Batchuluun, K., Ramadhani, D., Syaidah, R., Tsukada, T., Azuma, M., Kikuchi, M., and Yashiro, T. (2015). Expression of the heparin-binding growth factor midkine and its receptor, Ptprz1, in adult rat pituitary. *Cell and tissue research*, 359(3):909–914.

[69] Furnari, F. B., Cloughesy, T. F., Cavenee, W. K., and Mischel, P. S. (2015). Heterogeneity of epidermal growth factor receptor signalling networks in glioblastoma. *Nature reviews. Cancer*, 15(5):302–310.

[70] Gage, F. H., Ray, J., and Fisher, L. J. (1995). Isolation, characterization, and use of stem cells from the CNS. *Annual review of neuroscience*, 18:159–192.

[71] Galli, R., Binda, E., Orfanelli, U., Cipelletti, B., Gritti, A., De Vitis, S., Fiocco, R., Foroni, C., DiMeco, F., and Vescovi, A. (2004). Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Research*, 64(19):7011–7021.

[72] Gao, W.-L., Zhang, S.-Q., Zhang, H., Wan, B., and Yin, Z.-S. (2013). Chordin-like protein 1 promotes neuronal differentiation by inhibiting bone morphogenetic protein-4 in neural stem cells. *Molecular medicine reports*, 7(4):1143–1148.

[73] Ge, W., Martinowich, K., Wu, X., He, F., Miyamoto, A., Fan, G., Weinmaster, G., and Sun, Y. E. (2002). Notch signaling promotes astrogliogenesis via direct CSL-mediated glial gene activation. *Journal of neuroscience research*, 69(6):848–860.

[74] Gerdes, J. (1990). Ki-67 and other proliferation markers useful for immunohistological diagnostic and prognostic evaluations in human malignancies. *Seminars in Cancer Biology*, 1(3):199–206.

[75] 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., and Swanton, C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England journal of medicine*, 366(10):883–892.

[76] Gibson, P., Tong, Y., Robinson, G., Thompson, M. C., Currle, 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., Roussel, M. F., McKinnon, P. J., Gutmann, D. H., Ellison, D. W., Wechsler-Reya, R., and Gilbertson, R. J. (2010). Subtypes of medulloblastoma have distinct developmental origins. *Nature*, 468(7327):1095–1099.

[77] Giotopoulou, N., Valiakou, V., Papanikolaou, V., Dubos, S., Athanassiou, E., Tsezou, A., Zacharia, L. C., and Gkretsi, V. (2015). Ras suppressor-1 promotes apoptosis in breast cancer cells by inhibiting PINCH-1 and activating p53-upregulated-modulator of apoptosis (PUMA); verification from metastatic breast cancer human samples. *Clinical & experimental metastasis*, 32(3):255–265.

[78] Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., and Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research*, 17(6):877–885.

[79] Glabinski, A. R., Balasingam, V., Tani, M., Kunkel, S. L., Strieter, R. M., Yong, V. W., and Ransohoff, R. M. (1996). Chemokine monocyte chemoattractant protein-1 is expressed by astrocytes after mechanical injury to the brain. *Journal of immunology (Baltimore, Md. : 1950)*, 156(11):4363–4368.

[80] Goodenberger, M. L. and Jenkins, R. B. (2012). Genetics of adult glioma. *Cancer genetics*, 205(12):613–621.

[81] Gorup, D., Bohaček, I., Miličević, T., Pochet, R., Mitrečić, D., Križ, J., and Gajović, S. (2015). Increased expression and colocalization of GAP43 and CASP3 after brain ischemic lesion in mouse. *Neuroscience letters*, 597:176–182.

[82] Grange, C., Tapparo, M., Collino, F., Vitillo, L., Damasco, C., Deregibus, M. C., Tetta, C., Bussolati, B., and Camussi, G. (2011). Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. *Cancer Research*, 71(15):5346–5356.

[83] Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics*, 27(7):1017–1018.

[84] Gritti, A., Dal Molin, M., Foroni, C., and Bonfanti, L. (2009). Effects of developmental age, brain region, and time in culture on long-term proliferation and multipotency of neural stem cell populations. *The Journal of comparative neurology*, 517(3):333–349.

[85] Gritti, A., Parati, E. A., Cova, L., Frolichsthal, P., Galli, R., Wanke, E., Faravelli, L., Morassutti, D. J., Roisen, F., Nickel, D. D., and Vescovi, A. L. (1996). Multipotential stem cells from the adult mouse brain proliferate and self-renew in response to basic fibroblast growth factor. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 16(3):1091–1100.

[86] Group, M., Group, M., Co-chairs, committee, W., committee, S., Tissue and clinical data source sites: University of Cambridge/Cancer Research UK Cambridge Research Institute, Agency, B. C. C., University of Nottingham, London, K. C., Manitoba Institute of Cell Biology, Cancer genome/transcriptome characterization centres: University of Cambridge/Cancer Research UK Cambridge Research Institute, Agency, B. C. C., Data analysis subgroup: University of Cambridge/Cancer Research UK Cambridge Research Institute, and Agency, B. C. C. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, pages 1–7.

[87] Günther, H. S., Schmidt, N. O., Phillips, H. S., Kemming, D., Kharbanda, S., Soriano, R., Modrusan, Z., Meissner, H., Westphal, M., and Lamszus, K. (2008). Glioblastoma-derived stem cell-enriched cultures form distinct subgroups according to molecular and phenotypic criteria. *Oncogene*, 27(20):2897–2909.

[88] Guo, F., Lang, J., Sohn, J., Hammond, E., Chang, M., and Pleasure, D. (2015). Canonical Wnt signaling in the oligodendroglial lineage-puzzles remain. *Glia*, 63(10):1671–1693.

[89] Guo, K.-T., Fu, P., Juerchott, K., Motaln, H., Selbig, J., Lah, T., Tonn, J.-C., and Schichor, C. (2014). The expression of Wnt-inhibitor DKK1 (Dickkopf 1) is determined by intercellular crosstalk and hypoxia in human malignant gliomas. *Journal of cancer research and clinical oncology*, 140(8):1261–1270.

[90] Halliday, J., Helmy, K., Pattwell, S. S., Pitter, K. L., LaPlant, Q., Ozawa, T., and Holland, E. C. (2014). In vivo radiation response of proneural glioma characterized by protective p53 transcriptional program and proneural-mesenchymal shift. *Proceedings of the National Academy of Sciences of the United States of America*, 111(14):5248–5253.

[91] Hampton, D. W., Asher, R. A., Kondo, T., Steeves, J. D., Ramer, M. S., and Fawcett, J. W. (2007). A potential role for bone morphogenetic protein signalling in glial cell fate determination following adult central nervous system injury in vivo. *The European journal of neuroscience*, 26(11):3024–3035.

[92] Han, J., Kim, Y.-L., Lee, K.-W., Her, N.-G., Ha, T.-K., Yoon, S., Jeong, S.-I., Lee, J.-H., Kang, M.-J., Lee, M.-G., Ryu, B.-K., Baik, J.-H., and Chi, S.-G. (2013). ZNF313 is a novel cell cycle activator with an E3 ligase activity inhibiting cellular senescence by destabilizing p21(WAF1.). *Cell death and differentiation*, 20(8):1055–1067.

[93] Handl, J., Knowles, J., and Kell, D. B. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15):3201–3212.

[94] Hansen, K. D., Irizarry, R. A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics*, 13(2):204–216.

[95] Heimberger, A. B., McGary, E. C., Suki, D., Ruiz, M., Wang, H., Fuller, G. N., and Bar-Eli, M. (2005). Loss of the AP-2alpha transcription factor is associated with the grade of human gliomas. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 11(1):267–272.

[96] Hemmati, H. D., Nakano, I., Lazareff, J. A., Masterman-Smith, M., Geschwind, D. H., Bronner-Fraser, M., and Kornblum, H. I. (2003). Cancerous stem cells can arise from pediatric brain tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 100(25):15178–15183.

[97] Henikoff, J. G., Belsky, J. A., Krassovsky, K., MacAlpine, D. M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 108(45):18318–18323.

[98] Herschkowitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K. E., Jones, L. P., Assefnia, S., Chandrasekharan, S., Backlund, M. G., Yin, Y., Khramtsov, A. I., Bastein, R., Quackenbush, J., Glazer, R. I., Brown, P. H., Green, J. E., Kopelovich, L., Furth, P. A., Palazzo, J. P., Olopade, O. I., Bernard, P. S., Churchill, G. A., Van Dyke, T., and Perou, C. M. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome biology*, 8(5):R76.

[99] Hesselberth, J. R., Chen, X., Zhang, Z., Sabo, P. J., Sandstrom, R., Reynolds, A. P., Thurman, R. E., Neph, S., Kuehn, M. S., Noble, W. S., Fields, S., and Stamatoyannopoulos, J. A. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nature Methods*, 6(4):283–289.

[100] Hinojosa, A. E., Garcia-Bueno, B., Leza, J. C., and Madrigal, J. L. M. (2011). CCL2/MCP-1 modulation of microglial activation and proliferation. *Journal of neuroinflammation*, 8:77.

[101] Hjelmeland, A. B., Wu, Q., Wickman, S., Eyler, C., Heddleston, J., Shi, Q., Lathia, J. D., Macswords, J., Lee, J., McLendon, R. E., and Rich, J. N. (2010). Targeting A20 decreases glioma stem cell survival and tumor growth. *PLoS Biology*, 8(2):e1000319.

[102] Horie, M., Mitsumoto, Y., Kyushiki, H., Kanemoto, N., Watanabe, A., Taniguchi, Y., Nishino, N., Okamoto, T., Kondo, M., Mori, T., Noguchi, K., Nakamura, Y., Takahashi, E. i., and Tanigami, A. (2000). Identification and characterization of TMEFF2, a novel survival factor for hippocampal and mesencephalic neurons. *Genomics*, 67(2):146–152.

[103] Hu, B., Nandhu, M. S., Sim, H., Agudelo-Garcia, P. A., Saldivar, J. C., Dolan, C. E., Mora, M. E., Nuovo, G. J., Cole, S. E., and Viapiano, M. S. (2012). Fibulin-3 promotes glioma growth and resistance through a novel paracrine regulation of Notch signaling. *Cancer Research*, 72(15):3873–3885.

[104] Huse, J. T., Phillips, H. S., and Brennan, C. W. (2011). Molecular subclassification of diffuse gliomas: seeing order in the chaos. *Glia*, 59(8):1190–1199.

[105] Ida, M., Shuo, T., Hirano, K., Tokita, Y., Nakanishi, K., Matsui, F., Aono, S., Fujita, H., Fujiwara, Y., Kaji, T., and Oohira, A. (2006). Identification and functions of chondroitin sulfate in the milieu of neural stem cells. *The Journal of biological chemistry*, 281(9):5982–5991.

[106] Ignatova, T. N., Kukekov, V. G., Laywell, E. D., Suslov, O. N., Vrionis, F. D., and Steindler, D. A. (2002). Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro. *Glia*, 39(3):193–206.

[107] Ikushima, H., Todo, T., Ino, Y., Takahashi, M., Miyazawa, K., and Miyazono, K. (2009). Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. *Cell Stem Cell*, 5(5):504–514.

[108] Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–264.

[109] Jansson, L. C. and Åkerman, K. E. (2014). The role of glutamate and its receptors in the proliferation, migration, differentiation and survival of neural progenitor cells. *Journal of neural transmission (Vienna, Austria : 1996)*, 121(8):819–836.

[110] Jaworski, D. M., Kelly, G. M., Piepmeier, J. M., and Hockfield, S. (1996). BEHAB (brain enriched hyaluronan binding) is expressed in surgical samples of glioma and in intracranial grafts of invasive glioma cell lines. *Cancer Research*, 56(10):2293–2298.

[111] Jeffrey, P. L., Capes-Davis, A., Dunn, J. M., Tolhurst, O., Seeto, G., Hannan, A. J., and Lin, S. L. (2000). CROC-4: a novel brain specific transcriptional activator of c-fos expressed from proliferation through to maturation of multiple neuronal cell types. *Molecular and cellular neurosciences*, 16(3):185–196.

[112] Jeon, H.-M., Jin, X., Lee, J.-S., Oh, S.-Y., Sohn, Y.-W., Park, H.-J., Joo, K. M., Park, W.-Y., Nam, D.-H., DePinho, R. A., Chin, L., and Kim, H. (2008). Inhibitor of differentiation 4 drives brain tumor-initiating cell genesis through cyclin E and notch signaling. *Genes & Development*, 22(15):2028–2033.

[113] Jin, Y., Han, B., Chen, J., Wiedemeyer, R., Orsulic, S., Bose, S., Zhang, X., Karlan, B. Y., Giuliano, A. E., Cui, Y., and Cui, X. (2014). FOXC1 is a Critical Mediator of EGFR Function in Human Basal-like Breast Cancer. *Annals of surgical oncology*, 21 Suppl 4:758–766.

[114] Johnson, R. A., Wright, K. D., Poppleton, H., Mohankumar, K. M., Finkelstein, D., Pounds, S. B., Rand, V., Leary, S. E. S., White, E., Eden, C., Hogg, T., Northcott, P., Mack, S., Neale, G., Wang, Y.-D., Coyle, B., Atkinson, J., DeWire, M., Kranenburg, T. A., Gillespie, Y., Allen, J. C., Merchant, T., Boop, F. A., Sanford, R. A., Gajjar, A., Ellison, D. W., Taylor, M. D., Grundy, R. G., and Gilbertson, R. J. (2010). Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature*, 466(7306):632–636.

[115] Joo, K. M., Jin, J., Kim, E., Kim, K. H., Kim, Y., Kang, B. G., Kang, Y.-J., Lathia, J. D., Cheng, K. H., Song, P., Kim, H., Seol, H. J., Kong, D.-S., Lee, J.-I., Rich, J. N., Lee, J., and Nam, D.-H. (2012). MET Signaling Regulates Glioblastoma Stem Cells. *Cancer Research*.

[116] Jun, H. J., Bronson, R. T., and Charest, A. (2014). Inhibition of EGFR induces a c-MET-driven stem cell population in glioblastoma. *Stem cells (Dayton, Ohio)*, 32(2):338–348.

[117] Kaneda, A., Matsusaka, K., Aburatani, H., and Fukayama, M. (2012). Epstein-Barr Virus Infection as an Epigenetic Driver of Tumorigenesis. *Cancer Research*.

[118] Karantanos, T., Tanimoto, R., Edamura, K., Hirayama, T., Yang, G., Golstov, A. A., Wang, J., Kurosaka, S., Park, S., and Thompson, T. C. (2014). Systemic GLIPR1-ΔTM protein as a novel therapeutic approach for prostate cancer. *International journal of cancer. Journal international du cancer*, 134(8):2003–2013.

[119] Karasawa, T., Kawashima, A., Usui, F., Kimura, H., Shirasuna, K., Inoue, Y., Komada, T., Kobayashi, M., Mizushina, Y., Sagara, J., and Takahashi, M. (2015). Oligomerized CARD16 promotes caspase-1 assembly and IL-1$\beta$ processing. *FEBS open bio*, 5:348–356.

[120] Kathagen, A., Schulte, A., Balcke, G., Phillips, H. S., Martens, T., Matschke, J., Günther, H. S., Soriano, R., Modrusan, Z., Sandmann, T., Kuhl, C., Tissier, A., Holz, M., Krawinkel, L. A., Glatzel, M., Westphal, M., and Lamszus, K. (2013). Hypoxia and oxygenation induce a metabolic switch between pentose phosphate pathway and glycolysis in glioma stem-like cells. *Acta neuropathologica*, 126(5):763–780.

[121] Kim, E., Kim, M., Woo, D.-H., Shin, Y., Shin, J., Chang, N., Oh, Y. T., Kim, H., Rheey, J., Nakano, I., Lee, C., Joo, K. M., Rich, J. N., Nam, D.-H., and Lee, J. (2013). Phosphorylation of EZH2 Activates STAT3 Signaling via STAT3 Methylation and Promotes Tumorigenicity of Glioblastoma Stem-like Cells. *Cancer Cell*, 23(6):839–852.

[122] Kim, J., Woo, A. J., Chu, J., Snow, J. W., Fujiwara, Y., Kim, C. G., Cantor, A. B., and Orkin, S. H. (2010). A Myc Network Accounts for Similarities between Embryonic Stem and Cancer Cell Transcription Programs. *Cell*, 143(2):313–324.

[123] Kittaneh, M., Montero, A. J., and Glück, S. (2013). Molecular Profiling for Breast Cancer: A Comprehensive Review. *Biomarkers in cancer*, 5:61–70.

[124] Knelson, E. H., Nee, J. C., and Blobe, G. C. (2014). Heparan sulfate signaling in cancer. *Trends in biochemical sciences*, 39(6):277–288.

[125] Kondo, T. and Raff, M. (2000). Oligodendrocyte precursor cells reprogrammed to become multipotential CNS stem cells. *Science*, 289(5485):1754–1757.

[126] Kong, J. H., Yang, L., Dessaud, E., Chuang, K., Moore, D. M., Rohatgi, R., Briscoe, J., and Novitch, B. G. (2015). Notch activity modulates the responsiveness of neural progenitors to sonic hedgehog signaling. *Developmental cell*, 33(4):373–387.

[127] Kool, M., Korshunov, A., Remke, M., Jones, D. T. W., Schlanstein, M., Northcott, P. A., Cho, Y.-J., Koster, J., Schouten-van Meeteren, A., van Vuurden, D., Clifford, S. C., Pietsch, T., von Bueren, A. O., Rutkowski, S., McCabe, M., Collins, V. P., Bäcklund, M. L., Haberler, C., Bourdeaut, F., Delattre, O., Doz, F., Ellison, D. W., Gilbertson, R. J., Pomeroy, S. L., Taylor, M. D., Lichter, P., and Pfister, S. M. (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(4):473–484.

[128] Koyama-Nasu, R., Nasu-Nishimura, Y., Todo, T., Ino, Y., Saito, N., Aburatani, H., Funato, K., Echizen, K., Sugano, H., Haruta, R., Matsui, M., Takahashi, R., Manabe, E., Oda, T., and Akiyama, T. (2013). The critical role of cyclin D2 in cell cycle progression and tumorigenicity of glioblastoma stem cells. *Oncogene*, 32(33):3840–3845.

[129] Kozarewa, I., Ning, Z., Quail, M. A., Sanders, M. J., Berriman, M., and Turner, D. J. (2009). Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nature Methods*, 6(4):291–295.

[130] Kren, N. P., Zagon, I. S., and McLaughlin, P. J. (2015). Mutations in the opioid growth factor receptor in human cancers alter receptor function. *International journal of molecular medicine*, 36(1):289–293.

[131] Kuboyama, K., Fujikawa, A., Masumura, M., Suzuki, R., Matsumoto, M., and Noda, M. (2012). Protein tyrosine phosphatase receptor type z negatively regulates oligodendrocyte differentiation and myelination. *PLoS ONE*, 7(11):e48797.

[132] Kwong, L., Bijlsma, M. F., and Roelink, H. (2014). Shh-mediated degradation of Hhip allows cell autonomous and non-cell autonomous Shh signalling. *Nature communications*, 5:4849.

[133] Lamprianou, S., Chatzopoulou, E., Thomas, J.-L., Bouyain, S., and Harroch, S. (2011). A complex between contactin-1 and the protein tyrosine phosphatase PTPRZ controls the development of oligodendrocyte precursor cells. *Proceedings of the National Academy of Sciences of the United States of America*, 108(42):17498–17503.

[134] Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, 9:559.

[135] Lastowska, M., Al-Afghani, H., Al-Balool, H. H., Sheth, H., Mercer, E., Coxhead, J. M., Redfern, C. P. F., Peters, H., Burt, A. D., Santibanez-Koref, M., Bacon, C. M., Chesler, L., Rust, A. G., Adams, D. J., Williamson, D., Clifford, S. C., and Jackson, M. S. (2013). Identification of a neuronal transcription factor network involved in medulloblastoma development. *Acta neuropathologica communications*, 1:35.

[136] Lathia, J. D., Mack, S. C., Mulkearns-Hubert, E. E., Valentim, C. L. L., and Rich, J. N. (2015). Cancer stem cells in glioblastoma. *Genes & Development*, 29(12):1203–1217.

[137] Lau, S. K., Chu, P. G., and Weiss, L. M. (2004). CD163: a specific marker of macrophages in paraffin-embedded tissue samples. *American journal of clinical pathology*, 122(5):794–801.

[138] Lawrence, Y. R., Mishra, M. V., Werner-Wasik, M., Andrews, D. W., Showalter, T. N., Glass, J., Shen, X., Symon, Z., and Dicker, A. P. (2012). Improving prognosis of glioblastoma in the 21st century: who has benefited most? *Cancer*, 118(17):4228–4234.

[139] Lazarovici, A., Zhou, T., Shafer, A., Dantas Machado, A. C., Riley, T. R., Sandstrom, R., Sabo, P. J., Lu, Y., Rohs, R., Stamatoyannopoulos, J. A., and Bussemaker, H. J. (2013). Probing DNA shape and methylation state on a genomic scale with DNase I. *Proceedings of the National Academy of Sciences of the United States of America*, 110(16):6376–6381.

[140] Lee, A. M., Shi, Q., Pavey, E., Alberts, S. R., Sargent, D. J., Sinicrope, F. A., Berenberg, J. L., Goldberg, R. M., and Diasio, R. B. (2014a). DPYD variants as predictors of 5-fluorouracil toxicity in adjuvant colon cancer treatment (NCCTG N0147). *Journal of the National Cancer Institute*, 106(12).

[141] Lee, H. K., Hsu, A. K., Sajdak, J., Qin, J., and Pavlidis, P. (2004). Coexpression analysis of human genes across many microarray data sets. *Genome research*, 14(6):1085–1094.

[142] Lee, H. K., Laug, D., Zhu, W., Patel, J. M., Ung, K., Arenkiel, B. R., Fancy, S. P. J., Mohila, C., and Deneen, B. (2015). Apcdd1 stimulates oligodendrocyte differentiation after white matter injury. *Glia*, 63(10):1840–1849.

[143] Lee, J., Kotliarova, S., Kotliarov, Y., Li, A., Su, Q., Donin, N. M., Pastorino, S., Purow, B. W., Christopher, N., Zhang, W., Park, J. K., and Fine, H. A. (2006). Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell*, 9(5):391–403.

[144] Lee, S.-H., Kim, D.-Y., Kang, Y.-Y., Kim, H., Jang, J., Lee, M.-N., Oh, G. T., Kang, S.-W., and Choi, E. Y. (2014b). Developmental endothelial locus-1 inhibits MIF production through suppression of NF-$\kappa$B in macrophages. *International journal of molecular medicine*, 33(4):919–924.

[145] Lemmon, M. A., Schlessinger, J., and Ferguson, K. M. (2014). The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harbor perspectives in biology*, 6(4):a020768.

[146] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

[147] Li, L., Abdel Fattah, E., Cao, G., Ren, C., Yang, G., Goltsov, A. A., Chinault, A. C., Cai, W.-W., Timme, T. L., and Thompson, T. C. (2008). Glioma pathogenesis-related protein 1 exerts tumor suppressor activities through proapoptotic reactive oxygen species-c-Jun-NH2 kinase signaling. *Cancer Research*, 68(2):434–443.

[148] Li, S., Mattar, P., Dixit, R., Lawn, S. O., Wilkinson, G., Kinch, C., Eisenstat, D., Kurrasch, D. M., Chan, J. A., and Schuurmans, C. (2014). RAS/ERK signaling controls proneural genetic programs in cortical development and gliomagenesis. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 34(6):2169–2190.

[149] Ligon, K. L., Alberta, J. A., Kho, A. T., Weiss, J., Kwaan, M. R., Nutt, C. L., Louis, D. N., Stiles, C. D., and Rowitch, D. H. (2004). The oligodendroglial lineage marker OLIG2 is universally expressed in diffuse gliomas. *Journal of neuropathology and experimental neurology*, 63(5):499–509.

[150] Ligon, K. L., Huillard, E., Mehta, S., Kesari, S., Liu, H., Alberta, J. A., Bachoo, R. M., Kane, M., Louis, D. N., DePinho, R. A., Anderson, D. J., Stiles, C. D., and Rowitch, D. H. (2007). Olig2-regulated lineage-restricted pathway controls replication competence in neural stem cells and malignant glioma. *Neuron*, 53(4):503–517.

[151] Lim, D. A., Cha, S., Mayo, M. C., Chen, M.-H., Keles, E., VandenBerg, S., and Berger, M. S. (2007). Relationship of glioblastoma multiforme to neural stem cell regions predicts invasive and multifocal tumor phenotype. *Neuro-Oncology*, 9(4):424–429.

[152] Lin, C.-H., Chiu, L., Lee, H.-T., Chiang, C.-W., Liu, S.-P., Hsu, Y.-H., Lin, S.-Z., Hsu, C. Y., Hsieh, C.-H., and Shyu, W.-C. (2015). PACAP38/PAC1 signaling induces bone marrow-derived cells homing to ischemic brain. *Stem cells (Dayton, Ohio)*, 33(4):1153–1172.

[153] Lin, T.-n., Kim, G.-M., Chen, J.-J., Cheung, W.-M., He, Y. Y., and Hsu, C. Y. (2003). Differential regulation of thrombospondin-1 and thrombospondin-2 after focal cerebral ischemia/reperfusion. *Stroke; a journal of cerebral circulation*, 34(1):177–186.

[154] Liu, C., Sage, J. C., Miller, M. R., Verhaak, R. G. W., Hippenmeyer, S., Vogel, H., Foreman, O., Bronson, R. T., Nishiyama, A., Luo, L., and Zong, H. (2011). Mosaic analysis with double markers reveals tumor cell of origin in glioma. *Cell*, 146(2):209–221.

[155] Liu, G., Yuan, X., Zeng, Z., Tunici, P., Ng, H., Abdulkadir, I. R., Lu, L., Irvin, D., Black, K. L., and Yu, J. S. (2006). Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma. *Molecular Cancer*, 5:67.

[156] Liu, Y., Hayes, D. N., Nobel, A., and Marron, J. S. (2008). Statistical Significance of Clustering for High-Dimension, Low–Sample Size Data. *Journal of the American Statistical Association*, 103(483):1281–1293.

[157] Lottaz, C., Beier, D., Meyer, K., Kumar, P., Hermann, A., Schwarz, J., Junker, M., Oefner, P. J., Bogdahn, U., Wischhusen, J., Spang, R., Storch, A., and Beier, C. P. (2010). Transcriptional Profiles of CD133+ and CD133- Glioblastoma-Derived Cancer Stem Cell Lines Suggest Different Cells of Origin. *Cancer Research*, 70(5):2030–2040.

[158] Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., Scheithauer, B. W., and Kleihues, P. (2007). The 2007 WHO Classification of Tumours of the Central Nervous System. *Acta neuropathologica*, 114(2):97–109.

[159] Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12):550.

[160] Lu, C., Ward, P. S., Kapoor, G. S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C. R., Khanin, R., Figueroa, M. E., Melnick, A., Wellen, K. E., O'Rourke, D. M., Berger, S. L., Chan, T. A., Levine, R. L., Mellinghoff, I. K., and Thompson, C. B. (2012). IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*, 483(7390):474–478.

[161] Lu, Q. R., Park, J. K., Noll, E., Chan, J. A., Alberta, J., Yuk, D., Alzamora, M. G., Louis, D. N., Stiles, C. D., Rowitch, D. H., and Black, P. M. (2001). Oligodendrocyte lineage genes (OLIG) as molecular markers for human glial brain tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19):10851–10856.

[162] Manoranjan, B., Venugopal, C., McFarlane, N., Doble, B. W., Dunn, S. E., Scheinemann, K., and Singh, S. K. (2012). Medulloblastoma stem cells: Modeling tumor heterogeneity. *Cancer letters*.

[163] Manoranjan, B., Wang, X., Hallett, R. M., Venugopal, C., Mack, S. C., McFarlane, N., Nolte, S. M., Scheinemann, K., Gunnarsson, T., Hassell, J. A., Taylor, M. D., Lee, C., Triscott, J., Foster, C. M., Dunham, C., Hawkins, C., Dunn, S. E., and Singh, S. K. (2013). FoxG1 interacts with Bmi1 to regulate self-renewal and tumorigenicity of medulloblastoma stem cells. *Stem cells (Dayton, Ohio)*, 31(7):1266–1277.

[164] Mao, P., Joshi, K., Li, J., Kim, S.-H., Li, P., Santana-Santos, L., Luthra, S., Chandran, U. R., Benos, P. V., Smith, L., Wang, M., Hu, B., Cheng, S.-Y., Sobol, R. W., and Nakano, I. (2013). Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism involving aldehyde dehydrogenase 1A3. *Proceedings of the National Academy of Sciences of the United States of America*, 110(21):8644–8649.

[165] Marchetti, D., Mrak, R. E., Paulsen, D. D., and Sinnappah-Kang, N. D. (2007). Neurotrophin receptors and heparanase: a functional axis in human medulloblastoma invasion. *Journal of experimental & clinical cancer research : CR*, 26(1):5–23.

[166] Marko, N. F., Quackenbush, J., and Weil, R. J. (2011). Why Is There a Lack of Consensus on Molecular Subgroups of Glioblastoma? Understanding the Nature of Biological and Statistical Variability in Glioblastoma Expression Data. *PLoS ONE*, 6(7):e20826.

[167] Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? *Nature Publishing Group*, 12(5):323–334.

[168] McCune, K., Mehta, R., Thorat, M. A., Badve, S., and Nakshatri, H. (2010). Loss of ER$\alpha$ and FOXA1 expression in a progression model of luminal type breast cancer: insights from PyMT transgenic mouse model. *Oncology reports*, 24(5):1233–1239.

[169] McLeay, R. C. and Bailey, T. L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC bioinformatics*, 11:165.

[170] McLendon, R., Friedman, A., Bigner, D., Van Meir, E. G., Brat, D. J., M Mastrogianakis, G., Olson, J. J., Mikkelsen, T., Lehman, N., Aldape, K., Alfred Yung, W. K., Bogler, O., VandenBerg, S., Berger, M., Prados, M., Muzny, D., Morgan, M., Scherer, S., Sabo, A., Nazareth, L., Lewis, L., Hall, O., Zhu, Y., Ren, Y., Alvi, O., Yao, J., Hawes, A., Jhangiani, S., Fowler, G., San Lucas, A., Kovar, C., Cree, A., Dinh, H., Santibanez, J., Joshi, V., Gonzalez-Garay, M. L., Miller, C. A., Milosavljevic, A., Donehower, L., Wheeler, D. A., Gibbs, R. A., Cibulskis, K., Sougnez, C., Fennell, T., Mahan, S., Wilkinson, J., Ziaugra, L., Onofrio, R., Bloom, T., Nicol, R., Ardlie, K., Baldwin, J., Gabriel, S., Lander, E. S., Ding, L., Fulton, R. S., McLellan, M. D., Wallis, J., Larson, D. E., Shi, X., Abbott, R., Fulton, L., Chen, K., Koboldt, D. C., Wendl, M. C., Meyer, R., Tang, Y., Lin, L., Osborne, J. R., Dunford-Shore, B. H., Miner, T. L., Delehaunty, K., Markovic, C., Swift, G., Courtney, W., Pohl, C., Abbott, S., Hawkins, A., Leong, S., Haipek, C., Schmidt, H., Wiechert, M., Vickery, T., Scott, S., Dooling, D. J., Chinwalla, A., Weinstock, G. M., Mardis, E. R., Wilson, R. K., Getz, G., Winckler, W., Verhaak, R. G. W., Lawrence, M. S., O'Kelly, M., Robinson, J., Alexe, G., Beroukhim, R., Carter, S., Chiang, D., Gould, J., Gupta, S., Korn, J., Mermel, C., Mesirov, J., Monti, S., Nguyen, H., Parkin, M., Reich, M., Stransky, N., Weir, B. A., Garraway, L., Golub, T., Meyerson, M., Chin, L., Protopopov, A., Zhang, J., Perna, I., Aronson, S., Sathiamoorthy, N., Ren, G., Yao, J., Wiedemeyer, W. R., Kim, H., Won Kong, S., Xiao, Y., Kohane, I. S., Seidman, J., Park, P. J., Kucherlapati, R., Laird, P. W., Cope, L., Herman, J. G., Weisenberger, D. J., Pan, F., Van Den Berg, D., Van Neste, L., Mi Yi, J., Schuebel, K. E., Baylin, S. B., Absher, D. M., Li, J. Z., Southwick, A., Brady, S., Aggarwal, A., Chung, T., Sherlock, G., Brooks, J. D., Myers, R. M., Spellman, P. T., Purdom, E., Jakkula, L. R., Lapuk, A. V., Marr, H., Dorton, S., Gi Choi, Y., Han, J., Ray, A., Wang, V., Durinck, S., Robinson, M., Wang, N. J., Vranizan, K., Peng, V., Van Name, E., Fontenay, G. V., Ngai, J., Conboy, J. G., Parvin, B., Feiler, H. S., Speed, T. P., Gray, J. W., Brennan, C., Socci, N. D., Olshen, A., Taylor, B. S., Lash, A., Schultz, N., Reva, B., Antipin, Y., Stukalov, A., Gross, B., Cerami, E., Qing Wang, W., Qin, L.-X., Seshan, V. E., Villafania, L., Cavatore, M., Borsu, L., Viale, A., Gerald, W., Sander, C., Ladanyi, M., Perou, C. M., Neil Hayes, D., Topal, M. D., Hoadley, K. A., Qi, Y., Balu, S., Shi, Y., Wu, J., Penny, R., Bittner, M., Shelton, T., Lenkiewicz, E., Morris, S., Beasley, D., Sanders, S., Kahn, A., Sfeir, R., Chen, J., Nassau, D., Feng, L., Hickey, E., Zhang, J., Weinstein, J. N., Barker, A., Gerhard, D. S., Vockley, J., Compton, C., Vaught, J., Fielding, P., Ferguson, M. L., Schaefer, C., Madhavan, S., Buetow, K. H., Collins, F., Good, P., Guyer, M., Ozenberger, B., Peterson, J., and Thomson, E. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216):1061–1068.

[171] Mencarelli, M. A., Spanhol-Rosseto, A., Artuso, R., Rondinella, D., De Filippis, R., Bahi-Buisson, N., Nectoux, J., Rubinsztajn, R., Bienvenu, T., Moncla, A., Chabrol, B.,

Villard, L., Krumina, Z., Armstrong, J., Roche, A., Pineda, M., Gak, E., Mari, F., Ariani, F., and Renieri, A. (2010). Novel FOXG1 mutations associated with the congenital variant of Rett syndrome. *Journal of medical genetics*, 47(1):49–53.

[172] Menendez, J. A., Joven, J., Cufí, S., Corominas-Faja, B., Oliveras-Ferraros, C., Cuyàs, E., Martin-Castillo, B., López-Bonet, E., Alarcón, T., and Vazquez-Martin, A. (2013). The Warburg effect version 2.0: Metabolic reprogramming of cancer stem cells. *Cell cycle (Georgetown, Tex.)*, 12(8).

[173] Meyer, C. A. and Liu, X. S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nature reviews. Genetics*, 15(11):709–721.

[174] Meyer, M., Reimand, J., Lan, X., Head, R., Zhu, X., Kushida, M., Bayani, J., Pressey, J. C., Lionel, A. C., Clarke, I. D., Cusimano, M., Squire, J. A., Scherer, S. W., Bernstein, M., Woodin, M. A., Bader, G. D., and Dirks, P. B. (2015). Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America*.

[175] Michalopoulos, I., Pavlopoulos, G. A., Malatras, A., Karelas, A., Kostadima, M.-A., Schneider, R., and Kossida, S. (2012). Human gene correlation analysis (HGCA): a tool for the identification of transcriptionally co-expressed genes. *BMC research notes*, 5:265.

[176] Molofsky, A. V., Slutsky, S. G., Joseph, N. M., He, S., Pardal, R., Krishnamurthy, J., Sharpless, N. E., and Morrison, S. J. (2006). Increasing p16INK4a expression decreases forebrain progenitors and neurogenesis during ageing. *Nature*, 443(7110):448–452.

[177] Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine learning*, 52(1):91–118.

[178] Moyon, S., Dubessy, A. L., Aigrot, M. S., Trotter, M., Huang, J. K., Dauphinot, L., Potier, M. C., Kerninon, C., Melik Parsadaniantz, S., Franklin, R. J. M., and Lubetzki, C. (2015). Demyelination causes adult CNS progenitors to revert to an immature state and express immune cues that support their migration. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(1):4–20.

[179] Nagalla, S., Chou, J. W., Willingham, M. C., Ruiz, J., Vaughn, J. P., Dubey, P., Lash, T. L., Hamilton-Dutoit, S. J., Bergh, J., Sotiriou, C., Black, M. A., and Miller, L. D. (2013). Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. *Genome biology*, 14(4):R34.

[180] Nakamachi, T., Nakamura, K., Oshida, K., Kagami, N., Mori, H., Watanabe, J., Arata, S., Yofu, S., Endo, K., Wada, Y., Hori, M., Tsuchikawa, D., Kato, M., and Shioda, S. (2011). Pituitary adenylate cyclase-activating polypeptide (PACAP) stimulates proliferation of reactive astrocytes in vitro. *Journal of molecular neuroscience : MN*, 43(1):16–21.

[181] Newman, A. M. and Cooper, J. B. (2010). AutoSOME: a clustering method for identifying gene expression modules without prior knowledge of cluster number. *BMC bioinformatics*, 11:117.

[182] Nishiyama, A., Komitova, M., Suzuki, R., and Zhu, X. (2009). Polydendrocytes (NG2 cells): multifunctional cells with lineage plasticity. *Nature Reviews: Neuroscience*, 10(1):9–22.

[183] 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., and Taylor, M. D. (2011). Medulloblastoma Comprises Four Distinct Molecular Variants. *Journal of Clinical Oncology*, 29(11):1408–1414.

[184] Northcott, P. A., Shih, D. J. H., Peacock, J., Garzia, L., Morrissy, A. S., 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. G., Wang, X., Remke, M., Wu, X., Chiu, R. Y. B., Chu, A., Chuah, E., Corbett, R. D., Hoad, G. R., Jackman, S. D., Li, Y., Lo, A., Mungall, K. L., Nip, K. M., Qian, J. Q., Raymond, A. G. J., Thiessen, N. T., Varhol, R. J., Birol, I., Moore, R. A., Mungall, A. J., Holt, R., Kawauchi, D., Roussel, M. F., Kool, M., Jones, D. T. W., 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., Gillespie, G. Y., Di Rocco, C., Massimi, L., Michiels, E. M. C., 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., Bognar, 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., Ra, Y. S., Robinson, S., Zollo, M., Faria, C. C., Chan, J. A., Levy, M. L., Sorensen, P. H. B., 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. M., Korbel, J. O., Pfister, S. M., Marra, M. A., and Taylor, M. D. (2012). Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature*, 488(7409):49–56.

[185] Noushmehr, H., Weisenberger, D. J., Diefes, K., Phillips, H. S., Pujara, K., Berman, B. P., Pan, F., Pelloski, C. E., Sulman, E. P., Bhat, K. P., Verhaak, R. G. W., Hoadley, K. A., Hayes, D. N., Perou, C. M., Schmidt, H. K., Ding, L., Wilson, R. K., Van Den Berg, D., Shen, H., Bengtsson, H., Neuvial, P., Cope, L. M., Buckley, J., Herman, J. G., Baylin, S. B., Laird, P. W., and Aldape, K. (2010). Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell*, 17(5):510–522.

[186] Ogunrinu, T. A. and Sontheimer, H. (2010). Hypoxia increases the dependence of glioma cells on glutathione. *The Journal of biological chemistry*, 285(48):37716–37724.

[187] Ortensi, B., Osti, D., Pellegatta, S., Pisati, F., Brescia, P., Fornasari, L., Levi, D., Gaetani, P., Colombo, P., Ferri, A., Nicolis, S., Finocchiaro, G., and Pelicci, G. (2012). Rai is a new regulator of neural progenitor migration and glioblastoma invasion. *Stem cells (Dayton, Ohio)*, 30(5):817–832.

[188] Ortensi, B., Setti, M., Osti, D., and Pelicci, G. (2013). Cancer stem cell contribution to glioblastoma invasiveness. *Stem cell research & therapy*, 4(1):18.

[189] Ortiz, B., Fabius, A. W. M., Wu, W. H., Pedraza, A., Brennan, C. W., Schultz, N., Pitter, K. L., Bromberg, J. F., Huse, J. T., Holland, E. C., and Chan, T. A. (2014). Loss of the tyrosine phosphatase PTPRD leads to aberrant STAT3 activation and promotes gliomagenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(22):8149–8154.

[190] Ostrom, Q. T., Gittleman, H., Liao, P., Rouse, C., Chen, Y., Dowling, J., Wolinsky, Y., Kruchko, C., and Barnholtz-Sloan, J. (2014). CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007-2011. *Neuro-Oncology*, 16 Suppl 4:iv1–63.

[191] Ozawa, T., Riester, M., Cheng, Y.-K., Huse, J. T., Squatrito, M., Helmy, K., Charles, N., Michor, F., and Holland, E. C. (2014). Most Human Non-GCIMP Glioblastoma Subtypes Evolve from a Common Proneural-like Precursor Glioma. *Cancer Cell*, 26(2):288–300.

[192] Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Louis, D. N., Rozenblatt-Rosen, O., Suvà, M. L., Regev, A., and Bernstein, B. E. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, 344(6190):1396–1401.

[193] Perez-Janices, N., Blanco-Luquin, I., Tuñón, M. T., Barba-Ramos, E., Ibáñez, B., Zazpe-Cenoz, I., Martinez-Aguillo, M., Hernandez, B., Martínez-Lopez, E., Fernández, A. F., Mercado, M. R., Cabada, T., Escors, D., Megias, D., and Guerrero-Setas, D. (2015). EPB41L3, TSP-1 and RASSF2 as new clinically relevant prognostic biomarkers in diffuse gliomas. *Oncotarget*, 6(1):368–380.

[194] Perou, C. M., Jeffrey, S. S., van de Rijn, M., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenschikov, A., Williams, C. F., Zhu, S. X., Lee, J. C., Lashkari, D., Shalon, D., Brown, P. O., and Botstein, D. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16):9212–9217.

[195] Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A. L., Brown, P. O., and Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, 406(6797):747–752.

[196] Petterson, S. A., Dahlrot, R. H., Hermansen, S. K., K A Munthe, S., Gundesen, M. T., Wohlleben, H., Rasmussen, T., Beier, C. P., Hansen, S., and Kristensen, B. W. (2015). High levels of c-Met is associated with poor prognosis in glioblastoma. *Journal of neuro-oncology*, 122(3):517–527.

[197] Pfenninger, C. V., Roschupkina, T., Hertwig, F., Kottwitz, D., Englund, E., Bengzon, J., Jacobsen, S. E., and Nuber, U. A. (2007). CD133 is not present on neurogenic astrocytes in the adult subventricular zone, but on embryonic neural stem cells, ependymal cells, and glioblastoma cells. *Cancer Research*, 67(12):5727–5736.

[198] Phillips, H. S., Kharbanda, S., Chen, R., Forrest, W. F., Soriano, R. H., Wu, T. D., Misra, A., Nigro, J. M., Colman, H., and Soroceanu, L. (2006). Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell*, 9(3):157–173.

[199] Piccirillo, S. G. M., Reynolds, B. A., Zanetti, N., Lamorte, G., Binda, E., Broggi, G., Brem, H., Olivi, A., DiMeco, F., and Vescovi, A. L. (2006). Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. *Nature*, 444(7120):761–765.

[200] Pisapia, L., Barba, P., Cortese, A., Cicatiello, V., Morelli, F., and Del Pozzo, G. (2015). EBP1 protein modulates the expression of human MHC class II molecules in non-hematopoietic cancer cells. *International journal of oncology*, 47(2):481–489.

[201] Pohl, A. and Beato, M. (2014). bwtool: a tool for bigWig files. *Bioinformatics*, 30(11):1618–1619.

[202] Pollard, S. M. (2013). In vitro expansion of fetal neural progenitors as adherent cell lines. *Methods in molecular biology (Clifton, N.J.)*, 1059:13–24.

[203] Pollard, S. M., Conti, L., Sun, Y., Goffredo, D., and Smith, A. (2006). Adherent neural stem (NS) cells from fetal and adult forebrain. *Cerebral cortex (New York, N.Y. : 1991)*, 16 Suppl 1:i112–20.

[204] Pollard, S. M., Yoshikawa, K., Clarke, I. D., Danovi, D., Stricker, S., Russell, R., Bayani, J., Head, R., Lee, M., Bernstein, M., Squire, J. A., Smith, A., and Dirks, P. (2009). Glioma Stem Cell Lines Expanded in Adherent Culture Have Tumor-Specific Phenotypes and Are Suitable for Chemical and Genetic Screens. *Cell Stem Cell*, 4(6):568–580.

[205] Pontén, J. and Macintyre, E. H. (1968). Long term culture of normal and neoplastic human glia. *Acta pathologica et microbiologica Scandinavica*, 74(4):465–486.

[206] Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., He, X., and Perou, C. M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast cancer research : BCR*, 12(5):R68.

[207] Prat, A. and Perou, C. M. (2011). Deconstructing the molecular portraits of breast cancer. *Molecular oncology*, 5(1):5–23.

[208] Qiang, L., Wu, T., Zhang, H.-W., Lu, N., Hu, R., Wang, Y.-J., Zhao, L., Chen, F.-H., Wang, X.-T., You, Q.-D., and Guo, Q.-L. (2012). HIF-1$\alpha$ is critical for hypoxia-mediated maintenance of glioblastoma stem cells by activating Notch signaling pathway. *Cell death and differentiation*, 19(2):284–294.

[209] Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

[210] Raff, M. C., Williams, B. P., and Miller, R. H. (1984). The in vitro differentiation of a bipotential glial progenitor cell. *The EMBO journal*, 3(8):1857–1864.

[211] Rahbar, A., Stragliotto, G., Orrego, A., Peredo, I., Taher, C., Willems, J., and Soderberg-Naucler1, C. (2012). Low levels of Human Cytomegalovirus Infection in Glioblastoma Multiforme associates with patient survival; -a case-control study. *Herpesviridae*, 3(1):3.

[212] Ransohoff, R. M. and Engelhardt, B. (2012). The anatomical and cellular basis of immune surveillance in the central nervous system. *Nature reviews. Immunology*, 12(9):623–635.

[213] Rao, G., Pedone, C. A., Del Valle, L., Reiss, K., Holland, E. C., and Fults, D. W. (2004). Sonic hedgehog and insulin-like growth factor signaling synergize to induce medulloblastoma formation from nestin-expressing neural progenitors in mice. *Oncogene*, 23(36):6156–6162.

[214] Rao, M. S. and Mayer-Proschel, M. (1997). Glial-restricted precursors are derived from multipotent neuroepithelial stem cells. *Developmental biology*, 188(1):48–63.

[215] Rau, C. D., Wisniewski, N., Orozco, L. D., Bennett, B., Weiss, J., and Lusis, A. J. (2013). Maximal information component analysis: a novel non-linear network analysis method. *Frontiers in genetics*, 4:28.

[216] Ray, P. S., Wang, J., Qu, Y., Sim, M.-S., Shamonki, J., Bagaria, S. P., Ye, X., Liu, B., Elashoff, D., Hoon, D. S., Walter, M. A., Martens, J. W., Richardson, A. L., Giuliano, A. E., and Cui, X. (2010). FOXC1 is a potential prognostic biomarker with functional significance in basal-like breast cancer. *Cancer Research*, 70(10):3870–3876.

[217] Reuss, D. E., Sahm, F., Schrimpf, D., Wiestler, B., Capper, D., Koelsche, C., Schweizer, L., Korshunov, A., Jones, D. T. W., Hovestadt, V., Mittelbronn, M., Schittenhelm, J., Herold-Mende, C., Unterberg, A., Platten, M., Weller, M., Wick, W., Pfister, S. M., and von Deimling, A. (2015). ATRX and IDH1-R132H immunohistochemistry with subsequent copy number analysis and IDH sequencing as a basis for an "integrated" diagnostic approach for adult astrocytoma, oligodendroglioma and glioblastoma. *Acta neuropathologica*, 129(1):133–146.

[218] Reynolds, B. A. and Rietze, R. L. (2005). Neural stem cells and neurospheres–re-evaluating the relationship. *Nature Methods*, 2(5):333–336.

[219] Reynolds, B. A. and Weiss, S. (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science*, 255(5052):1707–1710.

[220] Rheinbay, E., Suvà, M. L., Gillespie, S. M., Wakimoto, H., Patel, A. P., Shahid, M., Oksuz, O., Rabkin, S. D., Martuza, R. L., Rivera, M. N., Louis, D. N., Kasif, S., Chi, A. S., and Bernstein, B. E. (2013). An Aberrant Transcription Factor Network Essential for Wnt Signaling and Stem Cell Maintenance in Glioblastoma. *Cell reports*, 3(5):1567–1579.

[221] Riemenschneider, M. J., Koy, T. H., and Reifenberger, G. (2004). Expression of oligodendrocyte lineage genes in oligodendroglial and astrocytic gliomas. *Acta neuropathologica*, 107(3):277–282.

[222] Rodriguez, M. S., Egaña, I., Lopitz-Otsoa, F., Aillet, F., Lopez-Mato, M. P., Dorronsoro, A., Dorronroso, A., Lobato-Gil, S., Sutherland, J. D., Barrio, R., Trigueros, C., and Lang, V. (2014). The RING ubiquitin E3 RNF114 interacts with A20 and modulates NF-$\kappa$B activity and T-cell activation. *Cell death & disease*, 5:e1399.

[223] Rousseau, A., Nutt, C. L., Betensky, R. A., Iafrate, A. J., Han, M., Ligon, K. L., Rowitch, D. H., and Louis, D. N. (2006). Expression of oligodendroglial and astrocytic lineage markers in diffuse gliomas: use of YKL-40, ApoE, ASCL1, and NKX2-2. *Journal of neuropathology and experimental neurology*, 65(12):1149–1156.

[224] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

[225] Ruan, J., Dean, A. K., and Zhang, W. (2010). A general co-expression network-based approach to gene expression analysis: comparison and applications. *BMC Systems Biology*, 4(1):8.

[226] Sakaguchi, S. (2004). Naturally arising CD4+ regulatory t cells for immunologic self-tolerance and negative control of immune responses. *Annual review of immunology*, 22:531–562.

[227] Sanai, N., Alvarez-Buylla, A., and Berger, M. S. (2005). Neural stem cells and the origin of gliomas. *The New England journal of medicine*, 353(8):811–822.

[228] Schoggins, J. W., Wilson, S. J., Panis, M., Murphy, M. Y., Jones, C. T., Bieniasz, P., and Rice, C. M. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, 472(7344):481–485.

[229] Scott-Drew, S. and ffrench Constant, C. (1997). Expression and function of thrombospondin-1 in myelinating glial cells of the central nervous system. *Journal of neuroscience research*, 50(2):202–214.

[230] Seoane, J., Le, H.-V., Shen, L., Anderson, S. A., and Massagué, J. (2004). Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. *Cell*, 117(2):211–223.

[231] Sgorbissa, A. and Brancolini, C. (2012). IFNs, ISGylation and cancer: Cui prodest? *Cytokine & growth factor reviews*, 23(6):307–314.

[232] Sharif, A., Legendre, P., Prévot, V., Allet, C., Romao, L., Studler, J.-M., Chneiweiss, H., and Junier, M.-P. (2007). Transforming growth factor alpha promotes sequential conversion of mature astrocytes into neural progenitors and stem cells. *Oncogene*, 26(19):2695–2706.

[233] Shaul, Y. D., Freinkman, E., Comb, W. C., Cantor, J. R., Tam, W. L., Thiru, P., Kim, D., Kanarek, N., Pacold, M. E., Chen, W. W., Bierie, B., Possemato, R., Reinhardt, F., Weinberg, R. A., Yaffe, M. B., and Sabatini, D. M. (2014). Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell*, 158(5):1094–1109.

[234] Shen, R., Mo, Q., Schultz, N., Seshan, V. E., Olshen, A. B., Huse, J., Ladanyi, M., and Sander, C. (2012). Integrative subtype discovery in glioblastoma using iCluster. *PLoS ONE*, 7(4):e35236.

[235] Shinawi, T., Hill, V. K., Krex, D., Schackert, G., Gentle, D., Morris, M. R., Wei, W., Cruickshank, G., Maher, E. R., and Latif, F. (2013). DNA methylation profiles of long- and short-term glioblastoma survivors. *Epigenetics*, 8(2):149–156.

[236] Shoshan, Y., Nishiyama, A., Chang, A., Mörk, S., Barnett, G. H., Cowell, J. K., Trapp, B. D., and Staugaitis, S. M. (1999). Expression of oligodendrocyte progenitor cell antigens by gliomas: implications for the histogenesis of brain tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 96(18):10361–10366.

[237] Singh, S. K., Clarke, I. D., Terasaki, M., Bonn, V. E., Hawkins, C., Squire, J., and Dirks, P. B. (2003). Identification of a cancer stem cell in human brain tumors. *Cancer Research*, 63(18):5821–5828.

[238] Smoll, N. R., Gautschi, O. P., Schatlo, B., Schaller, K., and Weber, D. C. (2012). Relative survival of patients with supratentorial low-grade gliomas. *Neuro-Oncology*, 14(8):1062–1069.

[239] Sneath, P. H. A. and Sokal., R. (1973). *Numerical taxonomy*. W. H. Freeman and Company, San Francisco.

[240] Snuderl, M., Fazlollahi, L., Le, L. P., Nitta, M., Zhelyazkova, B. H., Davidson, C. J., Akhavanfard, S., Cahill, D. P., Aldape, K. D., Betensky, R. A., Louis, D. N., and Iafrate, A. J. (2011). Mosaic Amplification of Multiple Receptor Tyrosine Kinase Genes in Glioblastoma. *Cancer Cell*, 20(6):810–817.

[241] Solomon, D. A., Kim, J.-S., Cronin, J. C., Sibenaller, Z., Ryken, T., Rosenberg, S. A., Ressom, H., Jean, W., Bigner, D., Yan, H., Samuels, Y., and Waldman, T. (2008). Mutational inactivation of PTPRD in glioblastoma multiforme and malignant melanoma. *Cancer Research*, 68(24):10300–10306.

[242] Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lønning, P. E., and Børresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19):10869–10874.

[243] Soroceanu, L. and Cobbs, C. S. (2011). Is HCMV a tumor promoter? *Virus research*, 157(2):193–203.

[244] Soroceanu, L., Matlaf, L., Bezrookove, V., Harkins, L., Martinez, R., Greene, M., Soteropoulos, P., and Cobbs, C. S. (2011). Human cytomegalovirus US28 found in glioblastoma promotes an invasive and angiogenic phenotype. *Cancer Research*, 71(21):6643–6653.

[245] Sottoriva, A., Spiteri, I., Piccirillo, S. G. M., Touloumis, A., Collins, V. P., Marioni, J. C., Curtis, C., Watts, C., and Tavaré, S. (2013). Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences of the United States of America*.

[246] Srikanth, M., Kim, J., Das, S., and Kessler, J. A. (2014). BMP signaling induces astrocytic differentiation of clinically derived oligodendroglioma propagating cells. *Molecular cancer research : MCR*, 12(2):283–294.

[247] Stieber, D., Golebiewska, A., Evers, L., Lenkiewicz, E., Brons, N. H. C., Nicot, N., Oudin, A., Bougnaud, S., Hertel, F., Bjerkvig, R., Vallar, L., Barrett, M. T., and Niclou, S. P. (2013). Glioblastomas are composed of genetically divergent clones with distinct tumourigenic potential and variable stem cell-associated phenotypes. *Acta neuropathologica*, 127(2):203–219.

[248] Stoll, E. A., Habibi, B. A., Mikheev, A. M., Lasiene, J., Massey, S. C., Swanson, K. R., Rostomily, R. C., and Horner, P. J. (2011). Increased re-entry into cell cycle mitigates age-related neurogenic decline in the murine subventricular zone. *Stem cells (Dayton, Ohio)*, 29(12):2005–2017.

[249] Stricker, S. H., Feber, A., Engstrom, P. G., Carén, H., Kurian, K. M., Takashima, Y., Watts, C., Way, M., Dirks, P., Bertone, P., Smith, A., Beck, S., and Pollard, S. M. (2013). Widespread resetting of DNA methylation in glioblastoma-initiating cells suppresses malignant cellular behavior in a lineage-dependent manner. *Genes & Development*, 27(6):654–669.

[250] Ström, A. C. and Weis, K. (2001). Importin-beta-like nuclear transport receptors. *Genome biology*, 2(6):REVIEWS3008.

[251] Sun, Y., Pollard, S., Conti, L., Toselli, M., Biella, G., Parkin, G., Willatt, L., Falk, A., Cattaneo, E., and Smith, A. (2008). Long-term tripotent differentiation capacity of human neural stem (NS) cells in adherent culture. *Molecular and Cellular Neuroscience*, 38(2):245–258.

[252] Suvà, M. L., Rheinbay, E., Gillespie, S. M., Patel, A. P., Wakimoto, H., Rabkin, S. D., Riggi, N., Chi, A. S., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Rivera, M. N., Rossetti, N., Kasif, S., Beik, S., Kadri, S., Tirosh, I., Wortman, I., Shalek, A. K., Rozenblatt-Rosen, O., Regev, A., Louis, D. N., and Bernstein, B. E. (2014). Reconstructing and Reprogramming the Tumor-Propagating Potential of Glioblastoma Stem-like Cells. *Cell*, 157(3):580–594.

[253] 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., and Weiss, W. A. (2012). Distinct neural stem cell populations give rise to disparate brain tumors in response to N-MYC. *Cancer Cell*, 21(5):601–613.

[254] Szerlip, N. J., Pedraza, A., Chakravarty, D., Azim, M., McGuire, J., Fang, Y., Ozawa, T., Holland, E. C., Huse, J. T., Jhanwar, S., Leversha, M. A., Mikkelsen, T., and Brennan, C. W. (2012). Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8):3041–3046.

[255] Talasila, K. M., Soentgerath, A., Euskirchen, P., Rosland, G. V., Wang, J., Huszthy, P. C., Prestegarden, L., Skaftnesmo, K. O., Sakariassen, P. Ø., Eskilsson, E., Stieber, D., Keunen, O., Brekka, N., Moen, I., Nigro, J. M., Vintermyr, O. K., Lund-Johansen,

M., Niclou, S., Mørk, S. J., Enger, P. Ø., Bjerkvig, R., and Miletic, H. (2013). EGFR wild-type amplification and activation promote invasion and development of glioblastoma independent of angiogenesis. *Acta neuropathologica*, 125(5):683–698.

[256] Tan, T. Z., Miow, Q. H., Huang, R. Y.-J., Wong, M. K., Ye, J., Lau, J. A., Wu, M. C., Bin Abdul Hadi, L. H., Soong, R., Choolani, M., Davidson, B., Nesland, J. M., Wang, L.-Z., Matsumura, N., Mandai, M., Konishi, I., Goh, B.-C., Chang, J. T., Thiery, J. P., and Mori, S. (2013). Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer. *EMBO molecular medicine*, 5(7):983–998.

[257] TCGA Research Network (2015). TCGA Research Network.

[258] Teschendorff, A. E., Naderi, A., Barbosa-Morais, N. L., and Caldas, C. (2006). PACK: Profile Analysis using Clustering and Kurtosis to find molecular classifiers in cancer. *Bioinformatics*, 22(18):2269–2275.

[259] The ENCODE Project and Consortium, T. E. P., data analysis coordination, O. c., data production, D. p. l., data analysis, L. a., group, W., scientific management, N. p. m., steering committee, P. i., Boise State University and University of North Carolina at Chapel Hill Proteomics groups (data production and analysis), Broad Institute Group (data production and analysis), Cold Spring Harbor, University of Geneva, Center for Genomic Regulation, Barcelona, RIKEN, Sanger Institute, University of Lausanne, Genome Institute of Singapore group (data production and analysis), Data coordination center at UC Santa Cruz (production data coordination), Duke University, EBI, University of Texas, Austin, University of North Carolina-Chapel Hill group (data production and analysis), Genome Institute of Singapore group (data production and analysis), HudsonAlpha Institute, Caltech, UC Irvine, Stanford group (data production and analysis), targeted experimental validation, L. B. N. L. g., data production, analysis, N. g., Sanger Institute, Washington University, Yale University, Center for Genomic Regulation, Barcelona, UCSC, MIT, University of Lausanne, CNIO group (data production and analysis), Stanford-Yale, Harvard, University of Massachusetts Medical School, University of Southern California/UC Davis group (data production and analysis), University of Albany SUNY group (data production and analysis), University of Chicago, Stanford group (data production and analysis), University of Heidelberg group (targeted experimental validation), University of Massachusetts Medical School Bioinformatics group (data production and analysis), University of Massachusetts Medical School Genome Folding group (data production and analysis), University of Washington, University of Massachusetts Medical Center group (data production and analysis), and Data Analysis Center (data analysis) (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 488(7414):57–74.

[260] Theodorou, V., Stark, R., Menon, S., and Carroll, J. S. (2013). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome research*, 23(1):12–22.

[261] Tian, C., Gong, Y., Yang, Y., Shen, W., Wang, K., Liu, J., Xu, B., Zhao, J., and Zhao, C. (2012). Foxg1 has an essential role in postnatal development of the dentate gyrus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(9):2931–2949.

[262] Tkocz, D., Crawford, N. T., Buckley, N. E., Berry, F. B., Kennedy, R. D., Gorski, J. J., Harkin, D. P., and Mullan, P. B. (2012). BRCA1 and GATA3 corepress FOXC1 to inhibit the pathogenesis of basal-like breast cancers. *Oncogene*, 31(32):3667–3678.

[263] Torsvik, A., Stieber, D., Enger, P. Ø., Golebiewska, A., Molven, A., Svendsen, A., Westermark, B., Niclou, S. P., Olsen, T. K., Chekenya Enger, M., and Bjerkvig, R. (2014). U-251 revisited: genetic drift and phenotypic consequences of long-term cultures of glioblastoma cells. *Cancer medicine*, 3(4):812–824.

[264] Trotman, L. C., Wang, X., Alimonti, A., Chen, Z., Teruya-Feldstein, J., Yang, H., Pavletich, N. P., Carver, B. S., Cordon-Cardo, C., Erdjument-Bromage, H., Tempst, P., Chi, S.-G., Kim, H.-J., Misteli, T., Jiang, X., and Pandolfi, P. P. (2007). Ubiquitination regulates PTEN nuclear import and tumor suppression. *Cell*, 128(1):141–156.

[265] Tsunekawa, Y., Britto, J. M., Takahashi, M., Polleux, F., Tan, S.-S., and Osumi, N. (2012). Cyclin D2 in the basal process of neural progenitors is linked to non-equivalent cell fates. *The EMBO journal*, 31(8):1879–1892.

[266] Tunici, P., Bissola, L., Lualdi, E., Pollo, B., Cajola, L., Broggi, G., Sozzi, G., and Finocchiaro, G. (2004). Genetic alterations and in vivo tumorigenicity of neurospheres derived from an adult glioblastoma. *Molecular Cancer*, 3:25.

[267] van den Heuvel, D. M. A., Hellemons, A. J. C. G. M., and Pasterkamp, R. J. (2013). Spatiotemporal expression of repulsive guidance molecules (RGMs) and their receptor neogenin in the mouse brain. *PLoS ONE*, 8(2):e55828.

[268] Vasconcelos, F. F. and Castro, D. S. (2014). Transcriptional control of vertebrate neurogenesis by the proneural factor Ascl1. *Frontiers in cellular neuroscience*, 8:412.

[269] Verginelli, F., Perin, A., Dali, R., Fung, K. H., Lo, R., Longatti, P., Guiot, M.-C., Del Maestro, R. F., Rossi, S., di Porzio, U., Stechishin, O., Weiss, S., and Stifani, S. (2013). Transcription factors FOXG1 and Groucho/TLE promote glioblastoma growth. *Nature communications*, 4:2956.

[270] Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., Miller, C. R., Ding, L., Golub, T., Mesirov, J. P., Alexe, G., Lawrence, M., Kelly, M. O., Tamayo, P., Weir, B. A., Gabriel, S., Winckler, W., Gupta, S., Jakkula, L., Feiler, H. S., Hodgson, J. G., James, C. D., Sarkaria, J. N., Brennan, C., Kahn, A., Spellman, P. T., Wilson, R. K., Speed, T. P., Gray, J. W., Meyerson, M., Getz, G., Perou, C. M., Hayes, D. N., and Network, T. C. G. A. R. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, 17(1):98–110.

[271] Walsh, J. G., Muruve, D. A., and Power, C. (2014). Inflammasomes in the CNS. *Nature Reviews: Neuroscience*, 15(2):84–97.

[272] Walter, M. J., Shen, D., Ding, L., Shao, J., Koboldt, D. C., Chen, K., Larson, D. E., McLellan, M. D., Dooling, D., Abbott, R., Fulton, R., Magrini, V., Schmidt, H., Kalicki-Veizer, J., O'Laughlin, M., Fan, X., Grillot, M., Witowski, S., Heath, S., Frater, J. L., Eades, W., Tomasson, M., Westervelt, P., DiPersio, J. F., Link, D. C., Mardis, E. R., Ley, T. J., Wilson, R. K., and Graubert, T. A. (2012). Clonal architecture of secondary acute myeloid leukemia. *The New England journal of medicine*, 366(12):1090–1098.

[273] Wang, C., Liu, F., Liu, Y.-Y., Zhao, C.-H., You, Y., Wang, L., Zhang, J., Wei, B., Ma, T., Zhang, Q., Zhang, Y., Chen, R., Song, H., and Yang, Z. (2011). Identification and characterization of neuroblasts in the subventricular zone and rostral migratory stream of the adult human brain. *Cell research*, 21(11):1534–1550.

[274] Wang, D. and Fawcett, J. (2012). The perineuronal net and the control of CNS plasticity. *Cell and tissue research*, 349(1):147–160.

[275] Wang, J., Pol, S. U., Haberman, A. K., Wang, C., O'Bara, M. A., and Sim, F. J. (2014a). Transcription factor induction of human oligodendrocyte progenitor fate and differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(28):E2885–94.

[276] Wang, J., Ray, P. S., Sim, M.-S., Zhou, X. Z., Lu, K. P., Lee, A. V., Lin, X., Bagaria, S. P., Giuliano, A. E., and Cui, X. (2012). FOXC1 regulates the functions of human basal-like breast cancer cells by activating NF-$\kappa$B signaling. *Oncogene*, 31(45):4798–4802.

[277] Wang, J., Wang, H., Li, Z., Wu, Q., Lathia, J. D., McLendon, R. E., Hjelmeland, A. B., and Rich, J. N. (2008). c-Myc is required for maintenance of glioma cancer stem cells. *PLoS ONE*, 3(11):e3769.

[278] Wang, Z., Zhang, J., Ye, M., Zhu, M., Zhang, B., Roy, M., Liu, J., and An, X. (2014b). Tumor suppressor role of protein 4.1B/DAL-1. *Cellular and molecular life sciences : CMLS*, 71(24):4815–4830.

[279] Waschek, J. A. (2013). VIP and PACAP: neuropeptide modulators of CNS inflammation, injury, and repair. *British journal of pharmacology*, 169(3):512–523.

[280] Weichselbaum, R. R., Ishwaran, H., Yoon, T., Nuyten, D. S. A., Baker, S. W., Khodarev, N., Su, A. W., Shaikh, A. Y., Roach, P., Kreike, B., Roizman, B., Bergh, J., Pawitan, Y., van de Vijver, M. J., and Minn, A. J. (2008). An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 105(47):18490–18495.

[281] Weigelt, B., Eberle, C., Cowell, C. F., Ng, C. K. Y., and Reis-Filho, J. S. (2014). Metaplastic breast carcinoma:more than a special type. *Nature reviews. Cancer*, pages 1–2.

[282] Wong, M.-T. and Chen, S. S.-L. (2014). Emerging roles of interferon-stimulated genes in the innate immune response to hepatitis C virus infection. *Cellular & molecular immunology*.

[283] Wu, A., Wei, J., Kong, L.-Y., Wang, Y., Priebe, W., Qiao, W., Sawaya, R., and Heimberger, A. B. (2010). Glioma cancer stem cells induce immunosuppressive macrophages/microglia. *Neuro-Oncology*, 12(11):1113–1125.

[284] Xuan, S., Baptista, C. A., Balas, G., Tao, W., Soares, V. C., and Lai, E. (1995). Winged helix transcription factor BF-1 is essential for the development of the cerebral hemispheres. *Neuron*, 14(6):1141–1152.

[285] Yan, K., Wu, Q., Yan, D. H., Lee, C. H., Rahim, N., Tritschler, I., DeVecchio, J., Kalady, M. F., Hjelmeland, A. B., and Rich, J. N. (2014). Glioma cancer stem cells secrete Gremlin1 to promote their maintenance within the tumor hierarchy. *Genes & Development*, 28(10):1085–1100.

[286] Zagon, I. S. and McLaughlin, P. J. (2014). Opioid growth factor and the treatment of human pancreatic cancer: a review. *World journal of gastroenterology : WJG*, 20(9):2218–2223.

[287] Zagon, I. S., Porterfield, N. K., and McLaughlin, P. J. (2013). Opioid growth factor - opioid growth factor receptor axis inhibits proliferation of triple negative breast cancer. *Experimental biology and medicine (Maywood, N.J.)*, 238(6):589–599.

[288] Zentner, G. E. and Henikoff, S. (2014). High-resolution digital profiling of the epigenome. *Nature reviews. Genetics*, 15(12):814–827.

[289] Zetser, A., Bashenko, Y., Miao, H.-Q., Vlodavsky, I., and Ilan, N. (2003). Heparanase affects adhesive and tumorigenic potential of human glioma cells. *Cancer Research*, 63(22):7733–7741.

[290] Zhang, X., Wang, B., and Li, J.-P. (2014). Implications of heparan sulfate and heparanase in neuroinflammation. *Matrix biology : journal of the International Society for Matrix Biology*, 35:174–181.

[291] Zhao, J., He, H., Zhou, K., Ren, Y., Shi, Z., Wu, Z., Wang, Y., Lu, Y., and Jiao, J. (2012). Neuronal transcription factors induce conversion of human glioma cells to neurons and inhibit tumorigenesis. *PLoS ONE*, 7(7):e41506.

[292] Zheng, H., Ying, H., Wiedemeyer, R., Yan, H., Quayle, S. N., Ivanova, E. V., Paik, J.-H., Zhang, H., Xiao, Y., Perry, S. R., Hu, J., Vinjamoori, A., Gan, B., Sahin, E., Chheda, M. G., Brennan, C., Wang, Y. A., Hahn, W. C., Chin, L., and DePinho, R. A. (2010). PLAGL2 regulates Wnt signaling to impede differentiation in neural stem cells and gliomas. *Cancer Cell*, 17(5):497–509.

[293] Zheng, S., Fu, J., Vegesna, R., Mao, Y., Heathcock, L. E., Torres-Garcia, W., Ezhilarasan, R., Wang, S., McKenna, A., Chin, L., Brennan, C. W., Yung, W. K. A., Weinstein, J. N., Aldape, K. D., Sulman, E. P., Chen, K., Koul, D., and Verhaak, R. G. W. (2013). A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival. *Genes & Development*, 27(13):1462–1472.

[294] Zhou, Q., Wang, S., and Anderson, D. J. (2000). Identification of a novel family of oligodendrocyte lineage-specific basic helix-loop-helix transcription factors. *Neuron*, 25(2):331–343.

[295] Zhou, W., Ke, S. Q., Huang, Z., Flavahan, W., Fang, X., Paul, J., Wu, L., Sloan, A. E., McLendon, R. E., Li, X., Rich, J. N., and Bao, S. (2015). Periostin secreted by glioblastoma stem cells recruits M2 tumour-associated macrophages and promotes malignant growth. *Nature Publishing Group*, 17(2):170–182.

[296] Zhu, T. S., Costello, M. A., Talsma, C. E., Flack, C. G., Crowley, J. G., Hamm, L. L., He, X., Hervey-Jumper, S. L., Heth, J. A., Muraszko, K. M., DiMeco, F., Vescovi, A. L., and Fan, X. (2011). Endothelial cells create a stem cell niche in glioblastoma by providing NOTCH ligands that nurture self-renewal of cancer stem-like cells. *Cancer Research*, 71(18):6061–6072.

# Chapter 7

# Appendix

## 7.1 Supplementary figures and tables

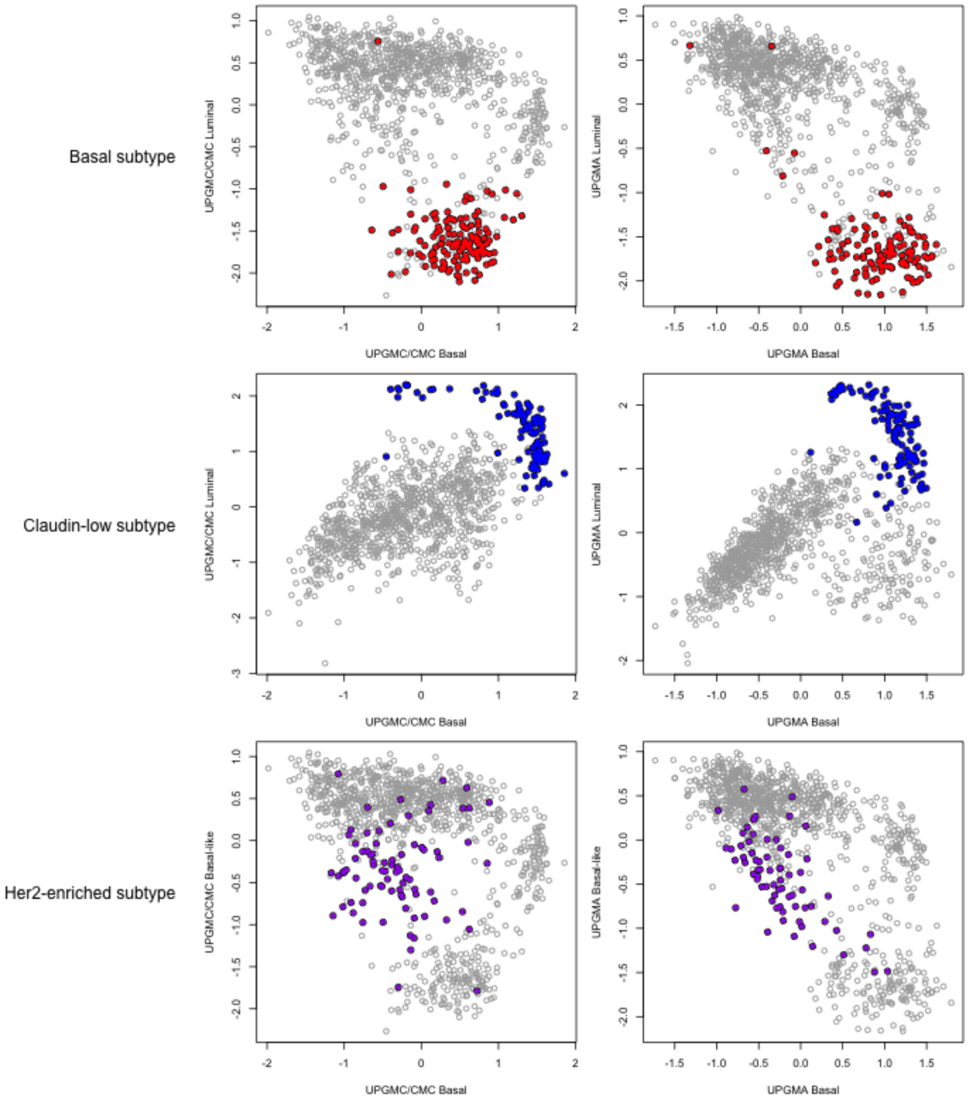


Fig. 7.1 Comparison of UPGMC and UPGMA derived modules and subtypes. Feature related coexpression modules derived from the same correlation matrix and aggregated using both UPGMC (Left column) and UPGMA (Right column) can replicate the BRCA basal, claudin-low and Her2-enriched subtype distinctions set out in Chapter 2. Samples classified as each subtype are highlighted in color for each subtype/plot row. Comparable CMC plots in the main text can be found in figure 3.4. Units used are z-score normalised log2 FPKM.

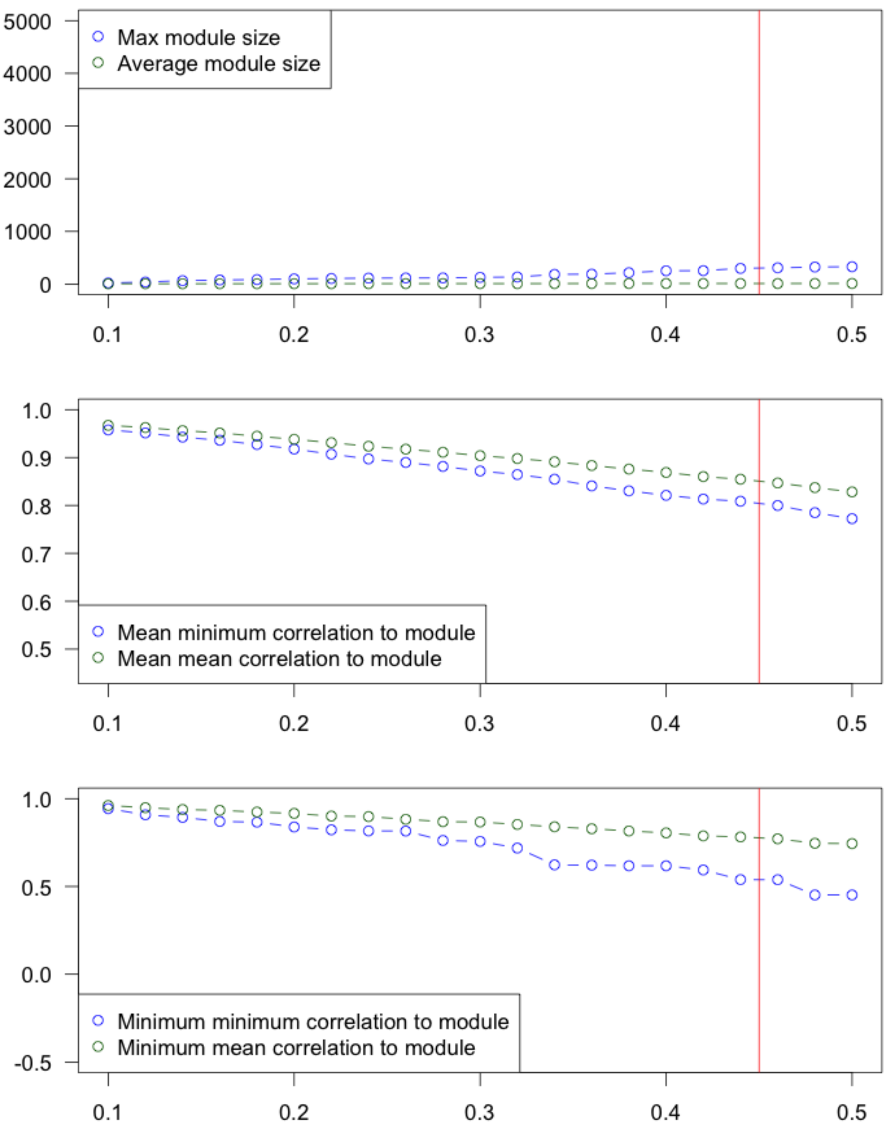


Fig. 7.2 Average linkage module metrics at a range of cut off heights. Coexpression modules derived using the UPGMA linkage method are examined using metrics for average and maximum number of genes per module (Top panel) as well as correlation to module z-score summarised as the mean (Middle panel) or minimum (Bottom panel) of all module metrics. The red line indicates the chosen cut off height of 0.45 for UPGMA derived modules.

| GO:ID | $p$-value | GO Term |
|---|---|---|
| GO:0002376 | 2.259E-39 | immune system process |
| GO:0006955 | 2.470E-38 | immune response |
| GO:0002682 | 1.110E-33 | regulation of immune system process |
| GO:0002684 | 7.879E-29 | positive regulation of immune system process |
| GO:0006952 | 3.879E-26 | defense response |
| GO:0045321 | 8.346E-26 | leukocyte activation |
| GO:0050776 | 2.122E-25 | regulation of immune response |
| GO:0051249 | 4.508E-25 | regulation of lymphocyte activation |
| GO:0046649 | 1.459E-24 | lymphocyte activation |
| GO:0001775 | 2.700E-24 | cell activation |
| GO:0002694 | 4.968E-24 | regulation of leukocyte activation |
| GO:0050865 | 1.335E-23 | regulation of cell activation |
| GO:0050778 | 1.909E-23 | positive regulation of immune response |
| GO:0050863 | 3.624E-23 | regulation of T cell activation |
| GO:0034110 | 8.829E-23 | regulation of homotypic cell-cell adhesion |

Table 7.1 Top 15 gene ontology terms enriched within the consensus immune cell module. The gene universe used for this analysis was all genes contained within a coexpression module at the 0.2 cut off used in each CMC run (total 6620 genes).

| GO:ID | $p$-value | GO Term |
|---|---|---|
| GO:0007049 | 1.703E-86 | cell cycle |
| GO:0000278 | 2.802E-80 | mitotic cell cycle |
| GO:0022402 | 2.523E-78 | cell cycle process |
| GO:1903047 | 1.670E-75 | mitotic cell cycle process |
| GO:0000280 | 1.679E-73 | nuclear division |
| GO:0048285 | 1.012E-72 | organelle fission |
| GO:0007067 | 3.048E-64 | mitotic nuclear division |
| GO:0007059 | 7.610E-55 | chromosome segregation |
| GO:0051301 | 6.133E-54 | cell division |
| GO:1902589 | 3.847E-35 | single-organism organelle organization |
| GO:0006996 | 1.924E-34 | organelle organization |
| GO:0000819 | 1.196E-33 | sister chromatid segregation |
| GO:0044772 | 5.843E-33 | mitotic cell cycle phase transition |
| GO:0051276 | 8.735E-33 | chromosome organization |
| GO:0044770 | 2.983E-32 | cell cycle phase transition |

Table 7.2 Top 15 gene ontology terms enriched within the consensus mitosis module. The gene universe used for this analysis was all genes contained within a coexpression module at the 0.2 cut off used in each CMC run (total 6620 genes).

| Claudin-low / Low expression | Claudin-low / High expression |
|---|---|
| COL10A1 | SPON1 |
| MMP11 | DPYSL3 |
| WISP1 | NID1 |
| BGN | LAMB1 |
| INHBA | SRPX2 |
| P4HA3 | C14orf37 |
| FN1 | HTRA1 |
| CTHRC1 | PRRX1 |
| C1QTNF6 | MMP2 |
| LRRC15 | LOX |
| COL1A1 | MRC2 |
| COL5A1 | CTSK |
| COL5A2 | PCOLCE |
| FAP | ANTXR1 |
| POSTN | ADAMTS2 |
| LOXL1 | CRISPLD2 |
| FNDC1 | CMTM3 |
| VCAN | COL8A2 |
| OLFML2B | COL6A2 |
| COL1A2 | DACT1 |
| THY1 | LUM |
| COL3A1 | SPARC |
| CDH11 | EMILIN1 |
| ADAMTS6 | CHSY3 |
| MMP14 | COL6A1 |
| COL12A1 | SPOCK1 |
|  | SFRP2 |
|  | HTRA3 |
|  | SCARF2 |
|  | ITGA11 |

Table 7.3 Stromal module genes differentially expressed compared to the module average z-score for claudin-low like CMC subcluster samples.

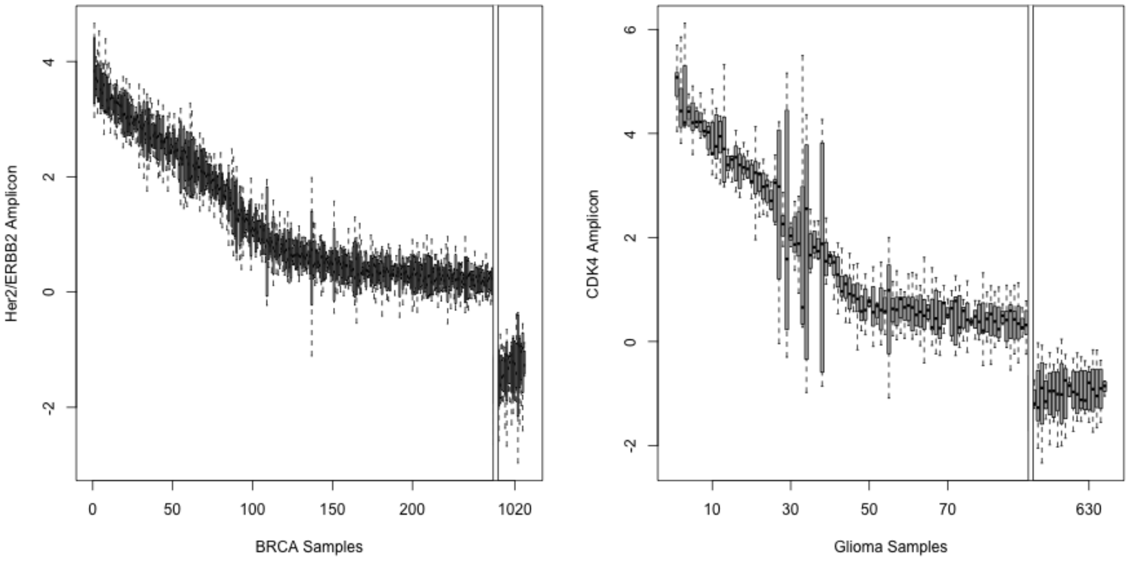


Fig. 7.3 Expression of CMC modules representative of high copy number amplifications of the *ERBB2* and *CDK4* amplicons in BRCA and glioma respectively. Based only on expression its possible to infer the copy number status of highly amplified loci. Y-axis units represent module z-score log2 FPKM.

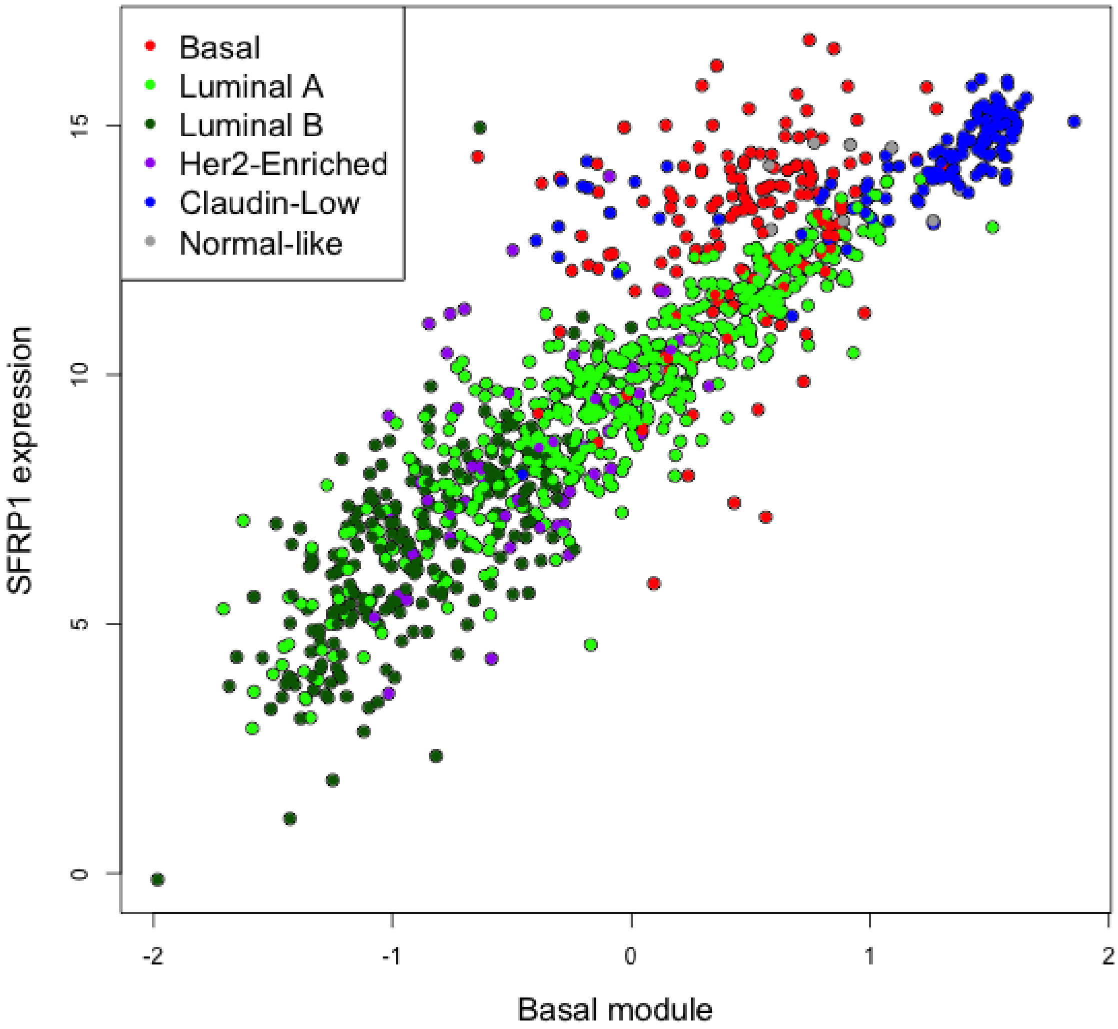


Fig. 7.4 Basal module vs. SFRP1 expression showing Basal expression variation in luminal samples and higher than expected SFRP1 expression in the basal samples. Colours indicate PAM50 classification and units used are log2 FPKM (y axis) and z-score normalised log2 FPKM (x axis).

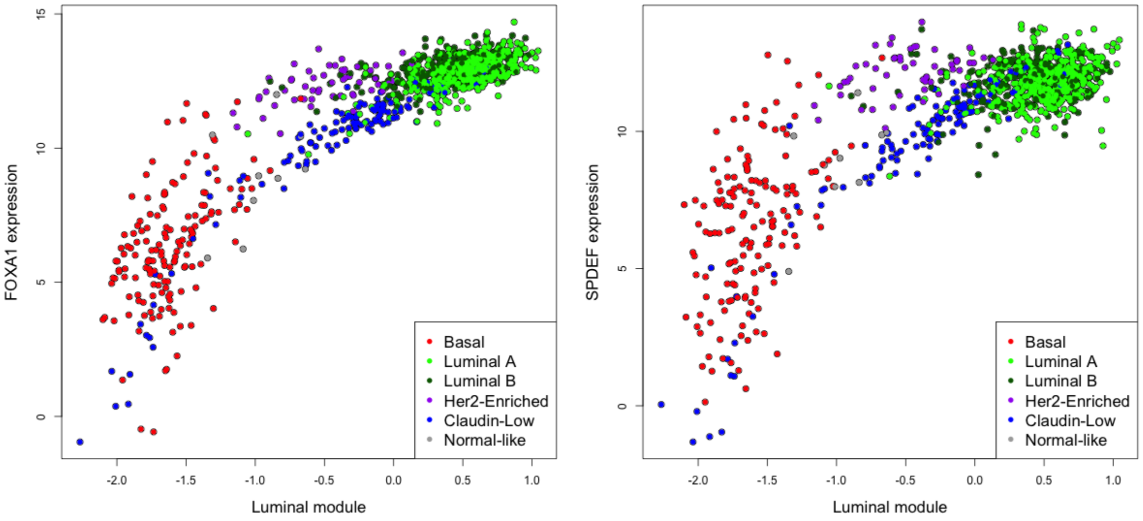


Fig. 7.5 Luminal module vs. FOXA1 and SPDEF expression showing retention of FOXA1 and SPDEF expression in HER2 enriched samples (Purple circles). Units used are log2 FPKM (Y axis) and z-score normalised log2 FPKMs (X axis).

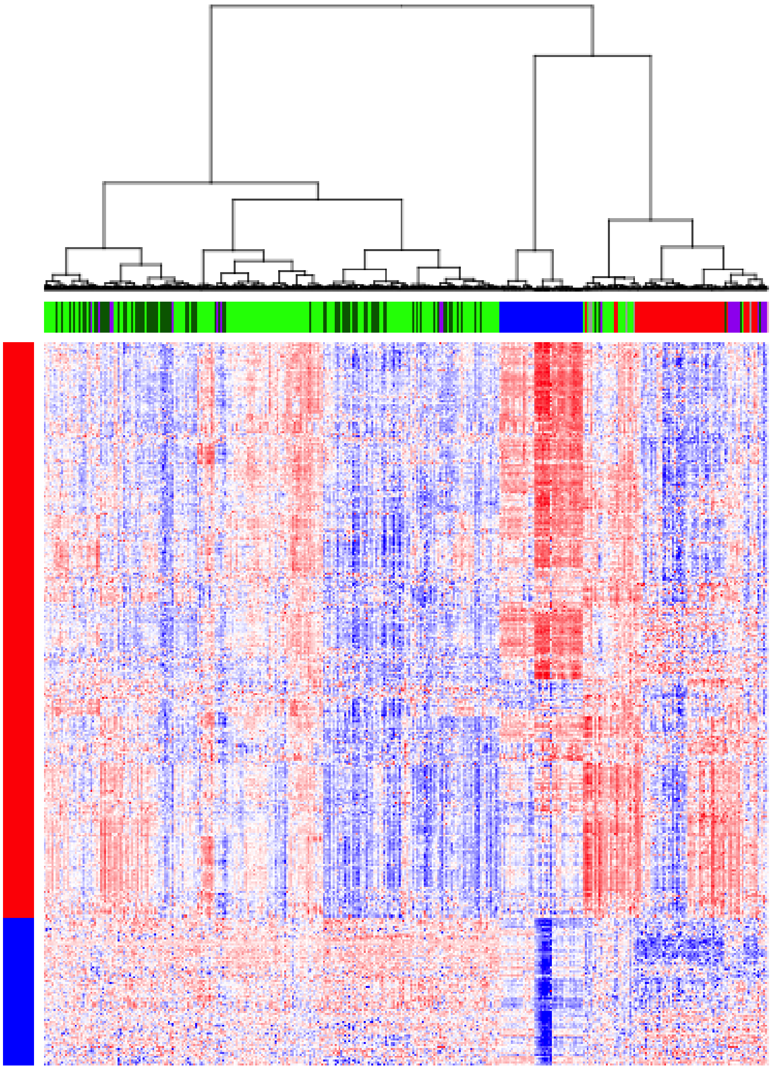


Fig. 7.6 Clustering of BRCA expression data for the identification of Claudin-low subtype samples. Clustering using Claudin-low marker genes identified by Herschkowitz *et al.* [98] (Red and blue row markers). Sample column markers indicate either PAM50 subtypes or the Claudin-low subtype samples inferred by this clustering (Blue samples). Units used are row mean normalised log2 FPKMs.

| Proneural module | Mesenchymal module |
|---|---|
| KLRC2 | POSTN |
| PRLHR | LTF |
| GRIN1 | CHI3L1 |
| HPSE2 | PLA2G2A |
| GABRG2 | HOXA7 |
| CSMD3 | MMP9 |
| TNR | IBSP |
| PCDH15 | ABCC3 |
| SPHKAP | HOXC10 |
| SVOP | HOXA10 |
| CACNG2 | HOXA4 |
| MYT1L | HOXB3 |
| CHGA | HOXA3 |
| SSTR1 | HOXD13 |
| GABRG1 | CA9 |
| GPR17 | HOXD10 |
| CPLX2 | CHI3L2 |
| CUX2 | NNMT |
| FAM123C | COL3A1 |
| VSTM2A | HOXA5 |
| INA | ESM1 |
| SLC1A6 | COL1A1 |
| GLRA3 | CLEC5A |
| KSR2 | HOXD11 |
| AGXT2L1 | SPOCD1 |
| C2orf85 | TNFSF12-TNFSF13 |
| SLC22A6 | CXCL10 |
| HRH3 | PDPN |
| LRTM2 | AQP5 |
| RIMS2 | CD163 |
| L1CAM | TREM1 |
| CBLN1 | SERPINA5 |
| CDH18 | IGFBP2 |
| PCDH11X | HOXA2 |
| WSCD2 | TIMP1 |
| CALN1 | SERPINA3 |
| SCRT1 | FMOD |
| ST8SIA3 | FCGR2B |
| ACTL6B | EMP3 |
| JPH3 | SRPX2 |
| HMP19 | SERPINE1 |
| GFRA1 | METTL7B |
| DACH2 | GDF15 |
| AFF2 | STC1 |
| GALNT13 | HOXB4 |
| PTPRT | COL6A3 |
| KLRC3 | HOXB2 |
| ATP8A2 | HOXC6 |
| DLL3 | ACTG2 |
| ELFN2 | HOXC13 |

Table 7.4 Table of top 50 most variable proneural and mesenchymal genes.

| Mesenchymal | Interferon | Immune Cell |
|---|---|---|
| AQP5 | CXCL11 | SERPINA1 |
| CHI3L1 | CXCL10 | SCIN |
| TREM1 | LOC400759 | RHOH |
| HK3 | ISG15 | FBP1 |
| RBP1 | IFI6 | SLC2A5 |
| CLEC5A | OASL | TLR8 |
| METTL7B | MX1 | C16orf54 |
| TMEM71 | RSAD2 | LOC100233209 |
| COL1A1 | BST2 | IL12RB1 |
| FN1 | OAS2 | CARD9 |
| IGFBP2 | GBP1 | CD84 |
| COL3A1 | IFI44 | C3 |
| ABCC3 | OAS1 | PTPRC |
| RAB36 | CMPK2 | MS4A7 |
| COL1A2 | EPSTI1 | WDFY4 |
| ADAM12 | PSMB9 | CSF3R |
| SH2D4A | SAMD9L | CCR5 |
| COL5A1 | TAP1 | SLC16A3 |
| RARRES2 | MX2 | PYCARD |
| PDLIM4 | STAT1 | GPR160 |
| FAM129A | IFIH1 | CYTIP |
| MIR155HG | HLA-A | NCF1 |
| BATF | HLA-B | LILRA1 |
| PTPN22 | PSMB8 | RIPK3 |
| FBXO17 | HLA-C | PIK3CG |
| LUM | PARP14 | FGL2 |
| ANXA1 | B2M | ADAM28 |
| EVC2 | | DAPP1 |
| C11orf63 | | CLEC7A |
| PYGL | | CIITA |
| C1S | | RNASE2 |
| MYO1G | | GAPT |
| PDCD1LG2 | | SLC11A1 |
| COL6A3 | | CMTM7 |
| TYMP | | FCGR1A |
| PDPN | | VSIG4 |
| FCGR2B | | ITGAL |
| TNFRSF12A | | FPR1 |
| PLAU | | TRAF3IP3 |
| CARD16 | | GPR34 |
| SRPX2 | | CSF2RB |
| APOBEC3G | | HLA-DRB1 |
| FCGR2C | | C17orf87 |
| CISH | | PTAFR |
| PLEK2 | | CYBB |
| COL6A2 | | LILRA2 |
| NNMT | | TFEC |
| PLAUR | | KCNK13 |
| CD248 | | ARHGAP9 |
| S100A4 | | HCST |

Table 7.5 Table of mesenchymal module genes split into smaller and highly correlated interferon, immune cell and mesenchymal modules.

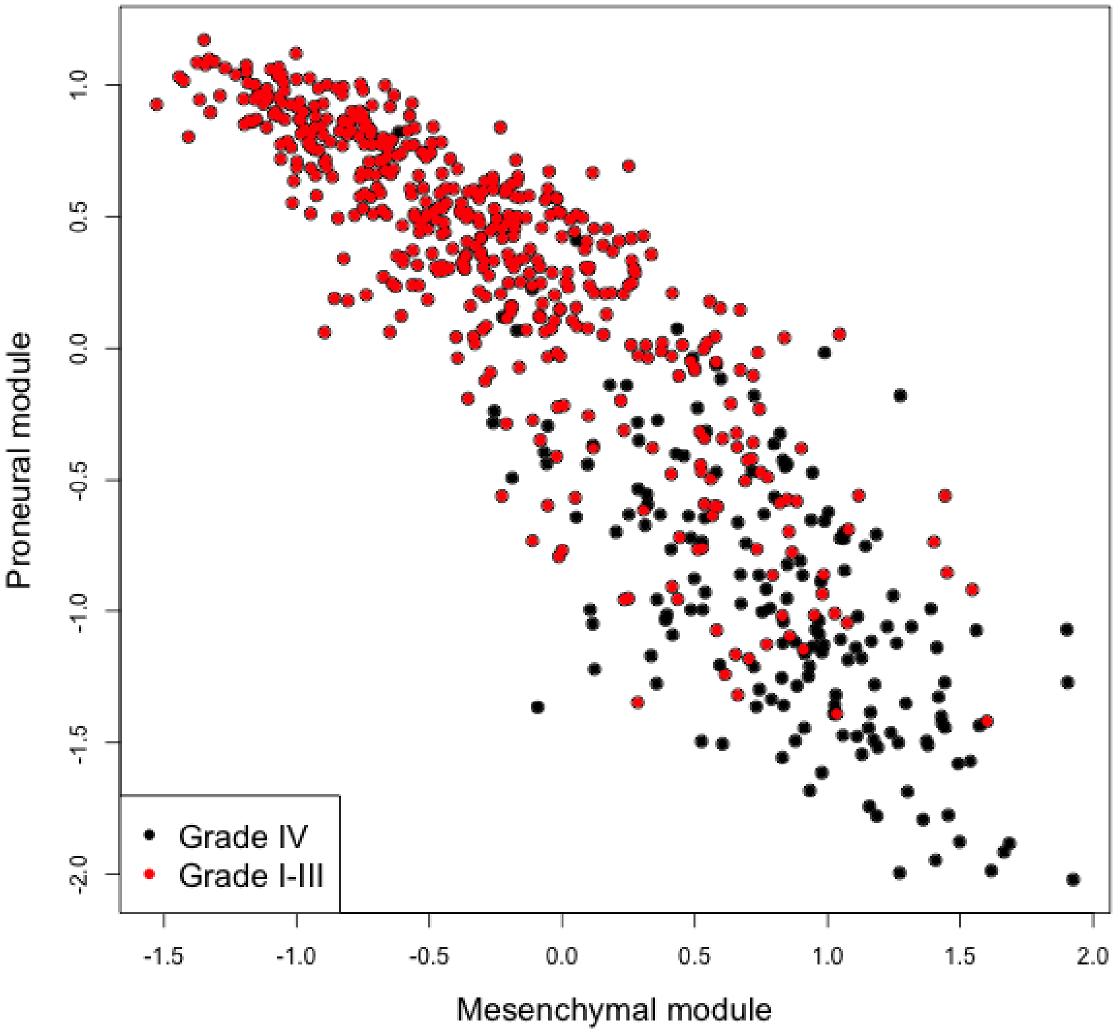


Fig. 7.7 Glioma proneural to mesenchymal axis showing tumour grade. Replicate figure 3.9a with coloring indicating tumour grade instead of Verhaak *et al.* subtypes. Units used are z-score normalised log2 FPKMs.

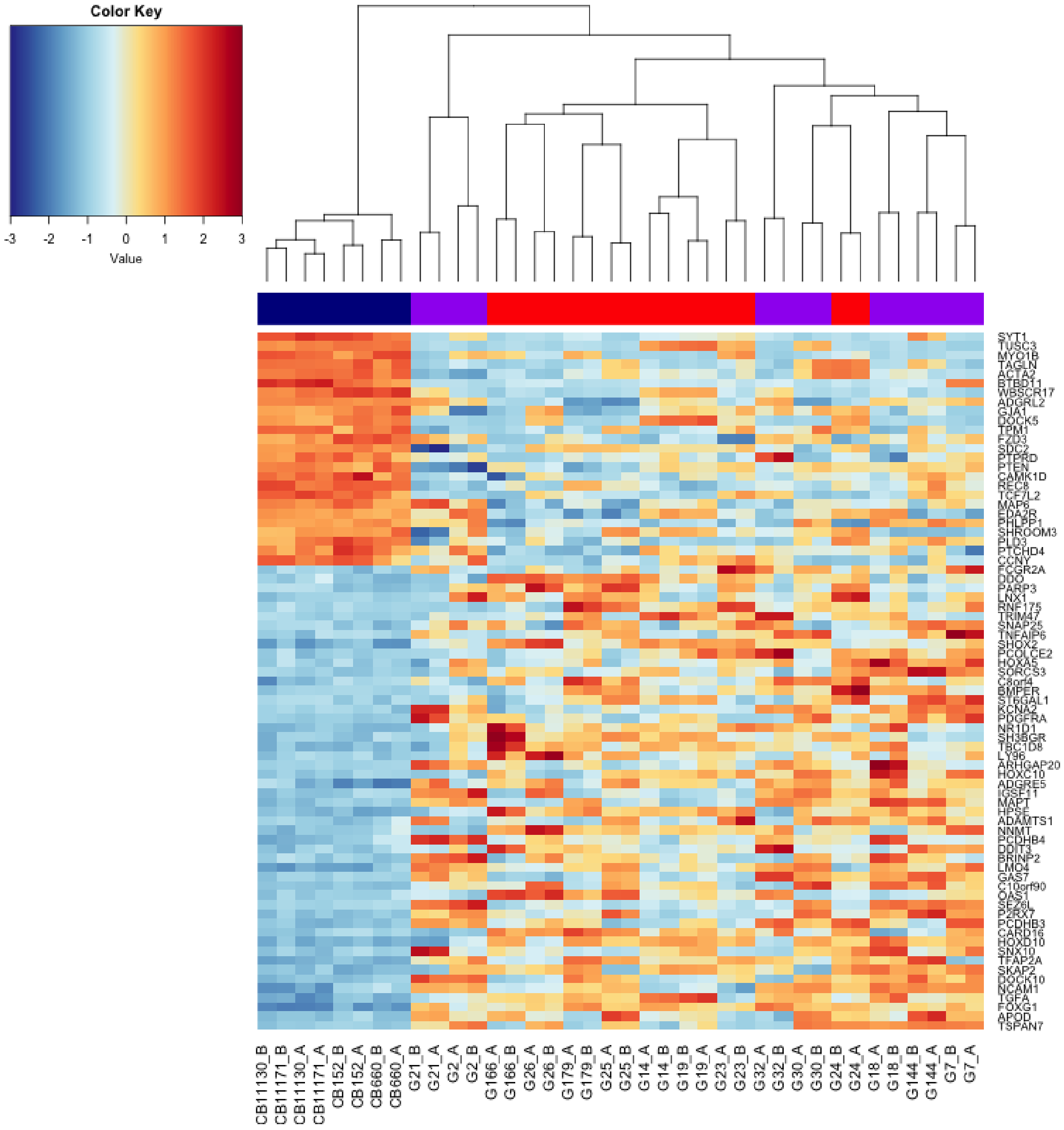


Fig. 7.8 Heatmap illustrating consensus genes differentially expressed between NS and GNS in analysis by Engström *et al.* [55] and new data presented in this document. Genes over and underexpressed in NS cells show relative consistency compared to GNS lines.

| GNS genes | Table 7.6 | | | | |
|---|---|---|---|---|---|
| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
| FOXG1 | 2290 | 1.72e-13 | PPARA | 5465 | 1.29e-03 |
| NRN1 | 51299 | 2.59e-10 | ATG101 | 60673 | 1.31e-03 |
| PCDHB9 | 56127 | 5.42e-08 | GEMIN6 | 79833 | 1.32e-03 |
| TTC39C | 125488 | 5.97e-08 | PPAN | 56342 | 1.33e-03 |
| ZFAND2A | 90637 | 9.27e-08 | MKRN1 | 23608 | 1.38e-03 |
| LDHA | 3939 | 2.98e-07 | HOOK2 | 29911 | 1.41e-03 |
| KLHL13 | 90293 | 5.52e-07 | DOCK10 | 55619 | 1.45e-03 |
| FAM102A | 399665 | 5.93e-07 | DIRAS1 | 148252 | 1.49e-03 |
| DYNLL2 | 140735 | 5.93e-07 | ATP6V1F | 9296 | 1.49e-03 |
| MT2A | 4502 | 6.08e-07 | P2RX7 | 5027 | 1.49e-03 |
| TNFRSF21 | 27242 | 1.02e-06 | HDHD3 | 81932 | 1.58e-03 |
| PMS2P3 | 5387 | 1.28e-06 | | 5380 | 1.62e-03 |
| NUDCD3 | 23386 | 1.55e-06 | TMEM116 | 89894 | 1.64e-03 |
| MTG2 | 26164 | 1.58e-06 | RAB29 | 8934 | 1.64e-03 |
| ADGRE5 | 976 | 1.68e-06 | TPI1P2 | 286016 | 1.64e-03 |
| THY1 | 7070 | 1.79e-06 | WIPI2 | 26100 | 1.69e-03 |
| CD9 | 928 | 2.02e-06 | CASP4 | 837 | 1.69e-03 |
| NKX2-2 | 4821 | 2.15e-06 | ST8SIA5 | 29906 | 1.74e-03 |
| LMO4 | 8543 | 2.87e-06 | POLR2J | 5439 | 1.74e-03 |
| MT1L | 4500 | 3.02e-06 | KIZ | 55857 | 1.77e-03 |
| WDR91 | 29062 | 3.02e-06 | PLOD3 | 8985 | 1.78e-03 |
| C12orf66 | 144577 | 3.02e-06 | INPP5K | 51763 | 1.79e-03 |
| SHOX2 | 6474 | 3.25e-06 | MT1B | 4490 | 1.80e-03 |
| TGFA | 7039 | 3.85e-06 | CBLL1 | 79872 | 1.80e-03 |
| FAM122C | 159091 | 5.09e-06 | AMER1 | 139285 | 1.83e-03 |
| ADAMTS9 | 56999 | 6.20e-06 | APOD | 347 | 1.84e-03 |
| RNF114 | 55905 | 7.20e-06 | BPGM | 669 | 1.85e-03 |
| APCDD1 | 147495 | 8.42e-06 | APOL1 | 8542 | 1.85e-03 |
| OGFR | 11054 | 9.17e-06 | ZDHHC4 | 55146 | 1.85e-03 |
| MR1 | 3140 | 1.07e-05 | ITGA6 | 3655 | 1.86e-03 |
| BCAM | 4059 | 1.11e-05 | DDX31 | 64794 | 1.88e-03 |
| HSF2BP | 11077 | 1.13e-05 | ARFRP1 | 10139 | 1.91e-03 |
| TFAP2A | 7020 | 1.29e-05 | MSI2 | 124540 | 1.91e-03 |
| WBSCR22 | 114049 | 1.43e-05 | LRRC23 | 10233 | 1.91e-03 |
| CNTNAP3 | 79937 | 1.43e-05 | | 727849 | 1.93e-03 |
| GCC1 | 79571 | 1.56e-05 | TTLL11 | 158135 | 1.94e-03 |
| PPM1K | 152926 | 1.76e-05 | ZNF786 | 136051 | 1.98e-03 |
| QSOX2 | 169714 | 1.93e-05 | ZNF212 | 7988 | 1.98e-03 |
| MT1G | 4495 | 2.52e-05 | C9orf114 | 51490 | 1.99e-03 |

Continued on next page

**Table 7.6 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| MT1H | 4496 | 2.52e-05 | ARFGAP1 | 55738 | 1.99e-03 |
| GATAD1 | 57798 | 2.85e-05 | ZNF783 | 100289678 | 1.99e-03 |
| ZSCAN25 | 221785 | 2.96e-05 | TNFRSF19 | 55504 | 2.00e-03 |
| STARD10 | 10809 | 3.21e-05 | MALSU1 | 115416 | 2.00e-03 |
| FBXW4 | 6468 | 3.33e-05 | NME3 | 4832 | 2.00e-03 |
| ARAP3 | 64411 | 3.45e-05 | URGCP | 55665 | 2.07e-03 |
| HOXD10 | 3236 | 3.45e-05 |  | 84054 | 2.07e-03 |
| URB1-AS1 | 84996 | 3.65e-05 | RPS21 | 6227 | 2.07e-03 |
| MT1X | 4501 | 3.91e-05 | PPFIBP2 | 8495 | 2.09e-03 |
| DDX56 | 54606 | 4.15e-05 | BRAT1 | 221927 | 2.20e-03 |
| ATP1A1 | 476 | 4.24e-05 | RAB11FIP4 | 84440 | 2.20e-03 |
| PTCD1 | 26024 | 4.32e-05 | SERPINE2 | 5270 | 2.22e-03 |
| MITF | 4286 | 4.51e-05 | MRM2 | 29960 | 2.22e-03 |
| CIART | 148523 | 5.72e-05 | TMED4 | 222068 | 2.25e-03 |
| TMEM186 | 25880 | 7.08e-05 | ZNF394 | 84124 | 2.26e-03 |
| GET4 | 51608 | 7.13e-05 | SEMA4D | 10507 | 2.28e-03 |
|  | 548321 | 7.19e-05 | MT3 | 4504 | 2.29e-03 |
| EIF3B | 8662 | 7.82e-05 | CSF1 | 1435 | 2.29e-03 |
| PSMG3 | 84262 | 8.18e-05 | KAT2A | 2648 | 2.30e-03 |
| CD82 | 3732 | 8.18e-05 | JUNB | 3726 | 2.32e-03 |
| MRPS24 | 64951 | 8.50e-05 | PRRX1 | 5396 | 2.33e-03 |
| DUSP15 | 128853 | 9.29e-05 | POFUT1 | 23509 | 2.33e-03 |
| HPSE | 10855 | 9.29e-05 | CBX7 | 23492 | 2.35e-03 |
| SNAI2 | 6591 | 1.04e-04 | USP36 | 57602 | 2.38e-03 |
| NSD1 | 64324 | 1.11e-04 | HILPDA | 29923 | 2.39e-03 |
|  | 326343 | 1.16e-04 | POSTN | 10631 | 2.39e-03 |
| PPFIBP1 | 8496 | 1.22e-04 | HEMK1 | 51409 | 2.43e-03 |
| MOCS1 | 4337 | 1.24e-04 | ABHD11 | 83451 | 2.46e-03 |
| SURF1 | 6834 | 1.31e-04 | BAALC | 79870 | 2.46e-03 |
| IFRD1 | 3475 | 1.31e-04 | MRPS33 | 51650 | 2.47e-03 |
| SLC37A3 | 84255 | 1.34e-04 | GRIK4 | 2900 | 2.50e-03 |
| YKT6 | 10652 | 1.36e-04 | FAM122B | 159090 | 2.51e-03 |
| FAM156A | 727866 | 1.56e-04 | UBE2E1 | 7324 | 2.52e-03 |
| MDH2 | 4191 | 1.60e-04 | PLEKHH2 | 130271 | 2.56e-03 |
| SIX1 | 6495 | 1.60e-04 | ARHGAP26 | 23092 | 2.56e-03 |
| HEY1 | 23462 | 1.61e-04 | DDX27 | 55661 | 2.58e-03 |
|  | 100133091 | 1.63e-04 | LRWD1 | 100616367 | 2.59e-03 |
| TSHZ3 | 57616 | 1.68e-04 | LSMEM1 | 286006 | 2.59e-03 |
| TBL2 | 26608 | 1.69e-04 | CHCHD10 | 400916 | 2.61e-03 |

Continued on next page

**Table 7.6 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| ANG | 283 | 1.73e-04 | C7orf26 | 79034 | 2.68e-03 |
| MRGBP | 55257 | 1.73e-04 | TPD52 | 7163 | 2.68e-03 |
| PCDHB10 | 56126 | 1.73e-04 | ZCWPW1 | 55063 | 2.68e-03 |
| NCAM1 | 4684 | 1.74e-04 | CPQ | 10404 | 2.69e-03 |
| DPM3 | 54344 | 1.91e-04 | ZNF232 | 7775 | 2.70e-03 |
| TBX2 | 6909 | 1.91e-04 | MAPT | 4137 | 2.70e-03 |
| PCDHB12 | 56124 | 1.95e-04 | HIP1R | 9026 | 2.74e-03 |
| NMNAT3 | 349565 | 1.95e-04 | NOL4 | 8715 | 2.75e-03 |
| HIST2H2BF | 440689 | 2.02e-04 | KIAA0040 | 9674 | 2.75e-03 |
| BUD31 | 8896 | 2.09e-04 | AAMP | 14 | 2.77e-03 |
| TOMM7 | 54543 | 2.10e-04 | RBCK1 | 10616 | 2.78e-03 |
| SMIM19 | 114926 | 2.15e-04 | SHFM1 | 7979 | 2.78e-03 |
| MRPS25 | 64432 | 2.17e-04 | STARD5 | 80765 | 2.78e-03 |
| RBM28 | 55131 | 2.38e-04 | ASPHD1 | 253982 | 2.79e-03 |
| LINC00921 | 283876 | 2.45e-04 | TCEA2 | 6919 | 2.83e-03 |
| GIGYF1 | 64599 | 2.51e-04 | PMS2 | 5395 | 2.85e-03 |
| GPR156 | 165829 | 2.53e-04 | PCDHB16 | 57717 | 2.85e-03 |
| CLCN4 | 1183 | 2.54e-04 | CCDC126 | 90693 | 2.89e-03 |
| MT1E | 4493 | 2.54e-04 | TMX4 | 56255 | 2.91e-03 |
| TBRG4 | 9238 | 2.60e-04 | PLA2G4C | 8605 | 2.92e-03 |
| TSPAN13 | 27075 | 2.60e-04 | CHST12 | 55501 | 2.92e-03 |
| NR1D1 | 9572 | 2.65e-04 | ACACB | 32 | 2.99e-03 |
| ZNF767P | 79970 | 2.74e-04 | CBX4 | 8535 | 3.01e-03 |
| MT1M | 4499 | 2.84e-04 | PQBP1 | 10084 | 3.03e-03 |
| KATNAL2 | 83473 | 3.00e-04 | LETMD1 | 25875 | 3.10e-03 |
| TBC1D8 | 11138 | 3.04e-04 | CORO2A | 7464 | 3.11e-03 |
| NID2 | 22795 | 3.04e-04 | BAP1 | 8314 | 3.15e-03 |
| TM4SF1 | 4071 | 3.04e-04 | RFTN2 | 130132 | 3.15e-03 |
| TMEM101 | 84336 | 3.13e-04 | POM121 | 9883 | 3.17e-03 |
| CLDN15 | 24146 | 3.16e-04 | FAM3C | 10447 | 3.22e-03 |
| OTUD7B | 56957 | 3.26e-04 | GIMAP2 | 26157 | 3.27e-03 |
| ASB6 | 140459 | 3.34e-04 | ULBP2 | 80328 | 3.27e-03 |
| GRK4 | 2868 | 3.43e-04 | KANK2 | 25959 | 3.31e-03 |
| SKAP2 | 8935 | 3.44e-04 | BRI3 | 25798 | 3.38e-03 |
| HOXD11 | 3237 | 3.71e-04 | DDX55 | 57696 | 3.44e-03 |
| VCAN | 1462 | 3.74e-04 | MREG | 55686 | 3.45e-03 |
| FAM110B | 90362 | 3.85e-04 | SFMBT1 | 51460 | 3.46e-03 |
| GUSB | 2990 | 4.22e-04 | ACTR3B | 57180 | 3.48e-03 |
| PILRB | 29990 | 4.30e-04 | GSTK1 | 373156 | 3.49e-03 |

Continued on next page

**Table 7.6 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| GRB10 | 2887 | 4.38e-04 | DHX8 | 1659 | 3.50e-03 |
| TMEM79 | 84283 | 4.41e-04 | RELL2 | 285613 | 3.51e-03 |
| FAM185A | 222234 | 4.45e-04 | TMEM205 | 374882 | 3.55e-03 |
| PDP2 | 57546 | 4.48e-04 | WRAP73 | 49856 | 3.62e-03 |
| FASTK | 10922 | 4.51e-04 | GCAT | 23464 | 3.63e-03 |
| ATP6V1B1 | 525 | 4.68e-04 | C8orf4 | 56892 | 3.63e-03 |
| EXOSC7 | 23016 | 4.95e-04 | PCDHB4 | 56131 | 3.65e-03 |
| FIS1 | 51024 | 5.01e-04 | MPP6 | 51678 | 3.66e-03 |
| AGFG2 | 3268 | 5.05e-04 | PCYOX1L | 78991 | 3.66e-03 |
| PSD3 | 23362 | 5.05e-04 | IL17RB | 55540 | 3.67e-03 |
| RARRES3 | 5920 | 5.05e-04 | FBXW4P1 | 26226 | 3.68e-03 |
| DNAJA3 | 9093 | 5.17e-04 | PFDN6 | 10471 | 3.68e-03 |
| RAB24 | 53917 | 5.18e-04 | AGAP3 | 116988 | 3.74e-03 |
| VEGFA | 7422 | 5.19e-04 | ABCF2 | 10061 | 3.77e-03 |
| CYTH1 | 9267 | 5.20e-04 | ABCG2 | 9429 | 3.77e-03 |
| ILF3-AS1 | 147727 | 5.36e-04 | SMURF1 | 57154 | 3.77e-03 |
| TOR1B | 27348 | 5.55e-04 | ANAPC2 | 29882 | 3.77e-03 |
| SNX21 | 90203 | 5.64e-04 | ALDH2 | 217 | 3.77e-03 |
| AIMP2 | 7965 | 5.70e-04 | FAM27E3 | 100131997 | 3.81e-03 |
| NUB1 | 51667 | 5.72e-04 | AP4M1 | 9179 | 3.93e-03 |
| CUEDC1 | 404093 | 5.91e-04 | SEMA3B | 7869 | 3.99e-03 |
| HIST1H3C | 8352 | 6.01e-04 | KCNIP3 | 30818 | 4.02e-03 |
| GLUL | 2752 | 6.04e-04 | COX5B | 1329 | 4.03e-03 |
| RCC1L | 81554 | 6.29e-04 | KRI1 | 65095 | 4.05e-03 |
| OXSM | 54995 | 6.29e-04 | HIBADH | 11112 | 4.11e-03 |
| LAS1L | 81887 | 6.32e-04 | TUSC2 | 11334 | 4.16e-03 |
|  | 25845 | 6.32e-04 | TLE2 | 7089 | 4.19e-03 |
| SIPA1L2 | 57568 | 6.35e-04 | CDK5 | 1020 | 4.28e-03 |
| FAM220A | 84792 | 6.36e-04 | PARL | 55486 | 4.36e-03 |
| TRIM4 | 89122 | 6.38e-04 | RNASEH1 | 246243 | 4.45e-03 |
| PRKRIP1 | 79706 | 6.42e-04 | HOXA2 | 3199 | 4.47e-03 |
| RRP9 | 9136 | 6.43e-04 |  | 80154 | 4.52e-03 |
| UPK3BL | 100134938 | 6.50e-04 | TRMO | 51531 | 4.53e-03 |
| PGM1 | 5236 | 6.50e-04 | FAM96B | 51647 | 4.55e-03 |
| TRMU | 55687 | 6.85e-04 | CA13 | 377677 | 4.56e-03 |
| BLCAP | 10904 | 6.85e-04 | RILPL2 | 196383 | 4.66e-03 |
| MRM1 | 79922 | 6.96e-04 | SLC25A35 | 399512 | 4.71e-03 |
| TRAF3IP3 | 80342 | 7.16e-04 | FKBP4 | 2288 | 4.71e-03 |
| DNAH9 | 1770 | 7.17e-04 | CRIPAK | 285464 | 4.71e-03 |

Continued on next page

**Table 7.6 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| HOXC6 | 3223 | 7.35e-04 | NAT9 | 26151 | 4.71e-03 |
| ZKSCAN5 | 23660 | 7.36e-04 | SAMD9 | 54809 | 4.72e-03 |
| RAD52 | 5893 | 7.55e-04 | PJA1 | 64219 | 4.72e-03 |
| GSC | 145258 | 7.77e-04 | LINC01089 | 338799 | 4.72e-03 |
| STK32B | 55351 | 7.85e-04 | NDRG1 | 10397 | 4.75e-03 |
| CAMK2N1 | 55450 | 7.90e-04 |  | 11068 | 4.75e-03 |
| FAM73B | 84895 | 7.92e-04 | RNF216 | 54476 | 4.75e-03 |
| CDC14B | 8555 | 7.92e-04 | SMIM14 | 201895 | 4.83e-03 |
| CDKN2C | 1031 | 7.94e-04 | DPY19L2P2 | 349152 | 4.86e-03 |
| VOPP1 | 81552 | 7.95e-04 | CEBPB | 1051 | 4.86e-03 |
| TUB | 7275 | 7.98e-04 | GUCY1B3 | 2983 | 4.86e-03 |
| MT1JP | 4498 | 8.02e-04 | MRPS18B | 28973 | 4.86e-03 |
| PCDHB3 | 56132 | 8.02e-04 | ZNF343 | 79175 | 4.86e-03 |
| SLC47A1 | 55244 | 8.08e-04 | NSUN5P2 | 260294 | 4.88e-03 |
| SLC41A1 | 254428 | 8.22e-04 | POLR1E | 64425 | 4.89e-03 |
| UCK1 | 83549 | 8.26e-04 | CYP2J2 | 1573 | 4.91e-03 |
| ACSS1 | 84532 | 8.36e-04 | DMTN | 2039 | 4.95e-03 |
| DDRGK1 | 65992 | 8.52e-04 | KLHDC8A | 55220 | 4.97e-03 |
|  | 5383 | 8.53e-04 | IDH3B | 3420 | 4.97e-03 |
| NDRG2 | 57447 | 8.72e-04 | PIGL | 9487 | 4.98e-03 |
| SURF6 | 6838 | 8.81e-04 | TSPAN7 | 7102 | 5.00e-03 |
| CSPG4P12 | 440300 | 8.81e-04 | ZNF774 | 342132 | 5.06e-03 |
| MEPCE | 56257 | 8.81e-04 | RSAD1 | 55316 | 5.09e-03 |
| NDEL1 | 81565 | 9.04e-04 | LHX2 | 9355 | 5.11e-03 |
| REXO4 | 57109 | 9.05e-04 | PDPR | 55066 | 5.11e-03 |
| SLC12A9 | 56996 | 9.17e-04 | RRBP1 | 6238 | 5.11e-03 |
| CDC25B | 994 | 9.57e-04 | HKR1 | 284459 | 5.11e-03 |
| STYXL1 | 51657 | 9.65e-04 | SIM2 | 6493 | 5.14e-03 |
| ANKS3 | 124401 | 9.88e-04 | HSPB8 | 26353 | 5.18e-03 |
| C8orf33 | 65265 | 9.94e-04 | TMEM175 | 84286 | 5.18e-03 |
| C7orf49 | 78996 | 1.01e-03 | MAVS | 57506 | 5.19e-03 |
| GCNT2 | 2651 | 1.01e-03 | ICA1 | 3382 | 5.25e-03 |
| TMEM208 | 29100 | 1.01e-03 | FAIM2 | 23017 | 5.26e-03 |
| DDR2 | 4921 | 1.01e-03 | FBXO31 | 79791 | 5.26e-03 |
|  | 441194 | 1.03e-03 | PDGFRA | 5156 | 5.27e-03 |
| GHDC | 84514 | 1.06e-03 | RAB3D | 9545 | 5.29e-03 |
| BIN3 | 55909 | 1.08e-03 | B3GNT9 | 84752 | 5.40e-03 |
| C10orf90 | 118611 | 1.08e-03 | HES1 | 3280 | 5.42e-03 |
| TRRAP | 8295 | 1.13e-03 | TRPS1 | 7227 | 5.51e-03 |

Continued on next page

**Table 7.6 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| C1orf109 | 54955 | 1.20e-03 | SULF2 | 55959 | 5.54e-03 |
| HS3ST5 | 222537 | 1.22e-03 | OLIG2 | 10215 | 5.55e-03 |
| TMEM51 | 55092 | 1.24e-03 | STX1A | 6804 | 5.56e-03 |
|  | 2310 | 1.24e-03 |  | 442578 | 5.68e-03 |
| NDUFB2 | 4708 | 1.24e-03 | SLC11A2 | 4891 | 5.71e-03 |

Table 7.6 Table of top 400 genes overexpressed in GNS compared to NS sorted by adjusted p-value.

| NS genes | Table 7.6 | | | | |
|---|---|---|---|---|---|
| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
| RANBP17 | 64901 | 1.11e-22 | NETO2 | 81831 | 2.63e-04 |
| TES | 26136 | 2.03e-14 | ARHGEF28 | 64283 | 2.74e-04 |
| RASGRF2 | 5924 | 4.16e-14 | TEAD1 | 7003 | 2.81e-04 |
| EPHA7 | 2045 | 4.16e-14 | ERCC6 | 2074 | 2.89e-04 |
| OTX2 | 5015 | 1.01e-13 | ATP6V1B2 | 526 | 2.96e-04 |
| TNFRSF10D | 8793 | 4.84e-13 | MICALL1 | 85377 | 2.96e-04 |
| CDCP1 | 64866 | 1.55e-12 | NMT2 | 9397 | 3.00e-04 |
| AFF2 | 2334 | 2.92e-12 | MACF1 | 23499 | 3.13e-04 |
| SYT1 | 6857 | 1.37e-11 | ADAMTS16 | 170690 | 3.26e-04 |
| ANO4 | 121601 | 1.95e-11 | SPCS3 | 60559 | 3.29e-04 |
| AK7 | 122481 | 5.04e-11 | FNBP1L | 54874 | 3.31e-04 |
| BTBD11 | 121551 | 1.03e-10 | PLEKHA5 | 54477 | 3.37e-04 |
| MCHR1 | 2847 | 2.28e-10 | LIMS1 | 3987 | 3.43e-04 |
| CRHBP | 1393 | 2.47e-10 | NHS | 4810 | 3.43e-04 |
| GREB1L | 80000 | 6.73e-10 | ZDHHC20 | 253832 | 3.43e-04 |
| IGF2BP1 | 10642 | 6.73e-10 | RBM24 | 221662 | 3.44e-04 |
| NELL2 | 4753 | 9.44e-10 | EXO5 | 64789 | 3.68e-04 |
| PBX3 | 5090 | 9.44e-10 | KIF1BP | 26128 | 3.71e-04 |
| NEFM | 4741 | 1.88e-09 | SORBS1 | 10580 | 3.85e-04 |
| EPB41L3 | 23136 | 2.98e-09 | USP12 | 219333 | 3.85e-04 |
| MGST1 | 4257 | 3.18e-09 | KIAA1217 | 56243 | 3.85e-04 |
| NEGR1 | 257194 | 7.97e-09 | TIAL1 | 7073 | 3.95e-04 |
| LRRC7 | 57554 | 8.47e-09 | LBH | 81606 | 4.01e-04 |
| WBSCR17 | 64409 | 9.58e-09 | WEE1 | 7465 | 4.01e-04 |
| GRPR | 2925 | 1.34e-08 | NDST2 | 8509 | 4.14e-04 |
| RSU1 | 6251 | 5.47e-08 | ENAH | 55740 | 4.25e-04 |
| RAB11FIP1 | 80223 | 6.16e-08 | SERPING1 | 710 | 4.28e-04 |
| REC8 | 9985 | 6.73e-08 | SPATS2L | 26010 | 4.40e-04 |
| NECAB1 | 64168 | 1.32e-07 | ATXN10 | 25814 | 4.41e-04 |
| RGMB | 285704 | 1.32e-07 | TEX2 | 55852 | 4.48e-04 |
| HS3ST3A1 | 9955 | 1.49e-07 | AP1S2 | 8905 | 4.95e-04 |
| NXN | 64359 | 1.59e-07 | FNDC3A | 22862 | 5.06e-04 |
| CELSR1 | 9620 | 1.91e-07 | PCSK5 | 5125 | 5.18e-04 |
| SLC18A3 | 6572 | 2.53e-07 | ALS2 | 57679 | 5.18e-04 |
| OCA2 | 4948 | 2.83e-07 | MSMO1 | 6307 | 5.51e-04 |
| DAPK1 | 1612 | 3.04e-07 | CACUL1 | 143384 | 5.55e-04 |
| MYO1B | 4430 | 3.40e-07 | HEPH | 9843 | 5.55e-04 |
| TLE4 | 7091 | 3.59e-07 | LNX2 | 222484 | 5.72e-04 |
| DOCK2 | 1794 | 3.80e-07 | ARHGAP32 | 9743 | 5.74e-04 |

Continued on next page

**Table 7.7 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| OXTR | 5021 | 6.08e-07 | UBAP1 | 51271 | 5.83e-04 |
| CNN1 | 1264 | 6.08e-07 | REEP3 | 221035 | 5.85e-04 |
| ANK3 | 288 | 6.08e-07 | ADK | 132 | 6.32e-04 |
| CACHD1 | 57685 | 6.08e-07 | BEND6 | 221336 | 6.32e-04 |
| CCDC6 | 8030 | 6.12e-07 | SPATA18 | 132671 | 6.35e-04 |
| ANXA3 | 306 | 6.55e-07 | CAP2 | 10486 | 6.60e-04 |
| SEMA3D | 223117 | 6.98e-07 | SLC9A7 | 84679 | 6.63e-04 |
| CAMK1D | 57118 | 9.12e-07 | SMAD7 | 4092 | 6.68e-04 |
| MSRB3 | 253827 | 9.12e-07 | ARHGAP12 | 94134 | 6.75e-04 |
| NXPH2 | 11249 | 9.12e-07 | C3AR1 | 719 | 6.82e-04 |
| PDE4D | 5144 | 1.33e-06 | PTPRD | 5789 | 6.86e-04 |
| INPP5A | 3632 | 1.33e-06 | SCD | 6319 | 7.04e-04 |
| FAM84B | 157638 | 1.39e-06 | TCN2 | 6948 | 7.09e-04 |
| ZFAND4 | 93550 | 1.45e-06 | PTEN | 5728 | 7.11e-04 |
| PREP | 5550 | 1.53e-06 | ADGRL2 | 23266 | 7.14e-04 |
| UACA | 55075 | 1.59e-06 | FAM118A | 55007 | 7.25e-04 |
| LPP | 4026 | 1.73e-06 | IDI1 | 3422 | 7.45e-04 |
| UBE2D1 | 7321 | 1.73e-06 | XPR1 | 9213 | 7.47e-04 |
| TCF7L2 | 6934 | 1.73e-06 | ADAM19 | 8728 | 7.56e-04 |
| ATRNL1 | 26033 | 1.79e-06 | E2F5 | 1875 | 7.94e-04 |
| AMIGO2 | 347902 | 1.85e-06 | AUTS2 | 26053 | 7.95e-04 |
| IRX3 | 79191 | 2.37e-06 | STK10 | 6793 | 7.95e-04 |
| NEDD4 | 4734 | 2.39e-06 | KMO | 8564 | 7.96e-04 |
| SORBS2 | 8470 | 2.48e-06 | WDFY1 | 57590 | 8.01e-04 |
| PDZD8 | 118987 | 2.84e-06 | BLOC1S2 | 282991 | 8.05e-04 |
| IPO5 | 3843 | 2.86e-06 | ARMT1 | 79624 | 8.26e-04 |
| STK32A | 202374 | 2.89e-06 | PACS1 | 55690 | 8.40e-04 |
| NHLRC2 | 374354 | 3.02e-06 | MAP6 | 4135 | 8.55e-04 |
| CAP1 | 10487 | 3.02e-06 | GBF1 | 8729 | 8.72e-04 |
| NETO1 | 81832 | 3.24e-06 | THUMPD2 | 80745 | 8.77e-04 |
| GFRA1 | 2674 | 3.54e-06 | PTPRU | 10076 | 9.80e-04 |
| WDFY2 | 115825 | 4.06e-06 | PALLD | 23022 | 9.80e-04 |
| WDFY3 | 23001 | 4.06e-06 | SFRP1 | 6422 | 1.00e-03 |
| THSD4 | 79875 | 4.12e-06 | ACTR1A | 10121 | 1.01e-03 |
|  | 8464 | 4.12e-06 | PTPRE | 5791 | 1.02e-03 |
| PDLIM1 | 9124 | 4.13e-06 | RBPMS | 11030 | 1.06e-03 |
| GNG12 | 55970 | 4.95e-06 | ANTXR1 | 84168 | 1.10e-03 |
| TAGLN | 6876 | 5.20e-06 | TRPC4 | 7223 | 1.13e-03 |
| LEPR | 3953 | 5.64e-06 | CCSER2 | 54462 | 1.14e-03 |

Continued on next page

**Table 7.7 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| ADGRL4 | 64123 | 5.66e-06 | ROR1 | 4919 | 1.20e-03 |
| EXT1 | 2131 | 5.92e-06 | BAG3 | 9531 | 1.21e-03 |
| SKIDA1 | 387640 | 6.27e-06 | WWTR1 | 25937 | 1.21e-03 |
| ABCC4 | 10257 | 7.05e-06 | B3GAT2 | 135152 | 1.22e-03 |
| SPOPL | 339745 | 7.20e-06 | LPCAT2 | 54947 | 1.25e-03 |
| EFR3B | 22979 | 7.20e-06 | BVES | 11149 | 1.26e-03 |
| VGLL3 | 389136 | 7.41e-06 | RAI14 | 26064 | 1.26e-03 |
| GTF2A1L | 11036 | 7.57e-06 | ATP11C | 286410 | 1.28e-03 |
| ACTA2 | 59 | 7.86e-06 | RPP30 | 10556 | 1.29e-03 |
| RFX3 | 5991 | 9.77e-06 | COL14A1 | 7373 | 1.33e-03 |
| TXLNB | 167838 | 1.03e-05 | LINC01006 | 129790 | 1.33e-03 |
| BCAR3 | 8412 | 1.16e-05 | GDI2 | 2665 | 1.33e-03 |
| ST8SIA2 | 8128 | 1.16e-05 | CTTNBP2NL | 55917 | 1.33e-03 |
| PDLIM5 | 10611 | 1.17e-05 | DCP1B | 196513 | 1.36e-03 |
| RFX7 | 64864 | 1.17e-05 | CHST7 | 56548 | 1.37e-03 |
| TM4SF18 | 116441 | 1.21e-05 | HMGCS1 | 3157 | 1.38e-03 |
| PPP1R21 | 129285 | 1.25e-05 | TERF1 | 7013 | 1.39e-03 |
| FAM21C | 253725 | 1.27e-05 | MINPP1 | 9562 | 1.40e-03 |
| EMB | 133418 | 1.37e-05 | SYTL5 | 94122 | 1.40e-03 |
| DACH1 | 1602 | 1.40e-05 | BST1 | 683 | 1.44e-03 |
| TPH1 | 7166 | 1.43e-05 | C11orf80 | 79703 | 1.49e-03 |
| MMP15 | 4324 | 1.46e-05 |  | 100133106 | 1.49e-03 |
| BASP1 | 10409 | 1.52e-05 | AJUBA | 84962 | 1.49e-03 |
| ITSN1 | 6453 | 1.60e-05 | AKT3 | 10000 | 1.50e-03 |
| PHKB | 5257 | 1.84e-05 | GJA1 | 2697 | 1.50e-03 |
| QDPR | 5860 | 1.94e-05 | GCNT1 | 2650 | 1.52e-03 |
| EPHX4 | 253152 | 2.08e-05 | AK5 | 26289 | 1.54e-03 |
| MAB21L1 | 4081 | 2.12e-05 | FAM49B | 51571 | 1.54e-03 |
| TPM1 | 7168 | 2.31e-05 | NFATC4 | 4776 | 1.57e-03 |
| BMPR1A | 657 | 2.35e-05 | HHEX | 3087 | 1.57e-03 |
| LGR4 | 55366 | 2.36e-05 | ARNTL2 | 56938 | 1.59e-03 |
| TRDMT1 | 1787 | 2.40e-05 | TMEM135 | 65084 | 1.64e-03 |
| PPP3CB | 5532 | 2.46e-05 | POPDC3 | 64208 | 1.64e-03 |
| AP3M1 | 26985 | 2.60e-05 | COTL1 | 23406 | 1.64e-03 |
| LOX | 4015 | 2.69e-05 | NTSR1 | 4923 | 1.64e-03 |
| FAM160B1 | 57700 | 2.76e-05 | PTPN14 | 5784 | 1.69e-03 |
| LCOR | 84458 | 2.86e-05 | ARL3 | 403 | 1.75e-03 |
| CCDC177 | 56936 | 2.88e-05 | CUL2 | 8453 | 1.78e-03 |
| NOX4 | 50507 | 3.07e-05 | C15orf41 | 84529 | 1.81e-03 |

Continued on next page

**Table 7.7 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| PDLIM7 | 9260 | 3.07e-05 | BTBD7 | 55727 | 1.84e-03 |
| MYO1E | 4643 | 3.15e-05 | UBE2J1 | 51465 | 1.84e-03 |
| FAM204A | 63877 | 3.33e-05 | CBLB | 868 | 1.86e-03 |
| RB1 | 5925 | 3.45e-05 | MKL2 | 57496 | 1.86e-03 |
| UNC5D | 137970 | 3.64e-05 | CUBN | 8029 | 1.86e-03 |
| ACTC1 | 70 | 3.64e-05 | ACTR3 | 10096 | 1.86e-03 |
| GYG2 | 8908 | 3.65e-05 | STARD4 | 134429 | 1.88e-03 |
| PLA2G3 | 50487 | 3.65e-05 | CAPG | 822 | 1.91e-03 |
| FARP1 | 10160 | 3.65e-05 | DHX32 | 55760 | 1.91e-03 |
| DOCK1 | 1793 | 3.86e-05 | TOR1AIP2 | 163590 | 1.93e-03 |
| SHOC2 | 8036 | 3.91e-05 | MLLT3 | 4300 | 1.93e-03 |
| MTCL1 | 23255 | 4.10e-05 | NDFIP2 | 54602 | 1.94e-03 |
| LRRIQ1 | 84125 | 4.10e-05 | TRHDE | 29953 | 1.96e-03 |
| SORCS2 | 57537 | 4.18e-05 | MEGF10 | 84466 | 1.98e-03 |
| KPNA3 | 3839 | 4.24e-05 | MYO5C | 55930 | 1.98e-03 |
| SFXN3 | 81855 | 4.32e-05 | ATXN1 | 6310 | 1.98e-03 |
| ME1 | 4199 | 4.54e-05 | WDR11 | 55717 | 1.98e-03 |
| FLNB | 2317 | 4.87e-05 | TDG | 6996 | 1.98e-03 |
| TUBGCP2 | 10844 | 6.06e-05 | ASCC3 | 10973 | 2.00e-03 |
|  | 11245 | 7.11e-05 | VAMP8 | 8673 | 2.00e-03 |
| CAPN2 | 824 | 7.17e-05 | TNFRSF10A | 8797 | 2.00e-03 |
| SMAD1 | 4086 | 7.17e-05 | GALNT7 | 51809 | 2.00e-03 |
| KCNS1 | 3787 | 7.29e-05 | PAN3 | 255967 | 2.07e-03 |
| DAAM1 | 23002 | 7.82e-05 | HTRA1 | 5654 | 2.11e-03 |
| MANEA | 79694 | 7.82e-05 | SRD5A1 | 6715 | 2.12e-03 |
| TNKS2 | 80351 | 8.28e-05 | MORC4 | 79710 | 2.15e-03 |
| CHRNB1 | 1140 | 8.50e-05 | FERMT1 | 55612 | 2.18e-03 |
| USP6NL | 9712 | 8.96e-05 | SDC2 | 6383 | 2.19e-03 |
| ANKS1B | 56899 | 9.17e-05 | HS3ST3B1 | 9953 | 2.30e-03 |
| MTHFD1L | 25902 | 9.98e-05 | BOK | 666 | 2.31e-03 |
| SIAH3 | 283514 | 1.01e-04 | SLC25A24 | 29957 | 2.33e-03 |
| RNF182 | 221687 | 1.01e-04 | TOM1L2 | 146691 | 2.33e-03 |
| NEDD1 | 121441 | 1.02e-04 | BMPR1B | 658 | 2.34e-03 |
| MMP10 | 4319 | 1.03e-04 | RAB4A | 5867 | 2.38e-03 |
| RAP1GDS1 | 5910 | 1.04e-04 | TACC1 | 6867 | 2.38e-03 |
| CHRFAM7A | 89832 | 1.04e-04 | ZHX2 | 22882 | 2.38e-03 |
| PBLD | 64081 | 1.05e-04 | GLRX | 2745 | 2.38e-03 |
| TP53INP1 | 94241 | 1.20e-04 | ECHDC1 | 55862 | 2.39e-03 |
| TMEM163 | 81615 | 1.22e-04 | XPNPEP1 | 7511 | 2.40e-03 |

Continued on next page

**Table 7.7 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| CCNY | 219771 | 1.24e-04 | HNRNPF | 3185 | 2.40e-03 |
| FLNC | 2318 | 1.25e-04 | OPHN1 | 4983 | 2.41e-03 |
| COL25A1 | 84570 | 1.26e-04 | CYFIP1 | 23191 | 2.46e-03 |
| MMS19 | 64210 | 1.26e-04 | ANXA11 | 311 | 2.46e-03 |
| TUSC3 | 7991 | 1.27e-04 | IGF1R | 3480 | 2.46e-03 |
| DNMBP | 23268 | 1.29e-04 | RFTN1 | 23180 | 2.49e-03 |
| IRX5 | 10265 | 1.32e-04 | TMEM132D | 121256 | 2.50e-03 |
| SCHIP1 | 29970 | 1.33e-04 | HOMER2 | 9455 | 2.51e-03 |
| SSPN | 8082 | 1.36e-04 | STX7 | 8417 | 2.58e-03 |
| INPP5F | 22876 | 1.36e-04 | IDE | 3416 | 2.62e-03 |
| FZD3 | 7976 | 1.40e-04 | CD99 | 4267 | 2.67e-03 |
| TNFRSF10C | 8794 | 1.45e-04 | TCTN3 | 26123 | 2.68e-03 |
| HACD1 | 9200 | 1.56e-04 | EPHB2 | 2048 | 2.70e-03 |
| PLD3 | 23646 | 1.63e-04 | C10orf76 | 79591 | 2.74e-03 |
| H2AFY2 | 55506 | 1.63e-04 | NT5C2 | 22978 | 2.74e-03 |
| ICK | 22858 | 1.63e-04 | WWC2 | 80014 | 2.76e-03 |
| B3GALT5-AS1 | 114041 | 1.65e-04 | FAM172A | 83989 | 2.78e-03 |
| NR2F2 | 7026 | 1.68e-04 | ZIC1 | 7545 | 2.78e-03 |
| ARHGAP10 | 79658 | 1.68e-04 | ISYNA1 | 51477 | 2.81e-03 |
| EFNB2 | 1948 | 1.68e-04 | LARGE1 | 9215 | 2.83e-03 |
| GRK5 | 2869 | 1.73e-04 | CAPZB | 832 | 2.90e-03 |
| B3GALT5 | 10317 | 1.75e-04 | TRAM2 | 9697 | 2.92e-03 |
| C11orf70 | 85016 | 1.82e-04 | DOCK5 | 80005 | 2.92e-03 |
| NUDT15 | 55270 | 1.87e-04 | FGD6 | 55785 | 2.94e-03 |
| DIAPH3 | 81624 | 1.95e-04 | BHLHE41 | 79365 | 3.02e-03 |
| HSPG2 | 3339 | 1.95e-04 | POLR1D | 51082 | 3.07e-03 |
| RUNX1T1 | 862 | 1.95e-04 | CCKBR | 887 | 3.10e-03 |
| PPP2R2D | 55844 | 2.03e-04 | TMOD3 | 29766 | 3.10e-03 |
| CCNJ | 54619 | 2.09e-04 | SESN1 | 27244 | 3.12e-03 |
| FNBP1 | 23048 | 2.15e-04 | TSPAN14 | 81619 | 3.12e-03 |
| ATE1 | 11101 | 2.15e-04 | SPTLC2 | 9517 | 3.12e-03 |
| MYL12A | 10627 | 2.17e-04 | SQLE | 6713 | 3.16e-03 |
| TMEM2 | 23670 | 2.18e-04 | RPS6KA6 | 27330 | 3.18e-03 |
| PTENP1 | 11191 | 2.23e-04 | AGPS | 8540 | 3.22e-03 |
| MAB21L2 | 10586 | 2.28e-04 | ZYG11A | 440590 | 3.28e-03 |
| ANTXR2 | 118429 | 2.36e-04 | KLF11 | 8462 | 3.29e-03 |
| PTCHD4 | 442213 | 2.37e-04 | MPZL1 | 9019 | 3.33e-03 |
| EIF2AK4 | 440275 | 2.37e-04 | SKP2 | 6502 | 3.33e-03 |
| UXS1 | 80146 | 2.38e-04 | FGF11 | 2256 | 3.38e-03 |

Continued on next page

**Table 7.7 – continued from previous page**

| Gene name | Entrez id | Adj. p-value | Gene name | Entrez id | Adj. p-value |
|---|---|---|---|---|---|
| ZMYND11 | 10771 | 2.45e-04 | PAK1 | 5058 | 3.40e-03 |
| TMEM98 | 26022 | 2.46e-04 | KDM5B | 10765 | 3.40e-03 |
| WDR17 | 116966 | 2.51e-04 | NCOA4 | 8031 | 3.46e-03 |
| GLUD1 | 2746 | 2.53e-04 | TMEM178A | 130733 | 3.48e-03 |
| PAX3 | 5077 | 2.60e-04 | FAM149B1 | 317662 | 3.50e-03 |

Table 7.7 Table of top 400 genes overexpressed in NS compared to GNS sorted by adjusted p-value.

| GOBPID | Pvalue | Term |
|---|---|---|
| GO:0034470 | 3.81e-03 | ncRNA processing |
| GO:0048704 | 3.81e-03 | embryonic skeletal system morphogenesis |
| GO:0034660 | 6.67e-03 | ncRNA metabolic process |
| GO:0048706 | 6.67e-03 | embryonic skeletal system development |
| GO:0007600 | 1.33e-02 | sensory perception |
| GO:0045333 | 1.33e-02 | cellular respiration |
| GO:0048663 | 1.33e-02 | neuron fate commitment |
| GO:0006820 | 1.45e-02 | anion transport |
| GO:0042254 | 1.98e-02 | ribosome biogenesis |
| GO:0043900 | 1.98e-02 | regulation of multi-organism process |
| GO:0022613 | 2.23e-02 | ribonucleoprotein complex biogenesis |
| GO:0045165 | 2.27e-02 | cell fate commitment |
| GO:0006364 | 2.48e-02 | rRNA processing |
| GO:0006986 | 2.48e-02 | response to unfolded protein |
| GO:0016072 | 2.48e-02 | rRNA metabolic process |
| GO:0034976 | 2.48e-02 | response to endoplasmic reticulum stress |
| GO:0035966 | 2.48e-02 | response to topologically incorrect protein |
| GO:0043903 | 2.48e-02 | regulation of symbiosis, encompassing mutualism through parasitism |
| GO:0050792 | 2.48e-02 | regulation of viral process |
| GO:0071241 | 2.48e-02 | cellular response to inorganic substance |
| GO:0071248 | 2.48e-02 | cellular response to metal ion |
| GO:0001501 | 2.51e-02 | skeletal system development |
| GO:0006139 | 2.86e-02 | nucleobase-containing compound metabolic process |
| GO:0055085 | 2.99e-02 | transmembrane transport |
| GO:0016070 | 3.15e-02 | RNA metabolic process |

Table 7.8 Top 25 Gene ontology terms for genes overexpressed in GNS cells. The gene universe used in this case was the combined GNS and NS differentially expressed gene sets.

| GOBPID | Pvalue | Term |
|---|---|---|
| GO:0032989 | 3.69e-06 | cellular component morphogenesis |
| GO:0000902 | 1.32e-05 | cell morphogenesis |
| GO:0009653 | 2.43e-05 | anatomical structure morphogenesis |
| GO:0032990 | 2.66e-05 | cell part morphogenesis |
| GO:0048858 | 2.66e-05 | cell projection morphogenesis |
| GO:0030030 | 1.41e-04 | cell projection organization |
| GO:0031175 | 1.55e-04 | neuron projection development |
| GO:0048869 | 2.05e-04 | cellular developmental process |
| GO:0048468 | 2.21e-04 | cell development |
| GO:0048666 | 2.34e-04 | neuron development |
| GO:0000904 | 2.56e-04 | cell morphogenesis involved in differentiation |
| GO:0048667 | 2.58e-04 | cell morphogenesis involved in neuron differentiation |
| GO:0030154 | 3.00e-04 | cell differentiation |
| GO:0048812 | 3.08e-04 | neuron projection morphogenesis |
| GO:0007167 | 3.34e-04 | enzyme linked receptor protein signaling pathway |
| GO:0007411 | 5.44e-04 | axon guidance |
| GO:0097485 | 5.44e-04 | neuron projection guidance |
| GO:0006936 | 5.61e-04 | muscle contraction |
| GO:0016043 | 8.38e-04 | cellular component organization |
| GO:0003012 | 1.05e-03 | muscle system process |
| GO:0032502 | 1.41e-03 | developmental process |
| GO:0048856 | 1.75e-03 | anatomical structure development |
| GO:0007369 | 1.82e-03 | gastrulation |
| GO:0044767 | 1.97e-03 | single-organism developmental process |
| GO:0006935 | 2.30e-03 | chemotaxis |

Table 7.9 Top 25 Gene ontology terms for genes overexpressed in NS cells. The gene universe used in this case was the combined GNS and NS differentially expressed gene sets.

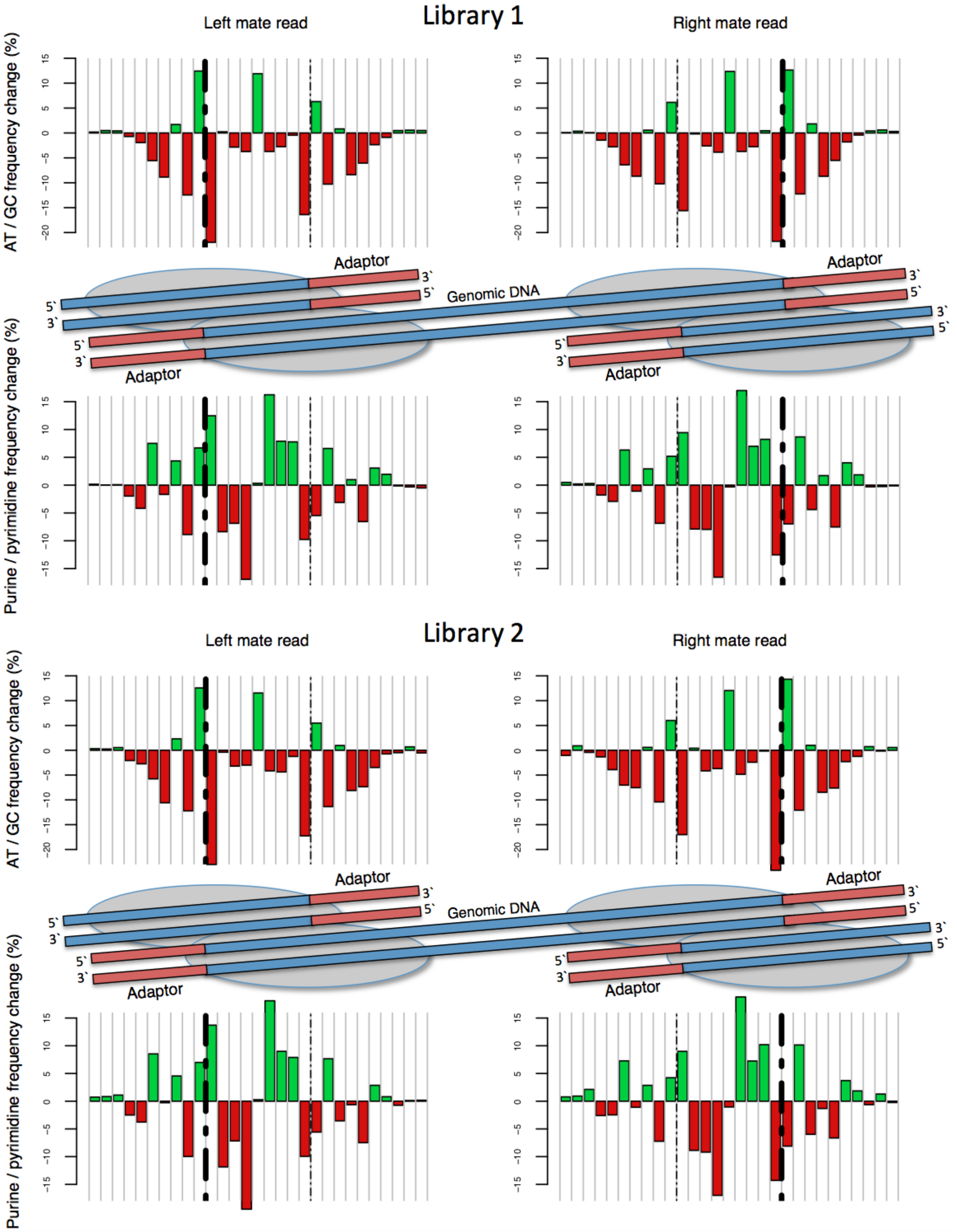


Fig. 7.9 Replicate transposase bias plots in a separate ATAC-seq library. See figure 5.3.

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| TAL1_f1 | TAL1 | 0 |
| ZN238_f1 | ZN238 | 0 |
| MA0091.1 | TAL1::TCF3 | 0 |
| ZNF238_DBD | | 0 |
| ZNF238_full | | 0 |
| MA0497.1 | MEF2C | 1.523e-303 |
| MA0095.2 | YY1 | 1.533e-295 |
| MA0052.2 | MEF2A | 2.902e-295 |
| TYY1_f2 | TYY1 | 1.059e-288 |
| TFE2_f2 | TFE2 | 1.64e-285 |
| MA0466.1 | CEBPB | 2.477e-280 |
| TFAP4_full | | 2.166e-277 |
| IRF1_si | IRF1 | 9.194e-277 |
| Atoh1_DBD | | 2.755e-276 |
| TAL1_f2 | TAL1 | 5.76e-272 |
| MA0522.1 | Tcf3 | 8.315e-268 |
| MA0481.1 | FOXP1 | 4.302e-267 |
| TCF4_DBD | | 6.17e-265 |
| FOXF1_f1 | FOXF1 | 6.928e-263 |
| TFAP4_DBD | | 7.133e-263 |
| FOXC1_DBD_1 | | 1.186e-261 |
| HTF4_f1 | HTF4 | 1.439e-258 |
| FOXJ3_f2 | FOXJ3 | 5.718e-258 |
| NEUROG2_DBD | | 3.467e-254 |
| FOXJ3_si | FOXJ3 | 1.398e-252 |
| MA0517.1 | STAT2::STAT1 | 5.819e-252 |
| NEUROG2_full | | 1.965e-251 |
| MA0102.3 | CEBPA | 4.019e-250 |
| MA0521.1 | Tcf12 | 4.672e-250 |
| MA0488.1 | JUN | 9.999e-247 |
| MEF2A_DBD | | 2.262e-245 |
| FOXC2_DBD_2 | | 3.04e-241 |
| MA0050.2 | IRF1 | 6.826e-241 |
| MYOD1_f1 | MYOD1 | 1.978e-239 |
| IRF2_f1 | IRF2 | 2.201e-239 |
| MEF2D_f1 | MEF2D | 2.034e-237 |
| MA0548.1 | AGL15 | 1.761e-236 |
| MA0058.2 | MAX | 4.795e-233 |
| Foxc1_DBD_1 | | 2.216e-232 |
| Tcf21_DBD | | 3.19e-231 |
| MA0545.1 | HLH-1 | 6.954e-231 |
| MEF2D_DBD | | 9.31e-229 |
| Foxj3_DBD_4 | | 1.795e-228 |
| IRF5_f1 | IRF5 | 2.513e-228 |
| MEF2A_f1 | MEF2A | 2.54e-228 |
| IRF7_DBD_1 | | 7.438e-227 |
| MA0537.1 | BLMP-1 | 1.926e-226 |
| MA0041.1 | Foxd3 | 9.534e-226 |
| YY2_full_1 | | 2.709e-225 |
| FOXD3_f1 | FOXD3 | 5.735e-225 |
| MA0593.1 | FOXP2 | 3.322e-224 |
| MA0045.1 | HMG-I/Y | 3.779e-222 |
| MA0480.1 | Foxo1 | 8.404e-221 |
| MITF_f1 | MITF | 2.148e-220 |
| Rarg_DBD_1 | | 3.533e-220 |
| MSC_full | | 6.09e-220 |
| IRF4_si | IRF4 | 1.752e-219 |
| Ascl2_DBD | | 2.207e-219 |
| MA0500.1 | Myog | 1.381e-218 |
| MA0559.1 | PI | 3.97e-218 |
| TCF3_DBD | | 1.258e-217 |
| FOXC2_DBD_3 | | 4.351e-217 |
| MA0558.1 | FLC | 5.403e-217 |
| IRF7_f1 | IRF7 | 7.573e-217 |
| FOXA2_f1 | FOXA2 | 7.903e-217 |
| FOXA1_f1 | FOXA1 | 3.035e-216 |
| IRF8_si | IRF8 | 7.135e-215 |
| MA0388.1 | SPT23 | 2.876e-214 |
| NDF1_f1 | NDF1 | 2.93e-214 |
| Nr2f6_DBD_1 | | 9.55e-214 |
| FOXO1_si | FOXO1 | 2.751e-213 |
| RARA_DBD_2 | | 4.898e-213 |
| MEF2C_f1 | MEF2C | 4.907e-213 |
| MEF2B_full | | 4.275e-212 |
| RARA_f1 | RARA | 4.303e-212 |
| FOXC2_f1 | FOXC2 | 2.793e-211 |
| RARB_full | | 5.792e-211 |
| FOXJ3_DBD_3 | | 8.88e-211 |
| FOXO3_si | FOXO3 | 7.724e-210 |
| NR2E3_f1 | NR2E3 | 3.513e-209 |
| MA0093.2 | USF1 | 5.235e-209 |
| NR1I3_si | NR1I3 | 2.173e-208 |
| MA0049.1 | hb | 6.383e-208 |
| FOXC1_DBD_3 | | 1.578e-206 |
| TFE3_f1 | TFE3 | 5.682e-206 |
| MA0160.1 | NR4A2 | 5.758e-206 |
| NR2F1_DBD_3 | | 7.26e-206 |
| OLIG3_DBD | | 8.352e-206 |
| MYF6_f1 | MYF6 | 1.086e-205 |
| MA0042.1 | FOXI1 | 1.843e-205 |
| MA0546.1 | PHA-4 | 2.489e-205 |
| CEBPA_do | CEBPA | 2.437e-204 |
| RARG_f1 | RARG | 1.105e-203 |
| MA0563.1 | SEP3 | 1.197e-203 |
| HMGA1_f1 | HMGA1 | 3.499e-203 |
| PRDM1_full | | 5.015e-203 |
| MA0277.1 | AZF1 | 7.924e-203 |
| MA0555.1 | SVP | 2.483e-202 |
| FOXJ2_DBD_3 | | 2.504e-202 |
| RARG_DBD_1 | | 1.085e-201 |
| FUBP1_f1 | FUBP1 | 2.345e-201 |
| FOXO1_DBD_1 | | 5.048e-201 |
| MA0377.1 | SFL1 | 5.114e-201 |
| RORA_f1 | RORA | 6.404e-201 |
| STAT2_f1 | STAT2 | 8.275e-201 |
| FOXP2_si | FOXP2 | 1.012e-200 |
| USF1_f1 | USF1 | 1.195e-200 |
| MA0492.1 | JUND | 2.39e-200 |
| FOXB1_DBD_3 | | 3.334e-200 |
| MA0508.1 | PRDM1 | 5.366e-200 |
| MA0561.1 | PIF4 | 6.851e-200 |
| FOXJ2_f1 | FOXJ2 | 1.092e-199 |
| PPARA_f1 | PPARA | 1.576e-199 |
| HXD13_f1 | HXD13 | 1.655e-199 |
| Rara_DBD_3 | | 5.001e-199 |
| PRDM1_f1 | PRDM1 | 8.143e-199 |
| NR2F1_full | | 2.314e-198 |
| MA0526.1 | USF2 | 6.429e-198 |
| FOXL1_full_2 | | 1.081e-197 |
| Rarb_DBD_1 | | 1.091e-197 |
| IRF8_DBD | | 2.074e-197 |
| Foxj3_DBD_3 | | 2.682e-197 |
| COT2_f1 | COT2 | 3.474e-197 |
| MA0147.2 | Myc | 5.394e-197 |
| CEBPB_f1 | CEBPB | 8.158e-197 |
| FOXQ1_f1 | FOXQ1 | 1.288e-196 |
| MA0296.1 | FKH1 | 1.968e-196 |
| HOXB13_DBD_1 | | 4.13e-196 |
| MA0148.3 | FOXA1 | 5.239e-196 |
| VDR_f1 | VDR | 1.524e-195 |
| CEBPD_f1 | CEBPD | 3.803e-195 |
| MA0071.1 | RORA_1 | 1.059e-194 |
| FOXF2_f1 | FOXF2 | 1.404e-194 |
| MA0157.1 | FOXO3 | 1.693e-194 |
| RARA_full_1 | | 2.061e-194 |
| MA0458.1 | slp1 | 4.074e-194 |
| FOXO4_f1 | FOXO4 | 4.771e-194 |
| HOXC13_DBD_1 | | 6.318e-194 |
| IRF3_f1 | IRF3 | 3.223e-193 |
| COT1_si | COT1 | 1.598e-192 |
| MA0113.2 | NR3C1 | 3.544e-192 |
| FOXA3_f1 | FOXA3 | 8.294e-192 |
| MA0556.1 | AP3 | 2.098e-191 |
| FOXM1_f1 | FOXM1 | 1.503e-190 |
| MA0560.1 | PIF3 | 3.037e-190 |
| MA0047.2 | Foxa2 | 4.872e-190 |
| TEAD4_f1 | TEAD4 | 1.436e-189 |
| MA0398.1 | SUM1 | 4.088e-189 |
| SPDEF_DBD_3 | | 5.464e-189 |
| TBX1_DBD_1 | | 1.241e-188 |
| PRGR_do | PRGR | 1.743e-188 |
| TBX3_f1 | TBX3 | 2.809e-188 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| NR6A1_do | NR6A1 | 1.081e-187 |
| Foxg1_DBD_1 | | 2.24e-187 |
| NFIA_full_2 | | 3.001e-187 |
| CDX1_DBD | | 5.506e-187 |
| FOXJ2_DBD_2 | | 4.219e-186 |
| Spic_DBD | | 4.455e-186 |
| HOXA13_full_1 | | 7.448e-186 |
| NR2F6_DBD_1 | | 1.272e-185 |
| NFAC2_f1 | NFAC2 | 1.242e-184 |
| RARG_full_1 | | 3.065e-184 |
| NHLH1_full | | 3.162e-184 |
| MA0051.1 | IRF2 | 1.101e-183 |
| HOXC10_DBD_1 | | 1.965e-183 |
| NFAC3_f1 | NFAC3 | 2.12e-183 |
| MA0554.1 | SOC1 | 2.96e-183 |
| TFEB_full | | 1.39e-182 |
| SPI1_full | | 2.142e-182 |
| HOXA13_DBD_1 | | 2.415e-182 |
| HOXA13_full_2 | | 3.122e-182 |
| MA0461.1 | Atoh1 | 3.395e-182 |
| IRF8_full | | 3.681e-182 |
| HOXC10_DBD_2 | | 4.75e-182 |
| RARB_f1 | RARB | 8.806e-182 |
| MA0446.1 | fkh | 1.39e-181 |
| NFIX_full_3 | | 1.586e-181 |
| YY1_full | | 2.861e-181 |
| IRF9_full | | 3.182e-181 |
| BPTF_si | BPTF | 6.358e-181 |
| SPIB_DBD | | 7.469e-181 |
| MYF6_full | | 9.555e-181 |
| HOXC11_full_2 | | 2.535e-180 |
| MYOG_f1 | MYOG | 8.318e-179 |
| FOXB1_full | | 9.034e-179 |
| ESRRB_DBD | | 1.58e-178 |
| HXA13_f1 | HXA13 | 4.355e-178 |
| MA0512.1 | Rxra | 5.568e-178 |
| ATF5_si | ATF5 | 8.691e-178 |
| HOXD13_DBD_1 | | 2.232e-177 |
| PRGR_f1 | PRGR | 6.211e-177 |
| OLIG1_DBD | | 5.177e-176 |
| TFEC_DBD | | 1.594e-175 |
| NR2F1_DBD_2 | | 2.108e-175 |
| FIGLA_DBD | | 2.706e-175 |
| Hoxd13_DBD_1 | | 3.289e-175 |
| CDX2_DBD | | 4.43e-175 |
| Rarb_DBD_2 | | 4.63e-175 |
| FOXC1_f1 | FOXC1 | 7.599e-175 |
| MEIS3_DBD_2 | | 1.006e-174 |
| ERR1_f1 | ERR1 | 1.794e-174 |
| CEBPG_si | CEBPG | 2.562e-174 |
| MEIS2_DBD_1 | | 4.728e-174 |
| MA0007.2 | AR | 1.071e-173 |
| MA0115.1 | NR1H2::RXRA | 2.525e-173 |
| ESRRA_DBD_1 | | 2.886e-173 |
| Meis2_DBD_2 | | 3.906e-173 |
| NR4A1_f1 | NR4A1 | 1.465e-172 |
| NFAC4_f1 | NFAC4 | 1.713e-172 |
| MA0100.2 | Myb | 1.901e-172 |
| CPEB1_full | | 2.542e-172 |
| USF1_DBD | | 4.103e-172 |
| FOXG1_DBD_1 | | 5.812e-172 |
| MA0287.1 | CUP2 | 6.836e-172 |
| RORA_DBD_2 | | 7.168e-172 |
| MA0464.1 | Bhlhe40 | 7.748e-172 |
| MA0562.1 | PIF5 | 1.247e-171 |
| TBX4_DBD_1 | | 2.075e-171 |
| ESRRA_DBD_4 | | 2.418e-171 |
| ESR2_si | ESR2 | 6.797e-171 |
| Meis3_DBD_2 | | 1.847e-170 |
| RXRG_f1 | RXRG | 2.094e-170 |
| PKNOX2_DBD | | 8.685e-170 |
| Rarg_DBD_3 | | 2.593e-169 |
| CDX1_f1 | CDX1 | 5.278e-169 |
| HOXD11_DBD_2 | | 1.954e-168 |
| OLIG2_DBD | | 2.378e-168 |
| PPARG_si | PPARG | 6.458e-168 |
| MA0040.1 | Foxq1 | 1.191e-167 |
| GCR_si | GCR | 2.894e-167 |
| Foxc1_DBD_2 | | 1.872e-166 |
| EHF_si | EHF | 1.911e-166 |
| IRF9_f1 | IRF9 | 2.907e-166 |
| SOX17_f2 | SOX17 | 5.496e-166 |
| TBX20_full_2 | | 1.08e-165 |
| MA0043.1 | HLF | 1.137e-165 |
| PKNOX1_DBD | | 2.012e-165 |
| ZN384_f1 | ZN384 | 4.673e-165 |
| MA0165.1 | Abd-B | 6.168e-165 |
| MA0174.1 | CG42234 | 6.168e-165 |
| Rara_DBD_2 | | 7.283e-165 |
| MEIS3_DBD_1 | | 1.305e-164 |
| ID4_DBD | | 1.476e-164 |
| FOXJ2_DBD_1 | | 1.521e-164 |
| MA0465.1 | CDX2 | 1.648e-164 |
| MA0030.1 | FOXF2 | 1.855e-164 |
| MA0019.1 | Ddit3::Cebpa | 5.685e-164 |
| MA0321.1 | INO2 | 1.391e-163 |
| OLIG2_full | | 2.718e-163 |
| NR4A2_si | NR4A2 | 1.757e-162 |
| MA0319.1 | HSF1 | 1.833e-162 |
| FOXI1_full_1 | | 2.157e-162 |
| FOXO4_DBD_2 | | 2.157e-162 |
| FOXO6_DBD_2 | | 2.157e-162 |
| MA0582.1 | RAV1 | 4.725e-162 |
| Nr2f6_DBD_2 | | 5.661e-162 |
| SPIC_full | | 1.451e-161 |
| NKX2-8_DBD | | 1.919e-161 |
| MA0247.2 | tin | 3.943e-161 |
| HOXC12_DBD_1 | | 4.47e-161 |
| NR4A2_full_3 | | 5.84e-161 |
| Hoxc10_DBD_2 | | 9.619e-161 |
| MA0010.1 | br_Z1 | 2.038e-160 |
| NFIX_full_2 | | 3.181e-160 |
| MA0451.1 | kni | 4.949e-160 |
| MA0152.1 | NFATC2 | 1.379e-159 |
| RXRG_DBD_2 | | 1.429e-159 |
| NR1D1_f1 | NR1D1 | 1.595e-159 |
| ESRRG_full_3 | | 1.671e-159 |
| TBX5_si | TBX5 | 1.821e-159 |
| MA0136.1 | ELF5 | 1.867e-159 |
| Srebf1_DBD | | 2.197e-159 |
| HOXC11_DBD_2 | | 5.532e-159 |
| NR1H2_f1 | NR1H2 | 8.98e-159 |
| MA0017.1 | NR2F1 | 1.03e-158 |
| Hoxd9_DBD_1 | | 1.157e-158 |
| Rarg_DBD_2 | | 1.926e-158 |
| MA0407.1 | THI2 | 2.577e-158 |
| ESRRA_DBD_5 | | 2.654e-158 |
| FEV_f1 | FEV | 2.898e-158 |
| MA0249.1 | twi | 3.965e-158 |
| Hic1_DBD_2 | | 4.713e-158 |
| NR2F6_DBD_2 | | 5.206e-158 |
| TGIF2_DBD | | 7.865e-158 |
| NR4A3_f1 | NR4A3 | 1.023e-157 |
| HOXA10_DBD_2 | | 5.134e-157 |
| SPIB_f1 | SPIB | 6.962e-157 |
| NR1I2_f2 | NR1I2 | 1.152e-156 |
| BHLHA15_DBD | | 2.501e-156 |
| RARA_DBD_1 | | 4.246e-156 |
| MA0092.1 | Hand1::Tcfe2a | 5.21e-156 |
| ONEC2_si | ONEC2 | 6.193e-156 |
| NEUROD2_full | | 9.252e-156 |
| Hoxa11_DBD_2 | | 2.378e-155 |
| IRF3_full | | 2.62e-155 |
| HXC6_f1 | HXC6 | 2.944e-155 |
| MEIS2_DBD_2 | | 3.498e-155 |
| NKX2-8_full | | 5.91e-155 |
| PO6F1_f1 | PO6F1 | 9.718e-155 |
| RXRG_full_1 | | 9.96e-155 |
| GLI3_si | GLI3 | 1.056e-154 |
| MA0141.2 | Esrrb | 1.151e-154 |
| HNF4A_full_4 | | 1.253e-154 |
| MA0083.2 | SRF | 1.437e-154 |
| MA0378.1 | SFP1 | 1.599e-154 |
| Esrra_DBD_2 | | 1.689e-154 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| NR2F6_full | | 1.855e-154 |
| HNF4A_DBD_1 | | 4.765e-154 |
| TBR1_full | | 7.054e-154 |
| RXRB_f1 | RXRB | 9.763e-154 |
| MA0346.1 | NHP6B | 1.084e-153 |
| MA0082.1 | squamosa | 2.46e-153 |
| THB_f1 | THB | 3.011e-153 |
| NHLH1_DBD | | 4.957e-153 |
| MA0156.1 | FEV | 6.239e-153 |
| NR2C1_si | NR2C1 | 6.392e-153 |
| MA0592.1 | ESRRA | 1.285e-152 |
| TBX20_full_1 | | 1.44e-152 |
| Atf4_DBD | | 1.911e-152 |
| RXRA_full_1 | | 5.162e-152 |
| Rarb_DBD_3 | | 5.34e-152 |
| ARI3A_do | ARI3A | 9.331e-152 |
| BHLHE22_DBD | | 1.352e-151 |
| HNF4A_full_1 | | 1.824e-151 |
| TBX5_DBD_1 | | 2.205e-151 |
| NFAT5_f1 | NFAT5 | 9.151e-151 |
| ZNF410_DBD | | 3.095e-150 |
| RORA_DBD_1 | | 3.264e-150 |
| HOXD12_DBD_1 | | 3.325e-150 |
| MGA_DBD_1 | | 3.733e-150 |
| ATF2+ATF4_f1 | ATF2+ATF4 | 3.952e-150 |
| TBX20_DBD_1 | | 4.356e-150 |
| EGR3_f1 | EGR3 | 4.976e-150 |
| NR2C2_DBD | | 5.77e-150 |
| TBX2_f1 | TBX2 | 8.723e-150 |
| MEIS2_do | MEIS2 | 1.08e-149 |
| NFIA+NFIB+NFIC_si | NFIA+NFIB+NFIC | 2.037e-149 |
| MA0368.1 | RIM101 | 3.543e-149 |
| MA0585.1 | SHP1 | 4.735e-149 |
| SPDEF_full_3 | | 6.04e-149 |
| NR1I2_si | NR1I2 | 1.005e-148 |
| IRF4_full | | 1.163e-148 |
| MA0534.1 | EcR::usp | 1.282e-148 |
| NR2E1_full_1 | | 1.389e-148 |
| MA0322.1 | INO4 | 2.244e-148 |
| COT1_f1 | COT1 | 4.172e-148 |
| TGIF2LX_full | | 1.397e-147 |
| YY2_full_2 | | 1.431e-147 |
| ERR2_f1 | ERR2 | 1.461e-147 |
| TBX21_full_2 | | 1.575e-147 |
| Nr2e1_DBD_1 | | 2.132e-147 |
| MA0390.1 | STB3 | 2.355e-147 |
| MA0499.1 | Myod1 | 2.654e-147 |
| THA_f1 | THA | 4.831e-147 |
| HIC2_DBD | | 7.735e-147 |
| TBX20_DBD_3 | | 1.737e-146 |
| TBX15_DBD_2 | | 1.883e-146 |
| Foxj3_DBD_2 | | 1.932e-146 |
| RARG_full_2 | | 3.189e-146 |
| Hoxd9_DBD_2 | | 3.323e-146 |
| RARG_DBD_3 | | 5.242e-146 |
| MA0459.1 | tll | 8.823e-146 |
| IKZF1_f1 | IKZF1 | 1.348e-145 |
| ESRRG_full_2 | | 1.602e-145 |
| ESR1_DBD | | 1.785e-145 |
| Hnf4a_DBD | | 1.938e-145 |
| MA0547.1 | SKN-1 | 3.168e-145 |
| NR2F6_f1 | NR2F6 | 4.543e-145 |
| MA0498.1 | Meis1 | 5.801e-145 |
| RHOXF1_full_2 | | 1.326e-144 |
| RARG_full_3 | | 1.972e-144 |
| GCR_do | GCR | 2.445e-144 |
| TFE3_DBD | | 3.194e-144 |
| HAND1_si | HAND1 | 7.663e-144 |
| MA0033.1 | FOXL1 | 8.573e-144 |
| RARA_full_3 | | 8.919e-144 |
| TGIF1_DBD | | 1.088e-143 |
| MA0409.1 | TYE7 | 1.875e-143 |
| DBP_si | DBP | 2.018e-143 |
| ELF5_f1 | ELF5 | 3.387e-143 |
| CEBPB_full | | 5.275e-143 |
| RXRB_DBD | | 6.028e-143 |
| ERR3_f1 | ERR3 | 6.126e-143 |
| AIRE_f2 | AIRE | 6.682e-143 |
| MA0133.1 | BRCA1 | 9.463e-143 |
| MA0359.1 | RAP1 | 1.59e-142 |
| NFATC1_full_1 | | 1.817e-142 |
| ESRRA_DBD_2 | | 2.23e-142 |
| USF2_f1 | USF2 | 7.299e-142 |
| MA0598.1 | EHF | 7.54e-142 |
| TBR1_DBD | | 9.636e-142 |
| NR2F1_DBD_1 | | 9.67e-142 |
| Pknox2_DBD | | 1.278e-141 |
| EOMES_DBD_1 | | 1.881e-141 |
| MA0595.1 | SREBF1 | 2.069e-141 |
| ATF4_DBD | | 2.959e-141 |
| Esrra_DBD_1 | | 3.813e-141 |
| TBX1_DBD_3 | | 5.676e-141 |
| SOX8_DBD_3 | | 5.753e-141 |
| RXRG_DBD_1 | | 7.774e-141 |
| SOX10_full_1 | | 1.94e-140 |
| NFIL3_DBD | | 3.536e-140 |
| Rxrb_DBD | | 3.762e-140 |
| ETS1_si | ETS1 | 5.276e-140 |
| PRDM4_full | | 5.471e-140 |
| TBX21_DBD_2 | | 8.464e-140 |
| CEBPD_DBD | | 1.101e-139 |
| MA0474.1 | Erg | 2.239e-139 |
| NR1I3_f2 | NR1I3 | 2.621e-139 |
| RARA_full_2 | | 2.864e-139 |
| MA0473.1 | ELF1 | 3.286e-139 |
| Rxra_DBD_1 | | 4.237e-139 |
| SREBF2_DBD | | 4.706e-139 |
| TBX5_DBD_2 | | 4.898e-139 |
| Tp53_DBD_1 | | 5.222e-139 |
| RARG_DBD_2 | | 1.102e-138 |
| HXD10_f1 | HXD10 | 1.602e-138 |
| ELF5_DBD | | 1.948e-138 |
| MA0048.1 | NHLH1 | 2.639e-138 |
| MA0503.1 | Nkx2-5 | 4.963e-138 |
| NFAC1_si | NFAC1 | 7.027e-138 |
| MA0013.1 | br_Z4 | 9.157e-138 |
| SPI1_si | SPI1 | 1.92e-137 |
| MA0119.1 | TLX1::NFIC | 2.118e-137 |
| Rara_DBD_1 | | 3.314e-137 |
| MEIS1_DBD | | 9.167e-137 |
| Meis2_DBD_1 | | 9.167e-137 |
| ELF3_f1 | ELF3 | 1.072e-136 |
| MA0270.1 | AFT2 | 1.917e-136 |
| COT2_f2 | COT2 | 2.037e-136 |
| THB_do | THB | 2.399e-136 |
| MA0596.1 | SREBF2 | 2.405e-136 |
| MEIS1_f2 | MEIS1 | 2.631e-136 |
| TBX21_full_1 | | 3.031e-136 |
| TBX4_DBD_2 | | 3.233e-136 |
| E2F2_DBD_1 | | 8.771e-136 |
| PITX2_si | PITX2 | 1.396e-135 |
| ZSCAN4_full | | 1.602e-135 |
| TBX2_full_2 | | 1.869e-135 |
| FOXJ3_DBD_2 | | 2.031e-135 |
| RARA_DBD_3 | | 3.374e-135 |
| RXRA_f1 | RXRA | 3.452e-135 |
| BRCA1_f1 | BRCA1 | 5.539e-135 |
| MA0255.1 | z | 6.77e-135 |
| TBX20_DBD_2 | | 8.464e-135 |
| MA0468.1 | DUX4 | 9.765e-135 |
| ELF5_full | | 9.915e-135 |
| MA0584.1 | SEP1 | 1.011e-134 |
| ELF3_full | | 1.561e-134 |
| SOX9_full_4 | | 3.448e-134 |
| GLI1_f1 | GLI1 | 4.178e-134 |
| MA0012.1 | br_Z3 | 4.506e-134 |
| MA0068.1 | Pax4 | 5.503e-134 |
| KLF3_f1 | KLF3 | 1.489e-133 |
| ANDR_do | ANDR | 1.557e-133 |
| Elf5_DBD | | 2.001e-133 |
| MA0484.1 | HNF4G | 2.357e-133 |
| NFIX_full_4 | | 3.446e-133 |
| TFAP4_si | TFAP4 | 4.152e-133 |
| MA0493.1 | Klf1 | 9.449e-133 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| NKX21_f1 | NKX21 | 9.602e-133 |
| MA0104.3 | Mycn | 1.001e-132 |
| NKX32_f1 | NKX32 | 3.504e-132 |
| HNF4A_full_2 | | 3.561e-132 |
| MA0475.1 | FLI1 | 2.105e-131 |
| MLXIPL_full | | 2.177e-131 |
| HOXC13_DBD_2 | | 3.265e-131 |
| RXRA_DBD_1 | | 3.813e-131 |
| HLTF_f1 | HLTF | 6.502e-131 |
| JDP2_DBD_2 | | 9.72e-131 |
| MA0018.2 | CREB1 | 1.001e-130 |
| HOXD8_DBD | | 1.374e-130 |
| MA0494.1 | Nr1h3::Rxra | 1.497e-130 |
| ELF3_DBD | | 1.938e-130 |
| MA0080.3 | Spi1 | 1.958e-130 |
| ETV5_f1 | ETV5 | 4.339e-130 |
| HNF4A_full_3 | | 1.128e-129 |
| STAT4_si | STAT4 | 1.249e-129 |
| MA0302.1 | GAT4 | 1.255e-129 |
| MA0288.1 | CUP9 | 1.33e-129 |
| SOX4_f1 | SOX4 | 1.68e-129 |
| ESRRA_DBD_6 | | 2.528e-129 |
| MA0304.1 | GCR1 | 4.014e-129 |
| ETV6_full_2 | | 6.02e-129 |
| MA0518.1 | Stat4 | 7.189e-129 |
| Cebpb_DBD | | 1.438e-128 |
| Jdp2_DBD_2 | | 3.408e-128 |
| FOXD3_DBD_1 | | 3.875e-128 |
| MA0072.1 | RORA_2 | 8.131e-128 |
| ELF1_f1 | ELF1 | 1.419e-127 |
| ESR2_do | ESR2 | 1.681e-127 |
| DDIT3_f1 | DDIT3 | 2.982e-127 |
| P63_si | P63 | 3.224e-127 |
| MA0306.1 | GIS1 | 4.551e-127 |
| THRB_DBD_3 | | 7.013e-127 |
| SOX4_DBD | | 1.483e-126 |
| MA0111.1 | Spz1 | 2.003e-126 |
| CEBPE_DBD | | 2.228e-126 |
| CREM_f1 | CREM | 4.42e-126 |
| MA0216.2 | CAD | 4.438e-126 |
| ESRRG_full_1 | | 5.108e-126 |
| SOX8_full_1 | | 5.297e-126 |
| Hoxd13_DBD_2 | | 5.409e-126 |
| ZNF143_DBD | | 5.608e-126 |
| MA0031.1 | FOXD1 | 9.167e-126 |
| MA0336.1 | MGA1 | 1.021e-125 |
| SRY_f1 | SRY | 1.312e-125 |
| MA0098.2 | Ets1 | 1.327e-125 |
| FOXJ3_DBD_1 | | 1.864e-125 |
| HNF4A_f1 | HNF4A | 1.883e-125 |
| CLOCK_DBD | | 1.891e-125 |
| MA0323.1 | IXR1 | 1.993e-125 |
| Vdr_DBD | | 3.191e-125 |
| NFAT5_DBD | | 4.181e-125 |
| RARG_do | RARG | 4.319e-125 |
| CEBPB_DBD | | 1.577e-124 |
| Hoxc10_DBD_1 | | 3.044e-124 |
| FOXO3_full_2 | | 3.091e-124 |
| FOXD2_DBD_1 | | 3.25e-124 |
| FOXD1_si | FOXD1 | 4.256e-124 |
| MA0328.1 | MATALPHA2 | 5.242e-124 |
| MA0182.1 | CG4328 | 5.405e-124 |
| TLX1_f1 | TLX1 | 6.609e-124 |
| ETV2_DBD | | 7.051e-124 |
| MA0389.1 | SRD1 | 1.012e-123 |
| HNF4G_f1 | HNF4G | 1.059e-123 |
| JDP2_full_2 | | 1.167e-123 |
| MA0144.2 | STAT3 | 1.246e-123 |
| MA0569.1 | MYC4 | 1.363e-123 |
| TBX1_DBD_2 | | 1.542e-123 |
| MA0301.1 | GAT3 | 1.781e-123 |
| SCRT1_DBD | | 1.962e-123 |
| MA0029.1 | Mecom | 2.117e-123 |
| HOXB13_DBD_2 | | 2.887e-123 |
| THRA_FL | | 3.154e-123 |
| Sox10_DBD_1 | | 5.431e-123 |
| MA0253.1 | vnd | 6.935e-123 |
| MA0130.1 | ZNF354C | 7.074e-123 |
| MA0543.1 | EOR-1 | 7.124e-123 |
| MA0114.2 | HNF4A | 1.145e-122 |
| ZNF75A_DBD | | 1.411e-122 |
| EVI1_f1 | EVI1 | 2.03e-122 |
| MA0022.1 | dl_1 | 2.394e-122 |
| HSF1_f2 | HSF1 | 3.577e-122 |
| NR2E1_full_2 | | 5.663e-122 |
| STF1_f1 | STF1 | 6.697e-122 |
| MCR_f1 | MCR | 7.155e-122 |
| Tp53_DBD_2 | | 9.374e-122 |
| ZEP1_f1 | ZEP1 | 1.695e-121 |
| VDR_full | | 2.024e-121 |
| MA0128.1 | EmBP-1 | 2.061e-121 |
| ENOA_si | ENOA | 2.748e-121 |
| ELF1_full | | 3.511e-121 |
| MA0269.1 | AFT1 | 1.056e-120 |
| MA0297.1 | FKH2 | 1.386e-120 |
| FOXD2_DBD_2 | | 1.386e-120 |
| FOXD3_DBD_2 | | 1.386e-120 |
| FOXL1_full_1 | | 1.386e-120 |
| FOXP3_DBD | | 1.386e-120 |
| Foxg1_DBD_3 | | 1.386e-120 |
| Foxk1_DBD_2 | | 1.386e-120 |
| MA0372.1 | RPH1 | 1.644e-120 |
| NFIB_full | | 1.652e-120 |
| SCRT2_DBD | | 2.44e-120 |
| MA0505.1 | Nr5a2 | 3.061e-120 |
| TGIF1_si | TGIF1 | 3.184e-120 |
| TEF_f1 | TEF | 3.612e-120 |
| TP63_DBD | | 4.819e-120 |
| MA0127.1 | PEND | 5.137e-120 |
| MA0122.1 | Nkx3-2 | 8.777e-120 |
| MA0108.2 | TBP | 1.603e-119 |
| MA0205.1 | Trl | 2.283e-119 |
| CEBPE_f1 | CEBPE | 3.725e-119 |
| ZNF282_DBD | | 5.437e-119 |
| MA0452.2 | KR | 5.997e-119 |
| ATF3_f1 | ATF3 | 6.977e-119 |
| NKX25_f1 | NKX25 | 7.295e-119 |
| PPARA_f2 | PPARA | 1.198e-118 |
| HBP1_f1 | HBP1 | 2.455e-118 |
| MA0159.1 | RXR::RAR_DR5 | 4.43e-118 |
| ETV7_si | ETV7 | 5.725e-118 |
| NR5A2_f1 | NR5A2 | 6.169e-118 |
| MA0566.1 | MYC2 | 6.577e-118 |
| E2F1_DBD_4 | | 7.613e-118 |
| NR3C1_DBD | | 7.802e-118 |
| YY2_DBD | | 9.923e-118 |
| BHLHE23_DBD | | 1.296e-117 |
| FOXB1_DBD_2 | | 1.587e-117 |
| Creb5_DBD | | 1.816e-117 |
| ELK4_f1 | ELK4 | 2.542e-117 |
| RHOXF1_DBD_2 | | 2.604e-117 |
| GCM1_full_1 | | 3.201e-117 |
| TBX15_DBD_1 | | 3.446e-117 |
| MA0103.2 | ZEB1 | 4.639e-117 |
| ETV4_f1 | ETV4 | 5.874e-117 |
| Sox11_DBD | | 6.141e-117 |
| MA0076.2 | ELK4 | 8.103e-117 |
| TBX2_full_1 | | 1.281e-116 |
| ELF4_full | | 2.614e-116 |
| NR3C2_DBD | | 4.765e-116 |
| ZNF524_full_1 | | 6.523e-116 |
| MA0423.1 | YER130C | 7.298e-116 |
| VDR_f2 | VDR | 1.423e-115 |
| MA0369.1 | RLM1 | 1.807e-115 |
| MA0408.1 | TOS8 | 2.054e-115 |
| E2F7_DBD | | 2.218e-115 |
| NANOG_f1 | NANOG | 2.485e-115 |
| PO3F2_si | PO3F2 | 3.251e-115 |
| THRB_DBD_2 | | 5.691e-115 |
| MA0538.1 | DAF-12 | 8.502e-115 |
| Creb3l2_DBD_1 | | 8.65e-115 |
| MA0440.1 | ZAP1 | 9.312e-115 |
| ZBTB6_si | ZBTB6 | 1.012e-114 |
| RXRA_full_2 | | 1.178e-114 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| TEAD3_si | TEAD3 | 2.325e-114 |
| Tp73_DBD | | 3.096e-114 |
| ZNF435_full | | 3.497e-114 |
| HXB1_f1 | HXB1 | 4.226e-114 |
| TFAP2C_DBD_2 | | 4.683e-114 |
| MA0472.1 | EGR2 | 6.58e-114 |
| MA0317.1 | HCM1 | 6.716e-114 |
| FOXG1_DBD_2 | | 7.281e-114 |
| TCF7L1_full | | 9.729e-114 |
| POU3F3_DBD_2 | | 1.27e-113 |
| ESRRA_DBD_3 | | 1.587e-113 |
| MA0001.2 | SEP4 | 2.124e-113 |
| HXA9_f1 | HXA9 | 2.152e-113 |
| HIC1_si | HIC1 | 2.561e-113 |
| HOXD12_DBD_4 | | 2.772e-113 |
| MGA_DBD_3 | | 4.049e-113 |
| RXRG_full_2 | | 4.197e-113 |
| SRF_do | SRF | 4.571e-113 |
| Foxg1_DBD_2 | | 4.789e-113 |
| MA0244.1 | slbo | 1.015e-112 |
| PPARG_f1 | PPARG | 1.019e-112 |
| MYBB_f1 | MYBB | 1.195e-112 |
| MESP1_DBD | | 1.235e-112 |
| MAX_f1 | MAX | 1.274e-112 |
| MA0263.1 | ttx-3::ceh-10 | 1.631e-112 |
| MA0027.1 | En1 | 1.665e-112 |
| POU3F1_DBD_2 | | 2.192e-112 |
| Foxk1_DBD_1 | | 2.877e-112 |
| MA0260.1 | che-1 | 4.907e-112 |
| NFYC_f1 | NFYC | 1.097e-111 |
| SOX15_full_1 | | 1.169e-111 |
| PPARD_f1 | PPARD | 1.744e-111 |
| MA0479.1 | FOXH1 | 2.105e-111 |
| EHF_full | | 3.031e-111 |
| NFIA_full_1 | | 3.077e-111 |
| OTX2_si | OTX2 | 3.236e-111 |
| MA0486.1 | HSF1 | 4.01e-111 |
| PIT1_f1 | PIT1 | 5.722e-111 |
| ZNF524_full_2 | | 6.377e-111 |
| AR_full | | 7.013e-111 |
| MA0502.1 | NFYB | 1.276e-110 |
| NFIA+NFIB+NFIC+NFIX_f2 | NFIA,B,C,X | 1.445e-110 |
| MA0314.1 | HAP3 | 1.767e-110 |
| MA0073.1 | RREB1 | 1.893e-110 |
| MAX_DBD_2 | | 2.41e-110 |
| MA0106.2 | TP53 | 3.679e-110 |
| MA0513.1 | SMAD2,3,4 | 4.346e-110 |
| Tp53_DBD_3 | | 6.357e-110 |
| THRB_DBD_1 | | 9.525e-110 |
| MA0109.1 | Hltf | 1.019e-109 |
| MLX_full | | 1.026e-109 |
| PROP1_f1 | PROP1 | 1.062e-109 |
| MA0137.3 | STAT1 | 1.215e-109 |
| MA0081.1 | SPIB | 2.105e-109 |
| MA0005.2 | AG | 2.622e-109 |
| HOXD13_DBD_2 | | 2.734e-109 |
| E2F8_DBD | | 2.891e-109 |
| ETS1_DBD_1 | | 3.159e-109 |
| MA0059.1 | MYC::MAX | 6.364e-109 |
| Hic1_DBD_1 | | 6.44e-109 |
| SOX9_full_1 | | 6.588e-109 |
| TBX1_DBD_5 | | 1.142e-108 |
| MA0293.1 | ECM23 | 1.737e-108 |
| MA0347.1 | NRG1 | 2.066e-108 |
| MA0009.1 | T | 2.098e-108 |
| MA0066.1 | PPARG | 2.578e-108 |
| NKX31_si | NKX31 | 4.15e-108 |
| TBP_f1 | TBP | 5.406e-108 |
| NFIX_full_1 | | 6.615e-108 |
| NFKB2_f1 | NFKB2 | 7.632e-108 |
| POU3F3_DBD_3 | | 7.792e-108 |
| SOX7_full_2 | | 8.709e-108 |
| MA0393.1 | STE12 | 2.887e-107 |
| FOXI1_f1 | FOXI1 | 2.896e-107 |
| CREB3_full_1 | | 3.056e-107 |
| ONECUT3_DBD | | 4.135e-107 |
| ETV6_full_1 | | 1.01e-106 |
| GNS motifs | Table 7.10 | |
| Motif name | TF name | Adjusted p-value |
| BMAL1_f1 | BMAL1 | 1.022e-106 |
| HSF2_DBD | | 1.687e-106 |
| NFYB_f1 | NFYB | 2.428e-106 |
| MA0403.1 | TBF1 | 2.517e-106 |
| LEF1_DBD | | 4.488e-106 |
| ATF7_DBD | | 4.489e-106 |
| GFI1_f1 | GFI1 | 7.132e-106 |
| LHX2_f1 | LHX2 | 8.073e-106 |
| ERF_DBD | | 9.196e-106 |
| MA0370.1 | RME1 | 1.31e-105 |
| MA0371.1 | ROX1 | 1.554e-105 |
| BATF3_DBD | | 5.842e-105 |
| SOX10_full_4 | | 5.958e-105 |
| HXD9_f1 | HXD9 | 5.962e-105 |
| MA0519.1 | Stat5a::Stat5b | 6.031e-105 |
| MA0011.1 | br_Z2 | 7.936e-105 |
| MA0514.1 | Sox3 | 8.047e-105 |
| MA0193.1 | Lag1 | 8.647e-105 |
| GABPA_f1 | GABPA | 1.246e-104 |
| MA0520.1 | Stat6 | 2.073e-104 |
| NR4A2_full_1 | | 2.198e-104 |
| NFYA_f1 | NFYA | 4.404e-104 |
| HXA10_f1 | HXA10 | 4.92e-104 |
| EGR2_si | EGR2 | 5.139e-104 |
| MA0065.2 | PPARG::RXRA | 7.334e-104 |
| Tcf7_DBD | | 8.536e-104 |
| MA0026.1 | Eip74EF | 9.332e-104 |
| MLXPL_f1 | MLXPL | 1.062e-103 |
| SOX21_DBD_1 | | 1.175e-103 |
| Nr2e1_DBD_2 | | 1.229e-103 |
| POU2F1_DBD_2 | | 1.406e-103 |
| GCM1_f1 | GCM1 | 1.475e-103 |
| SOX8_DBD_1 | | 1.815e-103 |
| HXA1_f1 | HXA1 | 3.247e-103 |
| E2F4_DBD_1 | | 4.148e-103 |
| ARI3A_f1 | ARI3A | 4.457e-103 |
| Egr1_mouse_mutantDBD | | 1.067e-102 |
| BRAC_si | BRAC | 1.452e-102 |
| MA0533.1 | SU(HW) | 1.61e-102 |
| Foxj3_DBD_1 | | 1.712e-102 |
| MA0120.1 | id1 | 2.15e-102 |
| E2F3_DBD_1 | | 2.922e-102 |
| ELK1_f1 | ELK1 | 2.964e-102 |
| NR1H4_f1 | NR1H4 | 3.631e-102 |
| MA0580.1 | DYT1 | 4.066e-102 |
| RORG_f1 | RORG | 4.129e-102 |
| FOXB1_DBD_1 | | 5.328e-102 |
| ATF1_si | ATF1 | 7.285e-102 |
| LEF1_f1 | LEF1 | 8.066e-102 |
| GABP1+GABP2_f1 | GABP1+GABP2 | 8.798e-102 |
| ZNF784_full | | 9.605e-102 |
| TLX1_f2 | TLX1 | 1.171e-101 |
| ELF1_DBD | | 1.207e-101 |
| RXRA_DBD_2 | | 1.234e-101 |
| HXA7_f1 | HXA7 | 1.572e-101 |
| FOXO3_full_3 | | 1.822e-101 |
| TBX1_DBD_4 | | 1.981e-101 |
| UBIP1_f1 | UBIP1 | 2.7e-101 |
| HOXA13_DBD_2 | | 2.802e-101 |
| MA0305.1 | GCR2 | 4.777e-101 |
| EGR2_full | | 5.935e-101 |
| Irx3_DBD | | 6.787e-101 |
| NFKB2_DBD | | 1.14e-100 |
| HNF4A_DBD_2 | | 1.261e-100 |
| SOX10_full_2 | | 1.417e-100 |
| P53_f2 | P53 | 1.818e-100 |
| ELK3_f1 | ELK3 | 2e-100 |
| THA_f2 | THA | 2.613e-100 |
| SOX8_DBD_5 | | 2.658e-100 |
| MA0056.1 | MZF1_1-4 | 3.756e-100 |
| FOXO1_DBD_2 | | 4.388e-100 |
| HOXC10_DBD_3 | | 4.961e-100 |
| NKX2-3_DBD | | 5.451e-100 |
| BHE40_f2 | BHE40 | 6.027e-100 |
| SUH_f1 | SUH | 9.512e-100 |
| MA0044.1 | HMG-1 | 9.673e-100 |
| ZNF713_full | | 1.131e-99 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| TFAP2A_DBD_4 | | 1.247e-99 |
| HSF4_DBD | | 1.661e-99 |
| CEBPZ_si | CEBPZ | 1.731e-99 |
| MA0060.2 | NFYA | 2.003e-99 |
| NKX28_f1 | NKX28 | 2.406e-99 |
| IRF7_DBD_2 | | 3.414e-99 |
| E2F3_DBD_3 | | 7.509e-99 |
| MA0110.2 | ATHB5 | 1.35e-98 |
| Sox10_DBD_3 | | 1.372e-98 |
| RARA_f2 | RARA | 2.185e-98 |
| NR2C2_f1 | NR2C2 | 2.201e-98 |
| MA0525.1 | TP63 | 3.408e-98 |
| TFAP2A_DBD_2 | | 5.946e-98 |
| STA5B_f1 | STA5B | 7.498e-98 |
| MA0032.1 | FOXC1 | 8.374e-98 |
| SPDEF_DBD_1 | | 8.642e-98 |
| FOXO4_DBD_3 | | 9.135e-98 |
| MA0454.1 | odd | 9.79e-98 |
| MA0143.3 | Sox2 | 1.112e-97 |
| POU2F2_DBD_2 | | 1.135e-97 |
| SOX9_full_3 | | 1.246e-97 |
| SOX10_full_5 | | 1.86e-97 |
| FLI1_full_1 | | 2.991e-97 |
| FOXC1_DBD_2 | | 3.374e-97 |
| FOXO1_DBD_3 | | 4.076e-97 |
| POU2F3_DBD_2 | | 4.388e-97 |
| ETS1_full_1 | | 1.172e-96 |
| MA0483.1 | Gfi1b | 1.364e-96 |
| ELF2_f1 | ELF2 | 1.488e-96 |
| GCM1_full_2 | | 1.967e-96 |
| MA0258.2 | ESR2 | 2.104e-96 |
| MA0121.1 | ARR10 | 2.467e-96 |
| AR_DBD | | 3.652e-96 |
| MA0343.1 | NDT80 | 4.046e-96 |
| MA0345.1 | NHP6A | 1.02e-95 |
| MA0315.1 | HAP4 | 1.508e-95 |
| MA0057.1 | MZF1_5-13 | 2.334e-95 |
| T_full | | 2.425e-95 |
| SOX8_full_3 | | 3.146e-95 |
| Mlx_DBD | | 3.628e-95 |
| TFAP2B_DBD_2 | | 4.425e-95 |
| MA0038.1 | Gfi1 | 9.156e-95 |
| SRY_DBD_1 | | 1.203e-94 |
| TF7L2_f1 | TF7L2 | 1.225e-94 |
| FOXP3_f1 | FOXP3 | 1.617e-94 |
| ERG_full_1 | | 2.041e-94 |
| YBOX1_f2 | YBOX1 | 2.829e-94 |
| MA0085.1 | Su(H) | 2.936e-94 |
| HSF2_si | HSF2 | 3.751e-94 |
| PTF1A_f1 | PTF1A | 4.672e-94 |
| SOX2_full_1 | | 5.411e-94 |
| SOX14_DBD_1 | | 5.934e-94 |
| HLF_si | HLF | 6.336e-94 |
| MA0523.1 | TCF7L2 | 6.762e-94 |
| MA0335.1 | MET4 | 7.909e-94 |
| POU4F1_DBD | | 1.146e-93 |
| SMAD1_si | SMAD1 | 1.224e-93 |
| IRX5_DBD | | 1.891e-93 |
| FOXO6_DBD_1 | | 1.917e-93 |
| LHX6_full_3 | | 2.6e-93 |
| MA0264.1 | ceh-22 | 2.606e-93 |
| SOX2_DBD_1 | | 2.632e-93 |
| MA0331.1 | MCM1 | 3.213e-93 |
| MA0162.2 | EGR1 | 3.455e-93 |
| MGA_DBD_2 | | 4.451e-93 |
| ZNF232_full | | 6.855e-93 |
| BHLHE41_full | | 6.88e-93 |
| IRF5_full_1 | | 7.214e-93 |
| IRX2_DBD | | 8.045e-93 |
| NR4A2_full_2 | | 9.433e-93 |
| MA0097.1 | bZIP911 | 1.46e-92 |
| SOX8_DBD_4 | | 1.656e-92 |
| Zfp740_DBD | | 1.687e-92 |
| STAT3_si | STAT3 | 2.113e-92 |
| NKX2-3_full | | 4.251e-92 |
| FOXC2_DBD_1 | | 5.814e-92 |
| GFI1B_f1 | GFI1B | 5.99e-92 |
| MA0568.1 | MYC3 | 6.041e-92 |
| Sox10_DBD_2 | | 6.912e-92 |
| SOX7_full_3 | | 7.417e-92 |
| EPAS1_si | EPAS1 | 9.678e-92 |
| SP8_DBD | | 1.066e-91 |
| TCF7_f1 | TCF7 | 1.237e-91 |
| MA0573.1 | ATHB9 | 1.354e-91 |
| MTF1_f1 | MTF1 | 1.67e-91 |
| MA0571.1 | ANT | 3.647e-91 |
| SMAD4_si | SMAD4 | 4.091e-91 |
| MA0581.1 | LEC2 | 4.245e-91 |
| MA0531.1 | CTCF | 5.532e-91 |
| ERG_f1 | ERG | 1.027e-90 |
| ZFHX3_f1 | ZFHX3 | 1.881e-90 |
| POU5F1P1_DBD_2 | | 2.36e-90 |
| SP1_DBD | | 3.593e-90 |
| MA0268.1 | ADR1 | 5.038e-90 |
| Rxra_DBD_2 | | 6.291e-90 |
| KLF13_full | | 7.094e-90 |
| STAT1_f2 | STAT1 | 7.44e-90 |
| ETS2_f1 | ETS2 | 8.633e-90 |
| MSX2_DBD_1 | | 1.369e-89 |
| EGR2_DBD | | 1.42e-89 |
| SPDEF_full_1 | | 1.959e-89 |
| MA0326.1 | MAC1 | 2.176e-89 |
| CREB3L1_full_2 | | 2.333e-89 |
| MA0340.1 | MOT3 | 2.843e-89 |
| MA0463.1 | Bcl6 | 3.247e-89 |
| MA0074.1 | RXRA::VDR | 3.387e-89 |
| MA0222.1 | exd | 4.599e-89 |
| SOX18_full_3 | | 5.065e-89 |
| Sox3_DBD_1 | | 5.664e-89 |
| MA0142.1 | Pou5f1::Sox2 | 6.141e-89 |
| ERG_DBD_1 | | 6.83e-89 |
| HSF1_full | | 8.488e-89 |
| Klf12_DBD | | 1.132e-88 |
| NFKB1_DBD | | 1.305e-88 |
| HOXD11_DBD_1 | | 1.342e-88 |
| ETV3_DBD | | 1.598e-88 |
| ZBTB4_si | ZBTB4 | 1.603e-88 |
| OVOL1_f1 | OVOL1 | 1.722e-88 |
| MA0105.3 | NFKB1 | 2.908e-88 |
| MA0046.1 | HNF1A | 2.952e-88 |
| SNAI2_DBD | | 3.305e-88 |
| P73_si | P73 | 3.565e-88 |
| ARI5B_f1 | ARI5B | 5.317e-88 |
| MA0532.1 | STAT92E | 5.541e-88 |
| MA0599.1 | KLF5 | 6.033e-88 |
| E2F2_DBD_2 | | 6.384e-88 |
| HOXA10_DBD_1 | | 8.658e-88 |
| Meis3_DBD_1 | | 1.099e-87 |
| SOX8_DBD_2 | | 1.297e-87 |
| CEBPG_full | | 1.606e-87 |
| MA0386.1 | TBP | 1.732e-87 |
| ALX3_full_1 | | 1.802e-87 |
| FOXO6_DBD_3 | | 2.031e-87 |
| MA0140.2 | TAL1::GATA1 | 2.356e-87 |
| PITX1_DBD | | 3.197e-87 |
| ZIC3_f1 | ZIC3 | 3.214e-87 |
| E2F2_DBD_3 | | 4.181e-87 |
| TF65_f2 | TF65 | 4.254e-87 |
| SOX18_f1 | SOX18 | 4.273e-87 |
| SOX10_full_3 | | 5.474e-87 |
| HOXC11_DBD_1 | | 6.893e-87 |
| HSF1_DBD | | 8.495e-87 |
| MA0070.1 | PBX1 | 9.55e-87 |
| MA0436.1 | YPR022C | 1.466e-86 |
| CDX2_f1 | CDX2 | 2.216e-86 |
| MA0578.1 | AtSPL8 | 2.413e-86 |
| MA0471.1 | E2F6 | 2.512e-86 |
| MA0589.1 | ZAP1 | 3.138e-86 |
| MA0035.3 | Gata1 | 4.365e-86 |
| EGR4_DBD_1 | | 8.513e-86 |
| POU4F2_DBD | | 8.953e-86 |
| MA0550.1 | BZR1 | 9.79e-86 |
| POU3F4_DBD_2 | | 1.964e-85 |
| TBX21_DBD_1 | | 1.978e-85 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| ARNTL_DBD | | 2.4e-85 |
| TFEB_f1 | TFEB | 2.43e-85 |
| Rhox11_DBD | | 2.472e-85 |
| PBX3_f2 | PBX3 | 2.557e-85 |
| MA0576.1 | AtMYB84 | 3.495e-85 |
| SOX9_full_5 | | 3.601e-85 |
| TFCP2_full_2 | | 4.342e-85 |
| TFAP2C_full_3 | | 4.805e-85 |
| PBX2_f1 | PBX2 | 5.198e-85 |
| NFIL3_si | NFIL3 | 5.37e-85 |
| TBX19_DBD | | 6.224e-85 |
| MA0025.1 | NFIL3 | 7.375e-85 |
| GCM2_DBD | | 8.295e-85 |
| SOX8_full_2 | | 1.029e-84 |
| MA0327.1 | MATA1 | 1.274e-84 |
| GABPA_full | | 2.012e-84 |
| NKX3-1_full | | 2.035e-84 |
| Hoxd9_DBD_3 | | 2.065e-84 |
| FOXI1_full_2 | | 2.414e-84 |
| POU4F3_DBD | | 3.426e-84 |
| MA0413.1 | USV1 | 3.874e-84 |
| MA0387.1 | SPT2 | 4.439e-84 |
| MA0294.1 | EDS1 | 4.613e-84 |
| SOX15_full_3 | | 4.84e-84 |
| FOXO4_DBD_1 | | 5.194e-84 |
| ARNT2_si | ARNT2 | 5.557e-84 |
| SOX9_full_6 | | 6.796e-84 |
| STA5A_do | STA5A | 7.43e-84 |
| MA0530.1 | CNC::maf-S | 1.092e-83 |
| MA0402.1 | SWI5 | 1.147e-83 |
| Sox17_DBD_1 | | 1.189e-83 |
| MA0235.1 | onecut | 1.353e-83 |
| ZIC2_f1 | ZIC2 | 2.233e-83 |
| EGR1_DBD | | 2.711e-83 |
| KLF8_f1 | KLF8 | 3.651e-83 |
| MA0079.3 | SP1 | 3.825e-83 |
| MA0507.1 | POU2F2 | 3.865e-83 |
| MA0015.1 | Cf2_II | 3.889e-83 |
| GCM1_DBD | | 4.302e-83 |
| MA0460.1 | ttk | 4.846e-83 |
| PO5F1_do | PO5F1 | 6.06e-83 |
| GATA3_si | GATA3 | 6.395e-83 |
| MA0482.1 | Gata4 | 7.706e-83 |
| HLF_full | | 7.927e-83 |
| SRF_full | | 9.013e-83 |
| MA0126.1 | ovo | 1.023e-82 |
| MA0239.1 | prd | 1.023e-82 |
| SOX14_DBD_3 | | 1.453e-82 |
| Ar_DBD | | 1.483e-82 |
| SRBP1_f2 | SRBP1 | 1.578e-82 |
| Sox17_DBD_2 | | 2.063e-82 |
| MEOX2_DBD_3 | | 3.487e-82 |
| PAX5_si | PAX5 | 6.672e-82 |
| MA0086.1 | sna | 7.288e-82 |
| Tcfap2a_DBD_2 | | 7.938e-82 |
| MA0504.1 | NR2C2 | 8.002e-82 |
| MA0266.1 | ABF2 | 9.096e-82 |
| NFAC1_do | NFAC1 | 9.342e-82 |
| KLF1_f1 | KLF1 | 1.072e-81 |
| TBX21_full_3 | | 1.097e-81 |
| KLF4_f2 | KLF4 | 1.499e-81 |
| Sox3_DBD_3 | | 1.509e-81 |
| EN1_DBD_2 | | 1.511e-81 |
| MA0357.1 | PHO4 | 1.599e-81 |
| SNAI2_f1 | SNAI2 | 1.662e-81 |
| ZEB1_do | ZEB1 | 1.662e-81 |
| LBX2_DBD_1 | | 2.074e-81 |
| MA0376.1 | RTG3 | 2.076e-81 |
| KLF14_DBD | | 3.427e-81 |
| GLI2_f1 | GLI2 | 4.518e-81 |
| SMAD2_si | SMAD2 | 6.003e-81 |
| EN1_full_2 | | 6.843e-81 |
| Sox1_DBD_1 | | 9.937e-81 |
| MA0262.1 | mab-3 | 1.007e-80 |
| Sox3_DBD_2 | | 1.343e-80 |
| ZN143_si | ZN143 | 1.434e-80 |
| PITX1_full_2 | | 1.62e-80 |
| MA0149.1 | EWSR1-FLI1 | 1.884e-80 |
| TWST1_f1 | TWST1 | 1.885e-80 |
| FOXO3_full_1 | | 1.953e-80 |
| MA0574.1 | AtMYB15 | 3.552e-80 |
| MA0383.1 | SMP1 | 4.165e-80 |
| MA0284.1 | CIN5 | 5.022e-80 |
| ELK1_DBD_2 | | 7.631e-80 |
| MZF1_f1 | MZF1 | 9.245e-80 |
| MA0016.1 | usp | 1.02e-79 |
| MA0267.1 | ACE2 | 1.023e-79 |
| POU3F2_DBD_1 | | 1.095e-79 |
| HES5_DBD_1 | | 1.139e-79 |
| CEBPG_DBD | | 1.188e-79 |
| DUXA_DBD | | 1.237e-79 |
| PAX6_f1 | PAX6 | 1.351e-79 |
| SP3_DBD | | 1.38e-79 |
| MA0039.2 | Klf4 | 1.727e-79 |
| HXB6_f1 | HXB6 | 1.828e-79 |
| SOX21_DBD_4 | | 3.519e-79 |
| TGIF1_f1 | TGIF1 | 4.684e-79 |
| MSX1_DBD_1 | | 4.942e-79 |
| FEV_DBD | | 6.973e-79 |
| EGR3_DBD | | 1.297e-78 |
| HES7_DBD | | 1.441e-78 |
| Pou2f2_DBD_1 | | 2.144e-78 |
| FOXK1_DBD | | 3.011e-78 |
| MYB_f1 | MYB | 3.784e-78 |
| ZIC1_f1 | ZIC1 | 4.036e-78 |
| MA0316.1 | HAP5 | 4.763e-78 |
| MA0088.1 | znf143 | 1.761e-77 |
| FLI1_DBD_1 | | 2.146e-77 |
| STAT6_do | STAT6 | 2.393e-77 |
| HES5_DBD_2 | | 3.189e-77 |
| POU4F2_full | | 3.338e-77 |
| HMGA2_f1 | HMGA2 | 4.154e-77 |
| ZNF306_full | | 4.744e-77 |
| POU1F1_DBD_1 | | 5.602e-77 |
| CREB3L1_DBD_1 | | 6.712e-77 |
| HOXD12_DBD_2 | | 7.891e-77 |
| MA0145.2 | Tcfcp2l1 | 8.925e-77 |
| MA0552.1 | PIL5 | 9.546e-77 |
| MA0597.1 | THAP1 | 1.792e-76 |
| MA0274.1 | ARR1 | 1.866e-76 |
| E2F4_DBD_2 | | 1.875e-76 |
| ONECUT1_DBD | | 2.062e-76 |
| SOX18_full_1 | | 2.285e-76 |
| ELK1_full_1 | | 2.591e-76 |
| NKX22_si | NKX22 | 3.559e-76 |
| PROP1_DBD | | 3.81e-76 |
| ELK4_DBD | | 5.003e-76 |
| Sox1_DBD_3 | | 5.457e-76 |
| SRY_DBD_3 | | 5.645e-76 |
| NFATC1_full_3 | | 5.731e-76 |
| HOXC12_DBD_2 | | 5.982e-76 |
| ZBTB7B_full | | 6.181e-76 |
| MA0572.1 | ATHB1 | 6.589e-76 |
| GLI2_DBD_1 | | 1.26e-75 |
| Hoxa11_DBD_1 | | 1.565e-75 |
| SOX9_full_2 | | 1.84e-75 |
| BHLHB2_DBD | | 1.84e-75 |
| SOX7_full_1 | | 2.032e-75 |
| NFKB1_f1 | NFKB1 | 3.28e-75 |
| KLF16_DBD | | 7.711e-75 |
| EOMES_f1 | EOMES | 8.494e-75 |
| MA0094.2 | Ubx | 8.59e-75 |
| HNF1A_f1 | HNF1A | 1.031e-74 |
| SMAD3_f1 | SMAD3 | 1.098e-74 |
| PBX1_do | PBX1 | 1.242e-74 |
| SOX2_f1 | SOX2 | 1.734e-74 |
| ZN219_f1 | ZN219 | 1.811e-74 |
| TEF_FL | | 2.343e-74 |
| Bhlhb2_DBD_1 | | 2.664e-74 |
| Alx1_DBD_1 | | 2.826e-74 |
| MA0281.1 | CBF1 | 3.082e-74 |
| Elk3_DBD | | 3.154e-74 |
| REL_do | REL | 3.298e-74 |
| MNT_DBD | | 3.425e-74 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| MA0023.1 | dl_2 | 5.693e-74 |
| Msx3_DBD_1 | | 7.157e-74 |
| SOX15_full_2 | | 9.623e-74 |
| ZBTB49_DBD | | 1.505e-73 |
| MA0101.1 | REL | 3.475e-73 |
| ZNF740_DBD | | 5.683e-73 |
| HNF6_f1 | HNF6 | 6.243e-73 |
| MA0116.1 | Zfp423 | 6.881e-73 |
| PO2F2_si | PO2F2 | 7.123e-73 |
| MA0004.1 | Arnt | 8.318e-73 |
| ZNF740_full | | 8.625e-73 |
| MA0355.1 | PHD1 | 1.076e-72 |
| KAISO_f1 | KAISO | 1.106e-72 |
| SP4_full | | 1.368e-72 |
| BCL6_f1 | BCL6 | 1.519e-72 |
| TCF4_full | | 1.662e-72 |
| SPDEF_DBD_2 | | 2.603e-72 |
| SRF_DBD | | 3.07e-72 |
| MA0515.1 | Sox6 | 3.131e-72 |
| RELB_si | RELB | 5.41e-72 |
| PO3F1_f1 | PO3F1 | 7.729e-72 |
| MA0062.2 | GABPA | 8.675e-72 |
| DBP_full | | 1.016e-71 |
| GLI2_DBD_2 | | 1.08e-71 |
| MA0278.1 | BAS1 | 1.103e-71 |
| HOXC11_full_1 | | 1.121e-71 |
| HXC8_f1 | HXC8 | 1.167e-71 |
| BHLHB3_full | | 1.534e-71 |
| MA0183.1 | CG7056 | 1.556e-71 |
| SOX2_DBD_2 | | 1.974e-71 |
| NFATC1_full_2 | | 1.991e-71 |
| BARHL2_full_2 | | 3.048e-71 |
| MA0333.1 | MET31 | 3.277e-71 |
| Barhl1_DBD_2 | | 3.818e-71 |
| Egr3_DBD | | 5.256e-71 |
| MA0003.2 | TFAP2A | 5.322e-71 |
| SOX14_DBD_2 | | 6.501e-71 |
| GATA5_f1 | GATA5 | 7.104e-71 |
| MA0443.1 | btd | 7.335e-71 |
| ONECUT2_DBD | | 8.607e-71 |
| MA0107.1 | RELA | 1.124e-70 |
| BHE41_f1 | BHE41 | 1.628e-70 |
| Barhl1_DBD_3 | | 1.768e-70 |
| SRY_DBD_4 | | 2.034e-70 |
| HOXD12_DBD_3 | | 2.55e-70 |
| EGR4_DBD_2 | | 3.595e-70 |
| MA0528.1 | ZNF263 | 5.632e-70 |
| ELK1_DBD_1 | | 8.592e-70 |
| PHOX2A_DBD | | 1.199e-69 |
| MA0385.1 | SOK2 | 1.513e-69 |
| CREB3L1_DBD_4 | | 2.044e-69 |
| BARX1_DBD_1 | | 2.175e-69 |
| MA0154.2 | EBF1 | 2.227e-69 |
| MA0540.1 | DPY-27 | 2.265e-69 |
| SOX15_f1 | SOX15 | 2.306e-69 |
| RREB1_si | RREB1 | 5.063e-69 |
| PAX9_DBD | | 6.188e-69 |
| ZBTB4!METH_f1 | ZBTB4!METH | 7.9e-69 |
| HEN1_si | HEN1 | 8.51e-69 |
| MA0163.1 | PLAG1 | 8.855e-69 |
| BARHL2_full_3 | | 1.143e-68 |
| ETV4_DBD | | 1.285e-68 |
| Nkx3-1_DBD | | 1.5e-68 |
| MA0197.1 | Oct | 1.967e-68 |
| MAFK_DBD_1 | | 2.769e-68 |
| ELK3_DBD | | 2.867e-68 |
| MA0516.1 | SP2 | 3.934e-68 |
| CDC5L_si | CDC5L | 4.527e-68 |
| MA0564.1 | ABI3 | 7.075e-68 |
| DLX2_f1 | DLX2 | 1.452e-67 |
| ALX1_si | ALX1 | 1.699e-67 |
| MA0188.1 | Dr | 2.024e-67 |
| PLAG1_f1 | PLAG1 | 2.349e-67 |
| MAX_DBD_1 | | 2.486e-67 |
| NKX3-2_DBD | | 2.712e-67 |
| MA0037.2 | GATA3 | 3.632e-67 |
| MA0350.1 | TOD6 | 4.393e-67 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| E2F3_DBD_2 | | 4.411e-67 |
| POU3F3_DBD_1 | | 4.848e-67 |
| MA0381.1 | SKN7 | 5.412e-67 |
| XBP1_DBD_1 | | 7.267e-67 |
| ZEP2_si | ZEP2 | 1.396e-66 |
| SOX18_full_2 | | 1.756e-66 |
| MNX1_DBD | | 2.52e-66 |
| EGR1_f2 | EGR1 | 2.59e-66 |
| BARHL2_DBD_2 | | 2.653e-66 |
| BARHL2_DBD_3 | | 3.738e-66 |
| PROX1_DBD | | 4.823e-66 |
| Uncx_DBD_1 | | 5.082e-66 |
| SOX21_DBD_3 | | 5.272e-66 |
| MYC_f1 | MYC | 5.895e-66 |
| TFCP2_f1 | TFCP2 | 6.454e-66 |
| AP2A_f2 | AP2A | 7.194e-66 |
| POU6F2_DBD_2 | | 9.925e-66 |
| CREB3L1_full_1 | | 1.526e-65 |
| MAFA_f1 | MAFA | 1.561e-65 |
| PAX6_DBD | | 2.053e-65 |
| PHOX2B_DBD | | 2.209e-65 |
| MA0587.1 | TCP16 | 2.218e-65 |
| ETV1_DBD | | 2.895e-65 |
| MA0551.1 | HY5 | 4.781e-65 |
| SOX2_full_2 | | 4.832e-65 |
| UNCX_DBD_1 | | 6.197e-65 |
| Bhlhb2_DBD_2 | | 6.203e-65 |
| POU2F2_DBD_1 | | 6.612e-65 |
| SOX2_full_3 | | 7.238e-65 |
| MA0135.1 | Lhx3 | 8.62e-65 |
| MA0138.2 | REST | 1.145e-64 |
| MA0524.1 | TFAP2C | 1.172e-64 |
| Sox1_DBD_4 | | 1.261e-64 |
| DBP_DBD | | 1.771e-64 |
| MA0485.1 | Hoxc9 | 2.066e-64 |
| WT1_f1 | WT1 | 2.79e-64 |
| LHX2_DBD_2 | | 3.242e-64 |
| XBP1_DBD_2 | | 3.627e-64 |
| MA0373.1 | RPN4 | 4.946e-64 |
| MA0418.1 | YAP6 | 5.29e-64 |
| MA0084.1 | SRY | 5.745e-64 |
| PHOX2B_full | | 7.3e-64 |
| SRY_DBD_2 | | 9.629e-64 |
| Sox1_DBD_2 | | 1.293e-63 |
| EOMES_DBD_2 | | 1.611e-63 |
| BCL6B_DBD | | 1.662e-63 |
| MA0069.1 | Pax6 | 2.267e-63 |
| ZBTB7C_full | | 2.296e-63 |
| PURA_f1 | PURA | 2.549e-63 |
| DLX2_DBD | | 3.785e-63 |
| MA0096.1 | bZIP910 | 6.006e-63 |
| SMAD3_DBD | | 6.487e-63 |
| MA0310.1 | HAC1 | 9.299e-63 |
| DLX3_do | DLX3 | 1.463e-62 |
| GSX2_DBD | | 1.496e-62 |
| MA0153.1 | HNF1B | 1.796e-62 |
| POU2F1_DBD_1 | | 1.817e-62 |
| INSM1_f1 | INSM1 | 2.051e-62 |
| PKNX1_si | PKNX1 | 2.292e-62 |
| GATA6_f2 | GATA6 | 2.694e-62 |
| LHX3_f1 | LHX3 | 4.438e-62 |
| CREB1_f1 | CREB1 | 5.158e-62 |
| TFCP2_full_1 | | 5.377e-62 |
| SPDEF_full_2 | | 6.465e-62 |
| TFAP2B_DBD_1 | | 6.61e-62 |
| MA0276.1 | ASH1 | 6.626e-62 |
| MA0164.1 | Nr2e3 | 1e-61 |
| BSX_DBD | | 1.403e-61 |
| UNCX_DBD_2 | | 1.48e-61 |
| MA0036.2 | GATA2 | 1.629e-61 |
| SOX2_DBD_3 | | 2.61e-61 |
| MA0445.1 | D | 3.735e-61 |
| POU3F1_DBD_1 | | 4.224e-61 |
| MA0583.1 | RAV1 | 4.257e-61 |
| HNF1B_f1 | HNF1B | 4.326e-61 |
| CART1_DBD | | 4.859e-61 |
| GRHL1_full | | 5.129e-61 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| MA0588.1 | TGA1 | 6.644e-61 |
| MA0173.1 | CG11617 | 7.65e-61 |
| MYCN_si | MYCN | 9.311e-61 |
| EMX2_DBD_2 | | 1.384e-60 |
| MA0112.2 | ESR1 | 1.491e-60 |
| HEY2_f1 | HEY2 | 1.872e-60 |
| REST_f1 | REST | 2.224e-60 |
| MA0453.1 | nub | 4.558e-60 |
| EN2_full | | 6.368e-60 |
| PRRX1_full_2 | | 6.711e-60 |
| NKX6-2_DBD | | 8.167e-60 |
| NR0B1_si | NR0B1 | 8.749e-60 |
| Lhx8_DBD_3 | | 9.035e-60 |
| Arx_DBD | | 1.513e-59 |
| GRHL1_DBD_1 | | 1.601e-59 |
| TFAP2A_DBD_5 | | 2.148e-59 |
| PO2F1_f1 | PO2F1 | 2.966e-59 |
| IRF5_full_2 | | 3.802e-59 |
| GATA1_si | GATA1 | 4.887e-59 |
| HNF1A_full | | 5.676e-59 |
| ESX1_DBD | | 6.341e-59 |
| E2F6_f1 | E2F6 | 8.399e-59 |
| MA0008.1 | HAT5 | 8.705e-59 |
| ESR1_do | ESR1 | 8.725e-59 |
| ISX_DBD_2 | | 1.147e-58 |
| PRRX1_full_1 | | 1.147e-58 |
| PRRX2_full | | 1.147e-58 |
| Vsx1_DBD | | 1.147e-58 |
| ISX_DBD_1 | | 1.485e-58 |
| ZN423_f1 | ZN423 | 1.634e-58 |
| MA0180.1 | Vsx2 | 1.669e-58 |
| SP1_f1 | SP1 | 2.704e-58 |
| SPZ1_f1 | SPZ1 | 2.77e-58 |
| MA0014.2 | PAX5 | 2.836e-58 |
| DLX1_DBD | | 2.911e-58 |
| GATA2_si | GATA2 | 2.92e-58 |
| TFAP2C_DBD_1 | | 3.557e-58 |
| EGR4_f1 | EGR4 | 3.661e-58 |
| EMX1_DBD_2 | | 5.973e-58 |
| DLX6_DBD | | 6.809e-58 |
| BARX2_si | BARX2 | 1.199e-57 |
| NRL_DBD | | 1.295e-57 |
| SRBP2_f1 | SRBP2 | 1.325e-57 |
| EN2_DBD | | 1.643e-57 |
| TEF_DBD | | 1.718e-57 |
| POU1F1_DBD_2 | | 2.412e-57 |
| MA0155.1 | INSM1 | 2.636e-57 |
| CREB3_full_2 | | 2.823e-57 |
| DLX3_DBD | | 3.308e-57 |
| DLX4_DBD | | 3.308e-57 |
| GLIS3_DBD | | 5.416e-57 |
| TFAP2A_DBD_1 | | 6.435e-57 |
| Zfp652_DBD | | 1.116e-56 |
| MA0542.1 | ELT-3 | 1.664e-56 |
| MAFK_full_1 | | 2.106e-56 |
| MA0286.1 | CST6 | 2.472e-56 |
| MA0054.1 | myb.Ph3 | 3.522e-56 |
| SOX21_DBD_2 | | 3.921e-56 |
| ATF6A_si | ATF6A | 4.181e-56 |
| MSX1_DBD_2 | | 4.509e-56 |
| MSX2_DBD_2 | | 4.509e-56 |
| ISX_full | | 4.558e-56 |
| LHX9_DBD_1 | | 4.558e-56 |
| RAXL1_DBD | | 4.558e-56 |
| SHOX2_DBD | | 4.558e-56 |
| SHOX_DBD | | 4.558e-56 |
| Shox2_DBD | | 4.558e-56 |
| HMX1_DBD | | 4.835e-56 |
| MA0028.1 | ELK1 | 6.402e-56 |
| MA0594.1 | Hoxa9 | 7.399e-56 |
| MSX1_full | | 7.947e-56 |
| Msx3_DBD_2 | | 7.947e-56 |
| Mafb_DBD_3 | | 9.036e-56 |
| PAX8_f1 | PAX8 | 1.031e-55 |
| MIXL1_full | | 1.335e-55 |
| NKX6-2_full | | 1.52e-55 |
| HMX2_DBD | | 1.533e-55 |
| GNS motifs | Table 7.10 | |
| Motif name | TF name | Adjusted p-value |
| MA0431.1 | YML081W | 1.712e-55 |
| MA0229.1 | inv | 2.599e-55 |
| MA0078.1 | Sox17 | 4.733e-55 |
| Tcfap2a_DBD_1 | | 6.012e-55 |
| NKX6-1_full | | 6.411e-55 |
| Nkx6-1_DBD | | 6.411e-55 |
| Mafb_DBD_1 | | 9.326e-55 |
| PRRX1_DBD | | 1.016e-54 |
| MA0351.1 | DOT6 | 1.736e-54 |
| ZN589_f1 | ZN589 | 1.975e-54 |
| Lhx8_DBD_1 | | 2.107e-54 |
| VSX2_si | VSX2 | 2.294e-54 |
| MTF1_DBD | | 3.675e-54 |
| HEY2_full | | 3.681e-54 |
| MA0139.1 | CTCF | 3.908e-54 |
| ARX_DBD | | 4.609e-54 |
| ISL2_DBD | | 5.26e-54 |
| GATA4_f1 | GATA4 | 7.252e-54 |
| LBX2_DBD_2 | | 9.149e-54 |
| FLI1_f1 | FLI1 | 1.173e-53 |
| EGR1_full | | 1.612e-53 |
| GLIS1_DBD | | 1.629e-53 |
| PAX5_DBD | | 1.901e-53 |
| HOMEZ_DBD | | 2.046e-53 |
| MA0125.1 | Nobox | 2.178e-53 |
| Alx4_DBD | | 2.403e-53 |
| MA0469.1 | E2F3 | 2.532e-53 |
| MA0575.1 | AtMYB77 | 2.589e-53 |
| MEOX2_DBD_2 | | 4.464e-53 |
| HXD4_f1 | HXD4 | 9.578e-53 |
| HNF1B_full_2 | | 1.337e-52 |
| LHX9_DBD_2 | | 1.427e-52 |
| ONECUT1_full | | 2.627e-52 |
| MA0382.1 | SKO1 | 3.308e-52 |
| NOBOX_si | NOBOX | 5.835e-52 |
| HMX3_DBD | | 7.985e-52 |
| SP1_f2 | SP1 | 8.307e-52 |
| CREB3L1_DBD_3 | | 8.832e-52 |
| MA0363.1 | REB1 | 1.041e-51 |
| ZN333_f1 | ZN333 | 1.447e-51 |
| PLAL1_si | PLAL1 | 1.868e-51 |
| PAX2_DBD | | 3.009e-51 |
| MA0421.1 | YDR026C | 3.315e-51 |
| HXB8_do | HXB8 | 3.486e-51 |
| KLF15_f1 | KLF15 | 3.944e-51 |
| Gbx1_DBD | | 5.386e-51 |
| Creb3l2_DBD_2 | | 5.686e-51 |
| PROP1_full | | 5.73e-51 |
| DLX5_FL | | 5.933e-51 |
| SNAI1_f1 | SNAI1 | 5.939e-51 |
| PO4F2_si | PO4F2 | 6.206e-51 |
| PDX1_DBD_1 | | 8.528e-51 |
| HESX1_f1 | HESX1 | 1.206e-50 |
| MA0447.1 | gt | 1.477e-50 |
| CREB3L1_DBD_2 | | 1.571e-50 |
| ERG_DBD_2 | | 1.739e-50 |
| POU5F1P1_DBD_1 | | 2.405e-50 |
| PAX1_DBD | | 2.564e-50 |
| MA0570.1 | ABF1 | 2.746e-50 |
| MA0496.1 | MAFK | 3.002e-50 |
| MA0434.1 | YPR013C | 3.89e-50 |
| POU6F2_DBD_1 | | 4.176e-50 |
| GBX2_full | | 4.49e-50 |
| Gbx2_DBD | | 4.579e-50 |
| E2F7_f1 | E2F7 | 4.791e-50 |
| MA0412.1 | UME6 | 5.26e-50 |
| ETV5_DBD | | 5.505e-50 |
| VSX1_full | | 7.689e-50 |
| Pou2f2_DBD_2 | | 8.156e-50 |
| MA0441.1 | ZMS1 | 1.569e-49 |
| AP2B_f1 | AP2B | 1.591e-49 |
| Dlx2_DBD | | 3.963e-49 |
| MYBL1_DBD_4 | | 4.64e-49 |
| MA0158.1 | HOXA5 | 6.966e-49 |
| MA0415.1 | YAP1 | 7.118e-49 |
| MA0334.1 | MET32 | 7.996e-49 |
| POU3F2_DBD_2 | | 8.076e-49 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| GRHL1_DBD_2 | | 1.004e-48 |
| TFAP2C_full_1 | | 1.139e-48 |
| CXXC1_si | CXXC1 | 2.226e-48 |
| HES1_f1 | HES1 | 2.351e-48 |
| MA0285.1 | CRZ1 | 3.117e-48 |
| HESX1_DBD_2 | | 3.505e-48 |
| ZN350_f1 | ZN350 | 3.792e-48 |
| ALX4_DBD | | 5.58e-48 |
| ITF2_f1 | ITF2 | 6.919e-48 |
| E2F1_DBD_3 | | 7.526e-48 |
| GBX1_DBD | | 8.316e-48 |
| PLAG1_si | PLAG1 | 1.037e-47 |
| POU2F3_DBD_1 | | 1.143e-47 |
| MA0450.1 | hkb | 1.491e-47 |
| Dbp_DBD | | 2.324e-47 |
| EN1_full_1 | | 2.502e-47 |
| COE1_f2 | COE1 | 3.688e-47 |
| LMX1B_DBD | | 4.44e-47 |
| DRGX_DBD | | 5.499e-47 |
| EN1_DBD_1 | | 1.145e-46 |
| PRRX1_f1 | PRRX1 | 1.178e-46 |
| SP3_f1 | SP3 | 1.218e-46 |
| NKX6-1_DBD | | 1.914e-46 |
| MA0166.1 | Antp | 2.129e-46 |
| MA0186.1 | Dfd | 2.129e-46 |
| MA0203.1 | Scr | 2.129e-46 |
| MA0215.1 | btn | 2.129e-46 |
| MA0225.1 | ftz | 2.129e-46 |
| GBX2_DBD_2 | | 2.852e-46 |
| MAF_f1 | MAF | 3.506e-46 |
| MAZ_f1 | MAZ | 5.593e-46 |
| E4F1_f1 | E4F1 | 6.656e-46 |
| MA0129.1 | TGA1A | 7.829e-46 |
| AP2C_f1 | AP2C | 1.011e-45 |
| CUX1_f1 | CUX1 | 1.027e-45 |
| ETS1_full_2 | | 1.269e-45 |
| HNF1B_full_1 | | 1.865e-45 |
| TFDP1_f1 | TFDP1 | 2.71e-45 |
| HESX1_DBD_1 | | 3.738e-45 |
| MA0435.1 | YPR015C | 4.165e-45 |
| MA0425.1 | YGR067C | 4.303e-45 |
| ESX1_full | | 6.734e-45 |
| PAX5_f1 | PAX5 | 1.074e-44 |
| Prrx2_DBD | | 1.302e-44 |
| MA0172.1 | CG11294 | 1.359e-44 |
| MA0178.1 | CG32105 | 1.359e-44 |
| MA0181.1 | Vsx1 | 1.359e-44 |
| MA0191.1 | HGTX | 1.359e-44 |
| MA0194.1 | Lim1 | 1.359e-44 |
| MA0206.1 | abd-A | 1.359e-44 |
| MA0208.1 | al | 1.359e-44 |
| MA0230.1 | lab | 1.359e-44 |
| MA0236.1 | otp | 1.359e-44 |
| MA0240.1 | repo | 1.359e-44 |
| MA0251.1 | unpg | 1.359e-44 |
| MA0257.1 | zen2 | 1.359e-44 |
| MA0448.1 | H2.0 | 1.359e-44 |
| MA0553.1 | SMZ | 2.549e-44 |
| Alx1_DBD_2 | | 2.872e-44 |
| POU3F4_DBD_1 | | 3.821e-44 |
| MEOX2_DBD_1 | | 7.113e-44 |
| XBP1_f1 | XBP1 | 9.052e-44 |
| GLIS3_f1 | GLIS3 | 1.581e-43 |
| HXA5_si | HXA5 | 1.678e-43 |
| GBX2_DBD_1 | | 3.158e-43 |
| ALX3_full_2 | | 9.577e-43 |
| E2F1_f2 | E2F1 | 1.248e-42 |
| ETS1_DBD_2 | | 1.259e-42 |
| MSX2_f1 | MSX2 | 1.565e-42 |
| ZBTB7A_DBD | | 1.582e-42 |
| HEY2_DBD | | 2.312e-42 |
| TBX21_DBD_3 | | 3.235e-42 |
| VSX2_DBD | | 3.755e-42 |
| VAX1_DBD | | 4.233e-42 |
| LHX6_full_2 | | 5.01e-42 |
| MA0124.1 | NKX3-1 | 7.203e-42 |
| MYBL2_DBD_2 | | 7.266e-42 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| GATA5_DBD | | 1.104e-41 |
| SP2_si | SP2 | 1.226e-41 |
| PAX2_f1 | PAX2 | 1.403e-41 |
| ERG_full_2 | | 1.567e-41 |
| SOX9_f1 | SOX9 | 1.581e-41 |
| MA0146.2 | Zfx | 1.621e-41 |
| MA0214.1 | bsh | 1.649e-41 |
| MA0248.1 | tup | 1.649e-41 |
| ALX3_DBD | | 1.855e-41 |
| BARX1_DBD_2 | | 2.192e-41 |
| AP2D_f1 | AP2D | 2.902e-41 |
| GATA4_DBD | | 3.189e-41 |
| SOX10_si | SOX10 | 3.812e-41 |
| VAX2_DBD | | 6.142e-41 |
| MA0549.1 | BES1 | 7.99e-41 |
| TFAP2C_full_2 | | 1.152e-40 |
| MAFF_DBD | | 1.63e-40 |
| MA0495.1 | MAFF | 1.99e-40 |
| E2F4_do | E2F4 | 2.037e-40 |
| Dlx1_DBD | | 3.156e-40 |
| MAFG_full | | 3.618e-40 |
| EBF1_full | | 4.065e-40 |
| MA0318.1 | HMRA2 | 5.693e-40 |
| CTCF_f2 | CTCF | 6.173e-40 |
| VENTX_DBD_2 | | 1.043e-39 |
| BARHL2_full_1 | | 1.817e-39 |
| GATA3_full | | 2.34e-39 |
| MA0338.1 | MIG2 | 2.668e-39 |
| SOX13_f1 | SOX13 | 2.949e-39 |
| LMX1B_full | | 7.446e-39 |
| MA0437.1 | YPR196W | 7.47e-39 |
| HXB7_si | HXB7 | 8.276e-39 |
| RAX_DBD | | 1.08e-38 |
| ZN148_si | ZN148 | 1.141e-38 |
| MA0349.1 | OPI1 | 1.406e-38 |
| PRRX2_f1 | PRRX2 | 1.661e-38 |
| MA0168.1 | B-H1 | 1.661e-38 |
| MA0169.1 | B-H2 | 1.661e-38 |
| MA0170.1 | C15 | 1.661e-38 |
| MA0171.1 | CG11085 | 1.661e-38 |
| MA0175.1 | CG13424 | 1.661e-38 |
| MA0176.1 | CG15696 | 1.661e-38 |
| MA0179.1 | CG32532 | 1.661e-38 |
| MA0192.1 | Hmx | 1.661e-38 |
| MA0196.1 | NK7.1 | 1.661e-38 |
| MA0226.1 | hbn | 1.661e-38 |
| MA0245.1 | slou | 1.661e-38 |
| MA0250.1 | unc-4 | 1.661e-38 |
| MA0444.1 | CG34031 | 1.661e-38 |
| MA0224.1 | exex | 2.234e-38 |
| MA0339.1 | MIG3 | 2.762e-38 |
| HEY1_DBD | | 2.947e-38 |
| PDX1_DBD_2 | | 4.963e-38 |
| MA0362.1 | RDS2 | 7.49e-38 |
| MAFK_DBD_2 | | 7.928e-38 |
| MAFK_full_2 | | 8.232e-38 |
| HOXB3_DBD | | 1.024e-37 |
| ZFX_f1 | ZFX | 1.283e-37 |
| MAFB_f1 | MAFB | 1.439e-37 |
| GATA3_DBD | | 3.457e-37 |
| LHX2_DBD_1 | | 6.46e-37 |
| MA0024.2 | E2F1 | 7.195e-37 |
| ZBT7B_si | ZBT7B | 7.456e-37 |
| MA0384.1 | SNT2 | 9.682e-37 |
| FLI1_full_2 | | 1.002e-36 |
| CUX2_DBD_1 | | 1.059e-36 |
| BARHL2_DBD_1 | | 1.683e-36 |
| MA0275.1 | ASG1 | 1.951e-36 |
| MA0118.1 | Macho-1 | 2.399e-36 |
| CENPB_full | | 2.654e-36 |
| Hlf_DBD | | 2.878e-36 |
| HOXB5_DBD | | 7.431e-36 |
| SP4_f1 | SP4 | 1.106e-35 |
| POU6F2_full | | 1.442e-35 |
| MA0433.1 | YOX1 | 1.598e-35 |
| GSX1_DBD | | 1.604e-35 |
| MEOX1_full | | 3.216e-35 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| MA0243.1 | sd | 3.708e-35 |
| CUX1_DBD_2 | | 1.506e-34 |
| ARNT_f1 | ARNT | 1.772e-34 |
| VENTX_DBD_1 | | 2.16e-34 |
| PAX2_si | PAX2 | 3.18e-34 |
| MYBL1_DBD_1 | | 8.257e-34 |
| MA0579.1 | CDC5 | 8.494e-34 |
| LMX1A_DBD | | 1.986e-33 |
| Lhx4_DBD | | 1.986e-33 |
| AHR_si | AHR | 2.199e-33 |
| MA0187.1 | Dll | 2.834e-33 |
| EVX1_DBD | | 3.229e-33 |
| E2F1_DBD_1 | | 4.346e-33 |
| MA0428.1 | YKL222C | 6.653e-33 |
| MA0211.1 | bap | 6.987e-33 |
| TFAP2B_DBD_3 | | 7.848e-33 |
| FLI1_DBD_2 | | 9.755e-33 |
| MA0279.1 | CAD1 | 1.17e-32 |
| MAFG_si | MAFG | 1.517e-32 |
| NF2L1_f1 | NF2L1 | 1.517e-32 |
| ELK1_full_2 | | 1.804e-32 |
| Barhl1_DBD_1 | | 4.16e-32 |
| MA0438.1 | YRM1 | 1.526e-31 |
| HOXA2_DBD | | 1.578e-31 |
| Uncx_DBD_2 | | 1.919e-31 |
| HSFY2_DBD_1 | | 1.994e-31 |
| MA0131.1 | HINFP | 2.706e-31 |
| MA0312.1 | HAP1 | 7.097e-31 |
| RUNX3_si | RUNX3 | 7.137e-31 |
| NOTO_DBD | | 7.943e-31 |
| E2F1_DBD_2 | | 1.483e-30 |
| MA0167.1 | Awh | 1.665e-30 |
| MA0177.1 | CG18599 | 1.665e-30 |
| MA0184.1 | CG9876 | 1.665e-30 |
| MA0195.1 | Lim3 | 1.665e-30 |
| MA0198.1 | OdsH | 1.665e-30 |
| MA0200.1 | Pph13 | 1.665e-30 |
| MA0202.1 | Rx | 1.665e-30 |
| MA0209.1 | ap | 1.665e-30 |
| MA0228.1 | ind | 1.665e-30 |
| MA0241.1 | ro | 1.665e-30 |
| MA0457.1 | PHDP | 1.665e-30 |
| PDX1_do | PDX1 | 2.136e-30 |
| RHOXF1_DBD_1 | | 2.268e-30 |
| VSX1_DBD | | 2.517e-30 |
| HINFP_f1 | HINFP | 2.996e-30 |
| KLF6_si | KLF6 | 4.24e-30 |
| MA0506.1 | NRF1 | 5.434e-30 |
| MA0259.1 | HIF1A::ARNT | 5.688e-30 |
| MA0077.1 | SOX9 | 5.699e-30 |
| TFAP2C_DBD_3 | | 7.802e-30 |
| MA0185.1 | Deaf1 | 1.083e-29 |
| MA0034.1 | Gamyb | 1.981e-29 |
| MA0400.1 | SUT2 | 2.134e-29 |
| Hoxd3_DBD | | 3.176e-29 |
| Lhx8_DBD_2 | | 4.067e-29 |
| CTCF_full | | 6.086e-29 |
| EMX2_DBD_1 | | 7.372e-29 |
| MA0117.1 | Mafb | 8.416e-29 |
| MA0565.1 | FUS3 | 9.924e-29 |
| Meox2_DBD | | 1.297e-28 |
| Tcfap2a_DBD_3 | | 1.371e-28 |
| RFX2_f1 | RFX2 | 2.196e-28 |
| MA0067.1 | Pax2 | 2.571e-28 |
| MA0309.1 | GZF3 | 5.032e-28 |
| OTX1_f1 | OTX1 | 6.65e-28 |
| HOXB2_DBD | | 1.395e-27 |
| En2_DBD | | 1.747e-27 |
| HMBOX1_DBD | | 2.977e-27 |
| MA0510.1 | RFX5 | 3.331e-27 |
| HOXA1_DBD | | 6.701e-27 |
| MA0219.1 | ems | 8.383e-27 |
| MA0221.1 | eve | 8.383e-27 |
| MA0238.1 | pb | 8.383e-27 |
| MA0256.1 | zen | 8.383e-27 |
| EVX2_DBD | | 1.752e-26 |
| PAX7_DBD | | 2.409e-26 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| MA0337.1 | MIG1 | 2.835e-26 |
| HSFY2_DBD_3 | | 3.788e-26 |
| CUX1_DBD_1 | | 4.87e-26 |
| MA0590.1 | LFY | 7.368e-26 |
| MA0449.1 | h | 8.068e-26 |
| SOX9_DBD | | 8.108e-26 |
| RHOXF1_full_1 | | 1.409e-25 |
| EMX1_DBD_1 | | 3.55e-25 |
| ZBT7A_f1 | ZBT7A | 9.774e-25 |
| Mafb_DBD_2 | | 1.162e-24 |
| TFAP2A_DBD_6 | | 2.267e-24 |
| MA0367.1 | RGT1 | 5.678e-24 |
| MA0467.1 | Crx | 7.813e-24 |
| PITX3_DBD | | 1.694e-23 |
| MA0332.1 | MET28 | 1.823e-23 |
| MA0237.2 | pan | 2.261e-23 |
| MA0544.1 | GEI-11 | 2.369e-23 |
| MA0439.1 | YRR1 | 2.423e-23 |
| MA0300.1 | GAT1 | 3.91e-23 |
| CUX2_DBD_2 | | 6.866e-23 |
| DPRX_DBD_2 | | 8.214e-23 |
| HSFY2_DBD_2 | | 1.685e-22 |
| Hoxa2_DBD | | 7.074e-22 |
| MA0414.1 | XBP1 | 7.903e-22 |
| MYBL2_DBD_4 | | 9.847e-22 |
| MA0213.1 | brk | 1.176e-21 |
| RFX2_DBD_2 | | 1.264e-21 |
| SOX5_f1 | SOX5 | 2.292e-21 |
| RFX1_f1 | RFX1 | 3.042e-21 |
| MA0416.1 | YAP3 | 4.377e-21 |
| MA0273.1 | ARO80 | 4.462e-21 |
| TFAP2A_DBD_3 | | 5.5e-21 |
| PITX1_full_1 | | 6.281e-21 |
| MA0577.1 | AtSPL3 | 8.032e-21 |
| HIF1A_si | HIF1A | 9.705e-21 |
| MA0364.1 | REI1 | 2.226e-20 |
| MA0430.1 | YLR278C | 2.845e-20 |
| MA0411.1 | UPC2 | 3.054e-20 |
| CUX1_DBD_3 | | 4.016e-20 |
| MA0509.1 | Rfx1 | 4.197e-20 |
| MA0586.1 | SPL14 | 5.17e-20 |
| MA0529.1 | BEAF-32 | 5.41e-20 |
| MA0308.1 | GSM1 | 7.458e-20 |
| PAX4_DBD | | 1.311e-19 |
| MYBL2_DBD_1 | | 2.237e-19 |
| ZBED1_DBD | | 3.462e-19 |
| ISL1_f1 | ISL1 | 4.921e-19 |
| PAX4_full | | 5.411e-19 |
| NRF1_full | | 5.884e-19 |
| MA0123.1 | abi4 | 2.383e-18 |
| GMEB2_DBD_2 | | 2.743e-18 |
| MA0391.1 | STB4 | 3.424e-18 |
| MA0422.1 | YDR520C | 4.759e-18 |
| NRF1_f1 | NRF1 | 4.894e-18 |
| MA0282.1 | CEP3 | 7.731e-18 |
| MA0348.1 | OAF1 | 1.325e-17 |
| GLIS2_DBD | | 3.077e-17 |
| MA0365.1 | RFX1 | 3.429e-17 |
| PAX7_full | | 5.606e-17 |
| RFX3_f1 | RFX3 | 9.902e-17 |
| MYBL2_DBD_3 | | 1.318e-16 |
| GMEB2_DBD_1 | | 2.308e-16 |
| RUNX3_DBD_3 | | 5.64e-16 |
| MA0265.1 | ABF1 | 1.227e-15 |
| MA0292.1 | ECM22 | 1.386e-15 |
| PAX3_DBD | | 1.402e-15 |
| NF2L2_si | NF2L2 | 2.234e-15 |
| MA0358.1 | PUT3 | 2.541e-15 |
| MA0290.1 | DAL81 | 7.778e-15 |
| MA0432.1 | YNR063W | 4.731e-14 |
| MA0295.1 | FHL1 | 9.102e-14 |
| RFX3_DBD_2 | | 1.759e-13 |
| MA0354.1 | PDR8 | 2.108e-13 |
| OTX2_DBD_1 | | 2.887e-13 |
| OTX2_DBD_2 | | 4.364e-13 |
| Otx1_DBD_2 | | 4.364e-13 |
| LHX6_full_1 | | 5.917e-13 |

| GNS motifs | Table 7.10 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| MA0567.1 | ERF1 | 1.387e-12 |
| RFX4_DBD_2 | | 1.506e-12 |
| CRX_si | CRX | 1.768e-12 |
| HINFP1_full_1 | | 1.9e-12 |
| Otx1_DBD_1 | | 1.958e-12 |
| FOS_si | FOS | 2.813e-12 |
| OTX1_DBD_1 | | 4.257e-12 |
| MA0420.1 | YBR239C | 5.042e-12 |
| MYBL1_DBD_2 | | 5.349e-12 |
| MYBL1_DBD_3 | | 6.196e-12 |
| MECP2_f1 | MECP2 | 8.995e-12 |
| TEAD1_f1 | TEAD1 | 1.865e-11 |
| MA0557.1 | FHY3 | 2.208e-11 |
| MA0424.1 | YER184C | 5.545e-11 |
| MA0600.1 | RFX2 | 9.527e-11 |
| GMEB2_DBD_4 | | 2.133e-10 |
| MA0392.1 | STB5 | 2.577e-10 |
| MA0283.1 | CHA4 | 3.888e-10 |
| RFX4_DBD_1 | | 1.004e-09 |
| MA0299.1 | GAL4 | 1.68e-09 |
| Rfx2_DBD_2 | | 1.986e-09 |
| DPRX_DBD_1 | | 1.998e-09 |
| E2F2_f1 | E2F2 | 2.496e-09 |
| GSC2_DBD | | 2.827e-09 |
| HINFP1_full_2 | | 4.663e-09 |
| TEAD1_full_2 | | 7.406e-09 |
| MA0271.1 | ARG80 | 1.846e-08 |
| OTX1_DBD_2 | | 2.68e-08 |
| RUNX2_DBD_2 | | 2.696e-08 |
| DMBX1_DBD | | 4.441e-08 |
| MA0325.1 | LYS14 | 1.99e-07 |
| TEAD3_DBD_1 | | 2.875e-07 |
| MA0395.1 | STP2 | 8.448e-07 |
| E2F5_do | E2F5 | 4.558e-06 |
| GMEB2_DBD_3 | | 4.7e-06 |
| MA0289.1 | DAL80 | 1.155e-05 |
| RUNX2_f1 | RUNX2 | 1.735e-05 |
| MA0380.1 | SIP4 | 2.171e-05 |
| MA0201.1 | Ptx1 | 7.578e-05 |
| GSC_full | | 7.687e-05 |
| MA0511.1 | RUNX2 | 0.0002088 |
| MA0456.1 | opa | 0.0003549 |
| MA0397.1 | STP4 | 0.0008826 |
| MA0344.1 | NHP10 | 0.0009088 |
| MA0527.1 | ZBTB33 | 0.001335 |
| MA0541.1 | EFL-1 | 0.001693 |
| MBD2_si | MBD2 | 0.001863 |
| MA0405.1 | TEA1 | 0.001947 |
| MA0396.1 | STP3 | 0.003194 |
| MA0280.1 | CAT8 | 0.003372 |
| ZIC3_full | | 0.005761 |
| MA0410.1 | UGA3 | 0.009072 |
| PEBB_f1 | PEBB | 0.0208 |
| MA0360.1 | RDR1 | 0.03791 |
| MA0470.1 | E2F4 | 0.04407 |

Table 7.10 Table of motifs enriched in GNS differentially accessible loci compared to background loci through the MEME suite tool FIMO [83].

| NS motifs | Table 7.11 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| MA0490.1 | JUNB | 1.263e-193 |
| MA0477.1 | FOSL1 | 1.295e-186 |
| MA0491.1 | JUND | 1.096e-185 |
| MA0476.1 | FOS | 5.823e-173 |
| MA0489.1 | JUN | 4.586e-139 |
| JUN_f1 | JUN | 5.438e-120 |
| MA0099.2 | JUN::FOS | 8.183e-120 |
| JDP2_full_1 | | 4.604e-112 |
| JDP2_DBD_1 | | 4.127e-95 |
| FOSL2_f1 | FOSL2 | 4.171e-82 |
| JUND_f1 | JUND | 9.569e-76 |
| FOSB_f1 | FOSB | 1.889e-71 |
| MA0303.1 | GCN4 | 2.605e-70 |
| Jdp2_DBD_1 | | 4.172e-69 |
| BATF_si | BATF | 3.661e-64 |
| MA0462.1 | BATF::JUN | 7.531e-46 |
| NFE2_DBD | | 2.141e-42 |
| JUNB_f1 | JUNB | 5.226e-31 |
| MA0478.1 | FOSL2 | 2.207e-28 |
| TEAD3_DBD_2 | | 6.669e-23 |
| RUNX2_DBD_3 | | 1.474e-22 |
| TEAD4_DBD | | 1.235e-21 |
| MA0272.1 | ARG81 | 2.791e-20 |
| RUNX3_DBD_2 | | 8.094e-19 |
| MA0406.1 | TEC1 | 1.251e-10 |
| RUNX3_full | | 6.227e-09 |
| TEAD1_full_1 | | 7.023e-08 |
| FOSL1_f2 | FOSL1 | 1.206e-05 |
| RUNX1_f1 | RUNX1 | 0.0006121 |
| MA0242.1 | run::Bgb | 0.04261 |

Table 7.11 Table of motifs enriched in NS differentially accessible loci compared to background loci through the MEME suite tool FIMO [83].

| Proneural motifs | Table 7.12 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| E2F2_DBD_3 | | 6.319E-13 |
| ALX3_full_1 | | 2.229E-11 |
| MEIS1_DBD | | 6.212E-11 |
| Meis2_DBD_1 | | 6.212E-11 |
| Meis3_DBD_1 | | 6.427E-11 |
| Alx1_DBD_1 | | 8.778E-11 |
| CDX1_f1 | CDX1 | 9.275E-10 |
| MIXL1_full | | 9.635E-10 |
| UNCX_DBD_2 | | 8.938E-09 |
| Tcf21_DBD | | 9.976E-09 |
| MA0229.1 | inv | 1.534E-08 |
| MA0386.1 | TBP | 2.672E-08 |
| ISX_full | | 3.982E-08 |
| LHX9_DBD_1 | | 3.982E-08 |
| RAXL1_DBD | | 3.982E-08 |
| SHOX2_DBD | | 3.982E-08 |
| SHOX_DBD | | 3.982E-08 |
| Shox2_DBD | | 3.982E-08 |
| Prrx2_DBD | | 4.627E-08 |
| MA0180.1 | Vsx2 | 9.477E-08 |
| MA0033.1 | FOXL1 | 4.069E-07 |
| Foxj3_DBD_3 | | 6.651E-07 |
| VSX2_DBD | | 4.907E-06 |
| ALX3_DBD | | 8.008E-06 |
| VSX1_DBD | | 8.507E-06 |
| LEF1_f1 | LEF1 | 9.158E-06 |
| MA0167.1 | Awh | 1.196E-05 |
| MA0177.1 | CG18599 | 1.196E-05 |
| MA0184.1 | CG9876 | 1.196E-05 |
| MA0195.1 | Lim3 | 1.196E-05 |
| MA0198.1 | OdsH | 1.196E-05 |
| MA0200.1 | Pph13 | 1.196E-05 |
| MA0202.1 | Rx | 1.196E-05 |
| MA0209.1 | ap | 1.196E-05 |
| MA0228.1 | ind | 1.196E-05 |
| MA0241.1 | ro | 1.196E-05 |
| MA0457.1 | PHDP | 1.196E-05 |
| LBX2_DBD_2 | | 1.554E-05 |
| HESX1_DBD_1 | | 1.569E-05 |
| MA0474.1 | Erg | 2.090E-05 |
| Pou2f2_DBD_1 | | 2.124E-05 |
| LMX1B_DBD | | 2.297E-05 |
| FOXO1_DBD_1 | | 2.525E-05 |
| MA0183.1 | CG7056 | 2.738E-05 |
| CTCF_full | | 2.994E-05 |
| HBP1_f1 | HBP1 | 3.210E-05 |
| POU2F1_DBD_2 | | 3.268E-05 |
| Uncx_DBD_2 | | 3.502E-05 |
| POU3F4_DBD_2 | | 3.515E-05 |
| MA0172.1 | CG11294 | 3.691E-05 |
| MA0178.1 | CG32105 | 3.691E-05 |
| MA0181.1 | Vsx1 | 3.691E-05 |
| MA0191.1 | HGTX | 3.691E-05 |
| MA0194.1 | Lim1 | 3.691E-05 |
| MA0206.1 | abd-A | 3.691E-05 |
| MA0208.1 | al | 3.691E-05 |
| MA0230.1 | lab | 3.691E-05 |
| MA0236.1 | otp | 3.691E-05 |
| MA0240.1 | repo | 3.691E-05 |
| MA0251.1 | unpg | 3.691E-05 |
| MA0257.1 | zen2 | 3.691E-05 |
| MA0448.1 | H2.0 | 3.691E-05 |
| PRRX1_DBD | | 3.982E-05 |
| POU3F2_DBD_1 | | 4.351E-05 |
| MA0157.1 | FOXO3 | 4.723E-05 |
| MA0297.1 | FKH2 | 5.095E-05 |
| FOXD2_DBD_2 | | 5.095E-05 |
| FOXD3_DBD_2 | | 5.095E-05 |
| FOXL1_full_1 | | 5.095E-05 |
| FOXP3_DBD | | 5.095E-05 |
| Foxg1_DBD_3 | | 5.095E-05 |
| Foxk1_DBD_2 | | 5.095E-05 |
| ESX1_DBD | | 8.888E-05 |
| SOX17_f2 | SOX17 | 9.536E-05 |
| MA0100.2 | Myb | 9.826E-05 |
| MA0433.1 | YOX1 | 1.008E-04 |
| ISX_DBD_2 | | 1.247E-04 |
| PRRX1_full_1 | | 1.247E-04 |
| PRRX2_full | | 1.247E-04 |
| Vsx1_DBD | | 1.247E-04 |
| POU6F2_DBD_2 | | 1.450E-04 |
| MA0317.1 | HCM1 | 2.093E-04 |
| SOX10_full_1 | | 2.203E-04 |
| E2F3_DBD_2 | | 2.215E-04 |
| LHX2_DBD_2 | | 2.351E-04 |
| MNX1_DBD | | 3.082E-04 |
| ETV2_DBD | | 3.702E-04 |
| LMX1A_DBD | | 3.735E-04 |
| Lhx4_DBD | | 3.735E-04 |
| E2F1_DBD_4 | | 4.309E-04 |
| SOX8_DBD_3 | | 5.207E-04 |
| MA0296.1 | FKH1 | 5.543E-04 |
| HXA5_si | HXA5 | 6.229E-04 |
| POU5F1P1_DBD_2 | | 7.988E-04 |
| MA0377.1 | SFL1 | 1.010E-03 |
| MA0408.1 | TOS8 | 1.180E-03 |
| TBP_f1 | TBP | 1.266E-03 |
| MA0294.1 | EDS1 | 1.277E-03 |
| POU2F2_DBD_2 | | 1.317E-03 |
| FOXB1_full | | 1.624E-03 |
| SOX9_full_4 | | 2.012E-03 |
| FOXQ1_f1 | FOXQ1 | 2.209E-03 |
| EN1_DBD_1 | | 2.303E-03 |
| Foxc1_DBD_2 | | 2.517E-03 |
| PROP1_DBD | | 2.593E-03 |
| POU3F3_DBD_3 | | 2.647E-03 |
| PAX4_full | | 3.125E-03 |
| ESX1_full | | 3.509E-03 |
| MA0520.1 | Stat6 | 3.742E-03 |
| ETS1_DBD_1 | | 4.137E-03 |
| MA0593.1 | FOXP2 | 4.582E-03 |
| EN2_DBD | | 4.641E-03 |
| POU3F1_DBD_2 | | 4.950E-03 |
| POU4F1_DBD | | 5.332E-03 |
| FOXJ3_DBD_1 | | 5.849E-03 |
| POU2F3_DBD_2 | | 5.861E-03 |
| POU3F3_DBD_1 | | 6.086E-03 |
| SOX8_full_1 | | 6.302E-03 |
| Gbx1_DBD | | 6.307E-03 |
| EMX2_DBD_1 | | 6.927E-03 |
| FOXP2_si | FOXP2 | 8.672E-03 |
| Ascl2_DBD | | 9.495E-03 |
| HOXA1_DBD | | 1.140E-02 |
| ISX_DBD_1 | | 1.164E-02 |
| POU3F3_DBD_2 | | 1.241E-02 |
| LHX6_full_1 | | 1.242E-02 |
| MSX2_DBD_1 | | 1.277E-02 |
| NFATC1_full_3 | | 1.327E-02 |
| MSX1_DBD_2 | | 1.342E-02 |
| MSX2_DBD_2 | | 1.342E-02 |
| FOXJ2_DBD_2 | | 1.347E-02 |
| LMX1B_full | | 1.440E-02 |
| MA0125.1 | Nobox | 1.462E-02 |
| ETS1_si | ETS1 | 1.495E-02 |
| RAX_DBD | | 1.612E-02 |
| DLX3_DBD | | 1.651E-02 |
| DLX4_DBD | | 1.651E-02 |
| MA0136.1 | ELF5 | 1.674E-02 |
| Lhx8_DBD_2 | | 2.001E-02 |
| PAX4_DBD | | 2.533E-02 |
| NANOG_f1 | NANOG | 2.808E-02 |
| MA0480.1 | Foxo1 | 2.880E-02 |
| POU2F2_DBD_1 | | 2.908E-02 |
| Arx_DBD | | 3.054E-02 |
| MA0345.1 | NHP6A | 3.187E-02 |
| Sox10_DBD_1 | | 3.398E-02 |
| GCR_si | GCR | 4.091E-02 |
| VSX2_si | VSX2 | 4.145E-02 |
| VAX1_DBD | | 4.687E-02 |

Table 7.12 Table of motifs enriched in proneural differentially accessible loci compared to background loci through the MEME suite tool FIMO [83].

| Mesenchymal motifs | Table 7.13 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| BACH1_si | BACH1 | 0.000E+00 |
| BATF_si | BATF | 0.000E+00 |
| FOSB_f1 | FOSB | 0.000E+00 |
| FOSL1_f2 | FOSL1 | 0.000E+00 |
| FOSL2_f1 | FOSL2 | 0.000E+00 |
| JUNB_f1 | JUNB | 0.000E+00 |
| JUND_f1 | JUND | 0.000E+00 |
| JUN_f1 | JUN | 0.000E+00 |
| MAFK_si | MAFK | 0.000E+00 |
| NFE2_f2 | NFE2 | 0.000E+00 |
| SMRC1_f1 | SMRC1 | 0.000E+00 |
| NF2L2_si | NF2L2 | 1.124E-266 |
| SNAI2_f1 | SNAI2 | 3.458E-247 |
| ZEB1_do | ZEB1 | 3.458E-247 |
| SNAI1_f1 | SNAI1 | 2.008E-199 |
| FOS_si | FOS | 3.925E-170 |
| ITF2_f1 | ITF2 | 5.105E-152 |
| RUNX1_f1 | RUNX1 | 6.050E-89 |
| TBX5_si | TBX5 | 3.806E-88 |
| AP2D_f1 | AP2D | 1.696E-78 |
| RUNX2_f1 | RUNX2 | 9.286E-70 |
| OTX2_si | OTX2 | 4.286E-70 |
| TWST1_f1 | TWST1 | 9.409E-68 |
| PITX2_si | PITX2 | 1.917E-63 |
| RXRB_f1 | RXRB | 1.667E-59 |
| SRBP1_f2 | SRBP1 | 1.019E-57 |
| ZIC3_f1 | ZIC3 | 9.655E-57 |
| THA_f1 | THA | 1.426E-56 |
| TFE3_f1 | TFE3 | 1.804E-56 |
| USF2_f1 | USF2 | 5.776E-55 |
| PAX2_f1 | PAX2 | 4.092E-53 |
| RARA_f1 | RARA | 1.021E-52 |
| MAFG_si | MAFG | 4.942E-52 |
| NF2L1_f1 | NF2L1 | 4.942E-52 |
| THB_f1 | THB | 4.248E-49 |
| AP2A_f2 | AP2A | 5.477E-48 |
| HESX1_f1 | HESX1 | 9.665E-48 |
| THB_do | THB | 1.569E-47 |
| VDR_f1 | VDR | 1.724E-47 |
| ZFX_f1 | ZFX | 5.068E-47 |
| RARB_f1 | RARB | 1.251E-46 |
| ESR1_do | ESR1 | 6.873E-46 |
| PAX5_si | PAX5 | 1.591E-45 |
| COT2_f1 | COT2 | 3.474E-45 |
| THA_f2 | THA | 1.502E-44 |
| NR1I2_f2 | NR1I2 | 8.209E-44 |
| ESR2_do | ESR2 | 1.729E-42 |
| TBX2_f1 | TBX2 | 2.122E-41 |
| AP2B_f1 | AP2B | 1.325E-40 |
| PPARA_f1 | PPARA | 2.929E-40 |
| BMAL1_f1 | BMAL1 | 5.442E-40 |
| PAX2_si | PAX2 | 4.163E-39 |
| ATF1_si | ATF1 | 1.684E-38 |
| FOXC2_f1 | FOXC2 | 2.731E-38 |
| VDR_f2 | VDR | 5.752E-38 |
| GLI1_f1 | GLI1 | 3.176E-37 |
| ESR2_si | ESR2 | 1.011E-36 |
| HEY2_f1 | HEY2 | 1.949E-36 |
| RARG_f1 | RARG | 2.020E-36 |
| PEBB_f1 | PEBB | 2.628E-36 |
| TLX1_f1 | TLX1 | 3.173E-36 |
| TBX3_f1 | TBX3 | 8.071E-36 |
| P53_f2 | P53 | 1.845E-35 |
| ATF6A_si | ATF6A | 2.018E-35 |
| PAX8_f1 | PAX8 | 2.082E-35 |
| ARNT2_si | ARNT2 | 7.592E-35 |
| NFKB1_f1 | NFKB1 | 1.818E-34 |
| KLF1_f1 | KLF1 | 7.955E-34 |
| ZIC2_f1 | ZIC2 | 9.474E-33 |
| RUNX3_si | RUNX3 | 1.143E-32 |
| KLF4_f2 | KLF4 | 1.743E-32 |
| USF1_f1 | USF1 | 6.881E-32 |
| MTF1_f1 | MTF1 | 4.603E-31 |
| TLX1_f2 | TLX1 | 3.762E-29 |
| TEAD1_f1 | TEAD1 | 1.095E-28 |
| ENOA_si | ENOA | 3.971E-28 |
| MITF_f1 | MITF | 1.249E-27 |
| TF65_f2 | TF65 | 1.284E-27 |
| SP4_f1 | SP4 | 4.219E-27 |
| PAX5_f1 | PAX5 | 1.570E-24 |
| NKX21_f1 | NKX21 | 1.695E-24 |
| ZIC1_f1 | ZIC1 | 1.812E-24 |

| Mesenchymall motifs | Table 7.13 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| COT2_f2 | COT2 | 2.889E-24 |
| NR4A3_f1 | NR4A3 | 3.472E-24 |
| SP3_f1 | SP3 | 5.087E-24 |
| ZN219_f1 | ZN219 | 6.771E-24 |
| MYC_f1 | MYC | 2.408E-23 |
| CXXC1_si | CXXC1 | 2.629E-23 |
| BPTF_si | BPTF | 3.028E-23 |
| BHE41_f1 | BHE41 | 3.597E-23 |
| ZN143_si | ZN143 | 5.086E-23 |
| NKX25_f1 | NKX25 | 6.021E-23 |
| SRBP2_f1 | SRBP2 | 1.269E-22 |
| NR1I3_si | NR1I3 | 3.084E-22 |
| RREB1_si | RREB1 | 5.434E-22 |
| RARG_do | RARG | 9.117E-22 |
| ZEP1_f1 | ZEP1 | 1.697E-21 |
| NR1I3_f2 | NR1I3 | 2.032E-21 |
| KLF3_f1 | KLF3 | 3.011E-21 |
| SP1_f2 | SP1 | 3.365E-21 |
| NR2C1_si | NR2C1 | 4.383E-21 |
| AP2C_f1 | AP2C | 6.068E-21 |
| PPARA_f2 | PPARA | 1.723E-20 |
| HES1_f1 | HES1 | 4.325E-20 |
| ATF2+ATF4_f1 | ATF2+ATF4 | 4.793E-20 |
| CREB1_f1 | CREB1 | 5.242E-20 |
| RXRA_f1 | RXRA | 5.356E-20 |
| ATF3_f1 | ATF3 | 1.478E-19 |
| SP1_f1 | SP1 | 1.737E-19 |
| HIC1_si | HIC1 | 3.056E-19 |
| RARA_f2 | RARA | 3.619E-19 |
| ERR3_f1 | ERR3 | 1.164E-18 |
| HINFP_f1 | HINFP | 1.632E-18 |
| MAFA_f1 | MAFA | 2.061E-18 |
| MECP2_f1 | MECP2 | 4.827E-18 |
| HMGA1_f1 | HMGA1 | 7.827E-18 |
| GLI3_si | GLI3 | 1.146E-17 |
| COE1_f2 | COE1 | 1.391E-17 |
| IKZF1_f1 | IKZF1 | 2.447E-17 |
| TGIF1_f1 | TGIF1 | 3.805E-17 |
| KLF15_f1 | KLF15 | 4.378E-17 |
| MBD2_si | MBD2 | 9.341E-17 |
| PLAL1_si | PLAL1 | 2.752E-16 |
| ZBTB6_si | ZBTB6 | 5.002E-16 |
| PURA_f1 | PURA | 1.188E-15 |
| MAFB_f1 | MAFB | 2.457E-15 |
| FUBP1_f1 | FUBP1 | 2.598E-15 |
| ERR2_f1 | ERR2 | 4.164E-15 |
| RXRG_f1 | RXRG | 5.504E-15 |
| COT1_f1 | COT1 | 5.656E-15 |
| DDIT3_f1 | DDIT3 | 5.782E-15 |
| ZBTB4_si | ZBTB4 | 1.067E-14 |
| YBOX1_f2 | YBOX1 | 1.250E-14 |
| MLXPL_f1 | MLXPL | 1.274E-14 |
| ZN423_f1 | ZN423 | 1.350E-14 |
| PPARG_f1 | PPARG | 4.878E-14 |
| XBP1_f1 | XBP1 | 5.873E-14 |
| NRF1_f1 | NRF1 | 9.087E-14 |
| MAX_f1 | MAX | 9.854E-14 |
| NR6A1_do | NR6A1 | 9.878E-14 |
| ZN589_f1 | ZN589 | 1.897E-13 |
| NFIA+NFIB+NFIC_f2 | NFIA+NFIB+NFIC | 3.255E-13 |
| ZBT7B_si | ZBT7B | 4.387E-13 |
| SOX5_f1 | SOX5 | 5.624E-13 |
| BHE40_f2 | BHE40 | 6.793E-13 |
| SP2_si | SP2 | 8.241E-13 |
| REL_do | REL | 8.389E-13 |
| E2F2_f1 | E2F2 | 1.018E-12 |
| GLIS3_f1 | GLIS3 | 2.479E-12 |
| OTX1_f1 | OTX1 | 3.058E-12 |
| CEBPD_f1 | CEBPD | 5.242E-12 |
| COT1_si | COT1 | 6.351E-12 |
| RORA_f1 | RORA | 6.447E-12 |
| P73_si | P73 | 7.118E-12 |
| NR2E3_f1 | NR2E3 | 1.654E-11 |
| E2F6_f1 | E2F6 | 2.562E-11 |
| ERR1_f1 | ERR1 | 2.758E-11 |
| CEBPB_f1 | CEBPB | 4.613E-11 |
| STF1_f1 | STF1 | 5.259E-11 |
| RORG_f1 | RORG | 7.065E-11 |
| MAZ_f1 | MAZ | 7.106E-11 |
| KLF8_f1 | KLF8 | 1.860E-10 |
| WT1_f1 | WT1 | 2.399E-10 |
| REST_f1 | REST | 2.898E-10 |

| Mesenchymal motifs | Table 7.13 | |
|---|---|---|
| Motif name | TF name | Adjusted p-value |
| ZBT7A_f1 | ZBT7A | 3.250E-10 |
| PPARG_si | PPARG | 3.590E-10 |
| CEBPG_si | CEBPG | 4.601E-10 |
| HIF1A_si | HIF1A | 5.847E-10 |
| RELB_si | RELB | 6.911E-10 |
| EGR4_f1 | EGR4 | 7.900E-10 |
| ZEP2_si | ZEP2 | 1.048E-09 |
| TFCP2_f1 | TFCP2 | 1.345E-09 |
| NFIA+NFIB+NFIC_si | NFIA+NFIB+NFIC | 6.887E-09 |
| HXD13_f1 | HXD13 | 7.011E-09 |
| HTF4_f1 | HTF4 | 8.936E-09 |
| AHR_si | AHR | 1.136E-08 |
| FOXJ3_si | FOXJ3 | 2.170E-08 |
| NR1D1_f1 | NR1D1 | 2.911E-08 |
| SPZ1_f1 | SPZ1 | 3.132E-08 |
| TFEB_f1 | TFEB | 4.352E-08 |
| HNF4G_f1 | HNF4G | 5.244E-08 |
| PPARD_f1 | PPARD | 6.021E-08 |
| ZN350_f1 | ZN350 | 7.304E-08 |
| NR5A2_f1 | NR5A2 | 7.616E-08 |
| SOX9_f1 | SOX9 | 8.544E-08 |
| NR2C2_f1 | NR2C2 | 1.048E-07 |
| NR1H4_f1 | NR1H4 | 1.166E-07 |
| HAND1_si | HAND1 | 1.407E-07 |
| NFKB2_f1 | NFKB2 | 1.551E-07 |
| SRY_f1 | SRY | 1.593E-07 |
| MYBB_f1 | MYBB | 2.592E-07 |
| IRF3_f1 | IRF3 | 3.547E-07 |
| CEBPA_do | CEBPA | 5.076E-07 |
| CREM_f1 | CREM | 7.876E-07 |
| MCR_f1 | MCR | 1.420E-06 |
| MAF_f1 | MAF | 1.772E-06 |
| ZBTB4!METH_f1 | ZBTB4!METH | 1.807E-06 |
| HSF2_si | HSF2 | 2.267E-06 |
| PAX6_f1 | PAX6 | 4.204E-06 |
| MYCN_si | MYCN | 5.525E-06 |
| NFYC_f1 | NFYC | 8.642E-06 |
| BRAC_si | BRAC | 1.398E-05 |
| INSM1_f1 | INSM1 | 1.468E-05 |
| Mesenchymal motifs | Table 7.13 | |
| Motif name | TF name | Adjusted p-value |
| TFE2_f2 | TFE2 | 2.353E-05 |
| ZN148_si | ZN148 | 3.732E-05 |
| IRF2_f1 | IRF2 | 3.892E-05 |
| PLAG1_f1 | PLAG1 | 4.216E-05 |
| P63_si | P63 | 6.064E-05 |
| SOX2_f1 | SOX2 | 6.299E-05 |
| NFAT5_f1 | NFAT5 | 8.117E-05 |
| GFI1_f1 | GFI1 | 9.153E-05 |
| E2F3_si | E2F3 | 1.028E-04 |
| NR1I2_si | NR1I2 | 1.034E-04 |
| NR1H2_f1 | NR1H2 | 1.490E-04 |
| MEF2C_f1 | MEF2C | 1.491E-04 |
| ARNT_f1 | ARNT | 1.532E-04 |
| KLF6_si | KLF6 | 1.674E-04 |
| GCM1_f1 | GCM1 | 2.986E-04 |
| GLI2_f1 | GLI2 | 3.075E-04 |
| EPAS1_si | EPAS1 | 4.525E-04 |
| PO6F1_f1 | PO6F1 | 5.016E-04 |
| SOX13_f1 | SOX13 | 7.101E-04 |
| HNF4A_f1 | HNF4A | 8.687E-04 |
| ELK3_f1 | ELK3 | 1.801E-03 |
| HXA7_f1 | HXA7 | 1.979E-03 |
| MEF2D_f1 | MEF2D | 2.473E-03 |
| STAT3_si | STAT3 | 2.961E-03 |
| HEN1_si | HEN1 | 3.474E-03 |
| GFI1B_f1 | GFI1B | 5.281E-03 |
| SMAD2_si | SMAD2 | 5.580E-03 |
| IRF4_si | IRF4 | 6.686E-03 |
| STAT2_f1 | STAT2 | 7.967E-03 |
| TYY1_f2 | TYY1 | 8.177E-03 |
| E2F4_do | E2F4 | 1.171E-02 |
| AIRE_f2 | AIRE | 1.319E-02 |
| RFX2_f1 | RFX2 | 1.383E-02 |
| IRF1_si | IRF1 | 1.832E-02 |
| NR4A1_f1 | NR4A1 | 1.975E-02 |
| HXA1_f1 | HXA1 | 2.462E-02 |
| CRX_si | CRX | 2.846E-02 |
| CEBPE_f1 | CEBPE | 3.089E-02 |

Table 7.13 Table of motifs enriched in mesenchymal differentially accessible loci compared to background loci through the MEME suite tool FIMO [83].

| Motif name | Alt name | ALL | GNS | NS | Pn | Mes | GNS vs. NS |
|---|---|---|---|---|---|---|---|
| ZNF238_full | | 264.24 | 419.88 | 329.12 | 307.24 | 289.69 | 90.76 |
| ZNF238_DBD | | 246.28 | 391.39 | 314.70 | 290.12 | 273.73 | 76.69 |
| MA0528 | ZNF263 | 232.09 | 349.22 | 276.51 | 215.31 | 159.59 | 72.71 |
| MA0091 | TAL1::TCF3 | 370.71 | 591.98 | 521.18 | 470.25 | 406.90 | 70.80 |
| ZN238_f1 | ZN238 | 253.21 | 390.76 | 330.03 | 284.41 | 261.48 | 60.73 |
| PITX2_si | PITX2 | 211.54 | 275.82 | 226.51 | 193.45 | 269.93 | 49.30 |
| Atoh1_DBD | | 197.46 | 313.38 | 274.81 | 279.63 | 228.36 | 38.57 |
| MA0554 | SOC1 | 292.11 | 395.11 | 358.73 | 330.39 | 355.73 | 36.38 |
| NEUROG2_full | | 252.04 | 394.30 | 361.74 | 331.74 | 278.34 | 32.56 |
| MA0466 | CEBPB | 281.19 | 419.99 | 388.74 | 346.69 | 362.81 | 31.25 |
| MA0052 | MEF2A | 207.29 | 306.08 | 275.07 | 276.15 | 258.48 | 31.01 |
| TAL1_f1 | TAL1 | 215.91 | 327.64 | 299.11 | 246.18 | 230.17 | 28.53 |
| MA0497 | MEF2C | 256.43 | 377.07 | 348.54 | 345.59 | 321.97 | 28.53 |
| OLIG1_DBD | | 145.34 | 233.62 | 207.31 | 215.65 | 179.73 | 26.32 |
| MA0149 | EWSR1-FLI1 | 369.74 | 526.33 | 500.54 | 421.29 | 353.70 | 25.79 |
| BHLHA15_DBD | | 155.83 | 243.18 | 218.11 | 233.59 | 185.81 | 25.07 |
| MA0050 | IRF1 | 196.76 | 318.58 | 295.15 | 245.74 | 236.91 | 23.43 |
| Srebf1_DBD | | 262.16 | 352.11 | 328.73 | 261.79 | 284.69 | 23.38 |
| NEUROG2_DBD | | 178.76 | 279.10 | 257.69 | 256.67 | 211.18 | 21.40 |
| MA0095 | YY1 | 226.89 | 319.12 | 298.33 | 253.97 | 226.54 | 20.79 |
| OLIG2_full | | 136.63 | 216.34 | 199.25 | 208.30 | 171.01 | 17.09 |
| MEF2C_f1 | MEF2C | 339.96 | 506.65 | 490.70 | 491.17 | 426.90 | 15.96 |
| MA0543 | EOR-1 | 266.62 | 384.49 | 368.58 | 319.27 | 244.48 | 15.91 |
| MEF2A_DBD | | 362.63 | 539.23 | 524.06 | 527.54 | 455.25 | 15.17 |
| BHLHE22_DBD | | 134.48 | 209.61 | 194.63 | 206.73 | 165.04 | 14.98 |
| MEF2D_DBD | | 331.49 | 493.10 | 478.37 | 473.99 | 413.29 | 14.73 |
| MEF2A_f1 | MEF2A | 336.93 | 505.15 | 490.96 | 496.42 | 424.91 | 14.19 |
| IKZF1_f1 | IKZF1 | 367.25 | 481.65 | 467.92 | 362.21 | 435.18 | 13.73 |
| Rarg_DBD_1 | | 229.46 | 308.30 | 295.63 | 241.66 | 275.24 | 12.67 |
| Tcf21_DBD | | 217.71 | 337.85 | 325.37 | 271.62 | 222.39 | 12.47 |
| MEF2D_f1 | MEF2D | 328.89 | 488.30 | 476.06 | 473.93 | 416.12 | 12.23 |
| NEUROD2_full | | 143.96 | 218.61 | 206.96 | 198.47 | 163.68 | 11.65 |
| TAL1_f2 | TAL1 | 190.89 | 284.44 | 273.42 | 226.52 | 207.41 | 11.02 |
| MEF2B_full | | 292.22 | 432.10 | 421.49 | 406.27 | 365.22 | 10.61 |
| OLIG3_DBD | | 155.61 | 243.18 | 233.04 | 234.78 | 196.31 | 10.13 |
| CEBPG_full | | 99.02 | 152.57 | 144.11 | 121.11 | 126.53 | 8.45 |
| OLIG2_DBD | | 152.35 | 237.34 | 229.82 | 236.32 | 195.37 | 7.52 |
| BHLHE23_DBD | | 121.65 | 184.77 | 177.86 | 179.85 | 147.41 | 6.90 |
| FOXC1_DBD_1 | | 199.26 | 287.96 | 281.43 | 273.00 | 250.31 | 6.53 |
| Nr2f6_DBD_1 | | 253.54 | 329.68 | 323.15 | 251.14 | 308.15 | 6.52 |
| MA0102 | | 220.18 | 341.66 | 335.17 | 299.92 | 285.36 | 6.49 |
| TFE2_f2 | TFE2 | 274.55 | 392.54 | 386.87 | 295.74 | 286.89 | 5.67 |
| IRF4_si | IRF4 | 314.15 | 465.55 | 460.17 | 388.93 | 381.07 | 5.38 |
| IRF8_full | | 196.00 | 303.98 | 299.98 | 229.66 | 231.11 | 4.00 |
| CEBPG_DBD | | 80.42 | 123.57 | 119.77 | 98.84 | 102.07 | 3.80 |
| RARG_DBD_1 | | 211.19 | 274.91 | 271.24 | 214.68 | 251.08 | 3.67 |
| IRF4_full | | 165.45 | 258.31 | 254.91 | 196.84 | 196.00 | 3.40 |
| CEBPD_DBD | | 86.62 | 130.79 | 128.39 | 102.93 | 111.56 | 2.40 |
| Cebpb_DBD | | 80.86 | 123.94 | 121.68 | 102.58 | 104.54 | 2.26 |
| MA0165 | Abd-B | 187.32 | 297.16 | 295.02 | 300.96 | 242.88 | 2.14 |

Table 7.14 Table of motif frequencies for motifs enriched in GNS accessible loci. Values indicate the number of motifs per accessible megabase alongside the relative change in the specified condition (rightmost column).

| Motif name | Alt name | ALL | GNS | NS | Pn | Mes | GNS vs. NS |
|---|---|---|---|---|---|---|---|
| MA0478 | FOSL2 | 345.33 | 331.62 | 1079.95 | 206.70 | 899.99 | -748.33 |
| MA0099 | JUN::FOS | 362.34 | 375.90 | 1018.45 | 264.71 | 858.13 | -642.55 |
| MA0491 | JUND | 281.78 | 263.63 | 900.91 | 167.28 | 752.82 | -637.28 |
| MA0477 | FOSL1 | 285.26 | 261.34 | 898.08 | 167.82 | 745.42 | -636.74 |
| MA0490 | JUNB | 272.62 | 253.03 | 874.60 | 159.78 | 723.85 | -621.57 |
| MA0489 | JUN | 282.18 | 270.53 | 879.44 | 176.36 | 736.66 | -608.90 |
| SMRC1_f1 | SMRC1 | 328.91 | 327.56 | 923.08 | 222.85 | 766.02 | -595.51 |
| MA0476 | FOS | 270.76 | 262.37 | 845.86 | 170.42 | 722.80 | -583.49 |
| MA0303 | GCN4 | 241.67 | 225.73 | 788.85 | 141.09 | 648.73 | -563.12 |
| JUNB_f1 | JUNB | 236.54 | 226.70 | 722.52 | 148.16 | 593.16 | -495.82 |
| FOSL2_f1 | FOSL2 | 219.26 | 203.38 | 687.72 | 129.78 | 567.57 | -484.34 |
| JUND_f1 | JUND | 219.01 | 206.37 | 684.20 | 131.07 | 569.46 | -477.83 |
| FOSL1_f2 | FOSL1 | 280.46 | 284.84 | 751.53 | 196.90 | 609.78 | -466.69 |
| TEAD4_DBD |  | 343.93 | 408.58 | 873.91 | 406.65 | 481.29 | -465.33 |
| JDP2_DBD_1 |  | 206.41 | 201.98 | 661.68 | 134.81 | 540.03 | -459.70 |
| NFE2_DBD |  | 209.38 | 204.05 | 656.02 | 129.50 | 550.19 | -451.96 |
| JDP2_full_1 |  | 199.93 | 195.53 | 645.96 | 127.96 | 529.14 | -450.43 |
| JUN_f1 | JUN | 205.93 | 194.99 | 644.00 | 124.85 | 546.91 | -449.00 |
| Jdp2_DBD_1 |  | 192.52 | 183.20 | 631.19 | 117.72 | 522.34 | -447.99 |
| TEAD1_full_1 |  | 332.40 | 395.82 | 824.34 | 394.84 | 460.62 | -428.52 |
| TEAD3_DBD_2 |  | 291.35 | 337.29 | 765.33 | 336.39 | 406.03 | -428.04 |
| RUNX1_f1 | RUNX1 | 376.39 | 462.34 | 872.47 | 437.52 | 511.97 | -410.13 |
| BATF_si | BATF | 211.95 | 215.59 | 623.18 | 146.24 | 537.69 | -407.59 |
| MA0501 | NFE2::MAF | 375.07 | 458.63 | 845.55 | 370.03 | 664.23 | -386.93 |
| FOSB_f1 | FOSB | 212.66 | 221.88 | 590.69 | 154.03 | 480.27 | -368.82 |
| MA0462 | BATF::JUN | 208.19 | 218.31 | 582.07 | 157.11 | 509.91 | -363.75 |
| MAFK_si | MAFK | 357.09 | 417.72 | 775.57 | 335.41 | 567.23 | -357.85 |
| MA0406 | TEC1 | 313.51 | 386.84 | 742.77 | 377.88 | 398.45 | -355.93 |
| RUNX3_full |  | 234.44 | 280.78 | 574.40 | 257.14 | 321.38 | -293.62 |
| BACH1_si | BACH1 | 316.17 | 391.00 | 670.22 | 315.94 | 483.80 | -279.22 |
| TEAD3_DBD_1 |  | 257.66 | 322.75 | 599.71 | 324.11 | 362.22 | -276.95 |
| RUNX3_DBD_2 |  | 213.44 | 256.47 | 532.24 | 232.11 | 291.43 | -275.77 |
| TEAD1_full_2 |  | 269.23 | 351.59 | 616.65 | 360.86 | 368.33 | -265.05 |
| MA0591 | Bach1::Mafk | 240.12 | 268.16 | 522.18 | 200.20 | 376.92 | -254.02 |
| MAFG_full |  | 279.24 | 348.65 | 597.01 | 316.85 | 372.07 | -248.36 |
| NFE2_f2 | NFE2 | 243.13 | 276.98 | 523.53 | 212.35 | 402.92 | -246.55 |
| RUNX2_DBD_3 |  | 186.32 | 223.60 | 464.74 | 195.71 | 254.22 | -241.14 |
| MA0495 |  | 355.09 | 471.46 | 705.97 | 452.76 | 478.81 | -234.51 |
| MA0242 | run::Bgb | 203.61 | 249.39 | 477.24 | 219.58 | 273.84 | -227.85 |
| MA0150 | Nfe2l2 | 210.80 | 247.10 | 466.44 | 193.67 | 365.43 | -219.34 |
| MA0419 | YAP7 | 204.18 | 259.83 | 477.94 | 223.19 | 392.28 | -218.10 |
| MA0514 | Sox3 | 351.89 | 500.42 | 715.25 | 479.05 | 433.81 | -214.83 |
| SOX9_f1 | SOX9 | 307.90 | 436.17 | 648.09 | 429.64 | 422.02 | -211.92 |
| MA0496 | MAFK | 360.88 | 491.06 | 693.73 | 475.28 | 475.70 | -202.67 |
| TF7L2_f1 | TF7L2 | 375.36 | 521.27 | 716.21 | 500.91 | 438.11 | -194.94 |
| NF2L2_si | NF2L2 | 231.73 | 272.41 | 467.22 | 216.85 | 350.53 | -194.81 |
| SOX2_f1 | SOX2 | 355.32 | 501.98 | 691.82 | 506.03 | 459.82 | -189.83 |
| MA0467 | Crx | 246.03 | 311.02 | 491.65 | 307.74 | 284.52 | -180.63 |
| MAFF_DBD |  | 193.55 | 236.08 | 415.22 | 206.64 | 260.75 | -179.14 |
| STAT6_do | STAT6 | 342.95 | 474.41 | 646.96 | 469.84 | 400.20 | -172.55 |

Table 7.15 Table of motif frequencies for motifs enriched in NS accessible loci. Values indicate the number of motifs per accessible megabase alongside the relative change in the specified condition (rightmost column).

| Motif name | Alt name | ALL | GNS | NS | Pn | Mes | Pn vs. Mes |
|---|---|---|---|---|---|---|---|
| MA0386 | TBP | 342.66 | 490.23 | 543.26 | 624.22 | 424.01 | 200.21 |
| MA0015 | Cf2_II | 368.29 | 518.18 | 570.09 | 650.73 | 458.74 | 191.99 |
| MA0398 | SUM1 | 351.82 | 533.14 | 548.84 | 600.41 | 434.90 | 165.51 |
| FOXB1_DBD_2 |  | 353.97 | 522.32 | 588.86 | 618.56 | 456.15 | 162.41 |
| ARI3A_do | ARI3A | 336.89 | 503.82 | 534.73 | 568.94 | 420.72 | 148.21 |
| MA0346 | NHP6B | 248.19 | 363.58 | 390.83 | 446.22 | 301.14 | 145.09 |
| POU3F3_DBD_3 |  | 305.40 | 469.74 | 518.74 | 528.96 | 388.75 | 140.20 |
| POU2F1_DBD_2 |  | 288.51 | 441.16 | 505.37 | 490.29 | 361.63 | 128.66 |
| MA0390 | STB3 | 347.39 | 525.58 | 589.47 | 559.64 | 436.57 | 123.07 |
| POU2F3_DBD_2 |  | 284.08 | 439.24 | 508.07 | 485.67 | 363.13 | 122.54 |
| MA0345 | NHP6A | 204.81 | 289.80 | 328.20 | 374.46 | 254.88 | 119.57 |
| FOXB1_DBD_3 |  | 318.30 | 480.11 | 500.23 | 522.86 | 403.86 | 119.00 |
| POU2F2_DBD_2 |  | 277.87 | 426.13 | 491.04 | 467.52 | 348.57 | 118.95 |
| POU5F1P1_DBD_2 |  | 280.21 | 433.09 | 505.85 | 476.00 | 357.40 | 118.60 |
| POU3F1_DBD_2 |  | 321.73 | 487.58 | 561.29 | 531.09 | 413.81 | 117.28 |
| POU3F3_DBD_1 |  | 274.56 | 417.84 | 510.82 | 462.12 | 346.16 | 115.95 |
| Pou2f2_DBD_1 |  | 271.08 | 412.22 | 497.49 | 451.56 | 335.97 | 115.59 |
| POU4F1_DBD |  | 264.40 | 403.40 | 466.22 | 453.98 | 339.22 | 114.76 |
| SOX10_si | SOX10 | 318.63 | 478.96 | 590.91 | 495.03 | 381.38 | 113.65 |
| POU3F3_DBD_2 |  | 314.18 | 473.05 | 533.77 | 517.68 | 404.28 | 113.40 |
| MA0593 | FOXP2 | 363.91 | 561.30 | 604.67 | 547.86 | 437.10 | 110.77 |
| POU3F2_DBD_1 |  | 254.70 | 389.71 | 463.35 | 436.30 | 325.81 | 110.48 |
| FOXD2_DBD_1 |  | 335.45 | 501.47 | 567.22 | 551.10 | 442.30 | 108.80 |
| CPEB1_full |  | 371.17 | 561.03 | 598.83 | 572.05 | 463.76 | 108.28 |
| NKX31_si | NKX31 | 270.56 | 399.48 | 423.41 | 452.44 | 344.98 | 107.47 |
| FOXC1_DBD_3 |  | 347.82 | 528.43 | 551.93 | 555.02 | 448.09 | 106.93 |
| POU4F3_DBD |  | 250.63 | 380.24 | 447.19 | 424.83 | 318.52 | 106.32 |
| MA0507 | POU2F2 | 342.53 | 501.83 | 644.30 | 540.42 | 434.37 | 106.05 |
| PROP1_DBD |  | 213.40 | 320.38 | 379.16 | 371.63 | 266.65 | 104.98 |
| POU3F1_DBD_1 |  | 285.75 | 428.78 | 539.78 | 466.42 | 362.81 | 103.61 |
| MA0453 | nub | 266.94 | 401.50 | 515.17 | 437.05 | 333.46 | 103.59 |
| PIT1_f1 | PIT1 | 261.38 | 399.05 | 454.77 | 439.06 | 335.66 | 103.41 |
| PO3F2_si | PO3F2 | 255.80 | 390.21 | 444.88 | 436.71 | 333.70 | 103.00 |
| MA0135 |  | 245.02 | 362.06 | 430.68 | 410.26 | 309.34 | 100.92 |
| FOXC1_DBD_2 |  | 280.05 | 424.82 | 470.44 | 465.13 | 366.45 | 98.69 |
| PO2F1_f1 | PO2F1 | 311.42 | 459.63 | 610.99 | 484.01 | 385.61 | 98.40 |
| MA0013 | br_Z4 | 279.88 | 419.44 | 470.49 | 451.53 | 353.67 | 97.86 |
| MA0296 | FKH1 | 284.57 | 447.90 | 469.92 | 449.21 | 351.78 | 97.42 |
| POU2F1_DBD_1 |  | 260.10 | 391.53 | 488.91 | 420.88 | 325.85 | 95.03 |
| POU3F2_DBD_2 |  | 273.08 | 411.11 | 526.97 | 440.35 | 345.78 | 94.57 |
| PO2F2_si | PO2F2 | 290.89 | 427.93 | 543.92 | 459.13 | 365.85 | 93.28 |
| FOXQ1_f1 | FOXQ1 | 280.06 | 429.15 | 449.41 | 444.81 | 351.99 | 92.82 |
| POU3F4_DBD_2 |  | 227.10 | 345.66 | 415.96 | 375.24 | 282.67 | 92.57 |
| ONEC2_si | ONEC2 | 327.06 | 483.94 | 525.28 | 513.06 | 420.58 | 92.48 |
| FOXB1_full |  | 246.34 | 380.89 | 403.16 | 382.12 | 289.86 | 92.26 |
| MA0388 | SPT23 | 314.97 | 474.47 | 527.45 | 465.92 | 374.05 | 91.86 |
| ARX_DBD |  | 194.90 | 287.25 | 359.26 | 335.26 | 244.31 | 90.95 |
| Foxc1_DBD_2 |  | 254.14 | 401.76 | 409.39 | 408.53 | 317.68 | 90.85 |
| HOXC13_DBD_1 |  | 280.32 | 430.99 | 448.93 | 442.20 | 351.71 | 90.49 |
| Arx_DBD |  | 196.69 | 290.00 | 364.05 | 337.52 | 247.42 | 90.10 |

Table 7.16 Table of motif frequencies for motifs enriched in proneural accessible loci. Values indicate the number of motifs per accessible megabase alongside the relative change in the specified condition (rightmost column).

| Motif name | Alt name | ALL | GNS | NS | Pn | Mes | Pn vs. Mes |
|---|---|---|---|---|---|---|---|
| MA0478 | FOSL2 | 345.33 | 331.62 | 1079.95 | 206.70 | 899.99 | -693.29 |
| MA0099 | JUN::FOS | 362.34 | 375.90 | 1018.45 | 264.71 | 858.13 | -593.42 |
| MA0491 | JUND | 281.78 | 263.63 | 900.91 | 167.28 | 752.82 | -585.54 |
| MA0477 | FOSL1 | 285.26 | 261.34 | 898.08 | 167.82 | 745.42 | -577.60 |
| MA0490 | JUNB | 272.62 | 253.03 | 874.60 | 159.78 | 723.85 | -564.07 |
| MA0489 | JUN | 282.18 | 270.53 | 879.44 | 176.36 | 736.66 | -560.30 |
| MA0476 | FOS | 270.76 | 262.37 | 845.86 | 170.42 | 722.80 | -552.38 |
| SMRC1_f1 | SMRC1 | 328.91 | 327.56 | 923.08 | 222.85 | 766.02 | -543.17 |
| MA0303 | GCN4 | 241.67 | 225.73 | 788.85 | 141.09 | 648.73 | -507.64 |
| JUNB_f1 | JUNB | 236.54 | 226.70 | 722.52 | 148.16 | 593.16 | -445.01 |
| JUND_f1 | JUND | 219.01 | 206.37 | 684.20 | 131.07 | 569.46 | -438.39 |
| FOSL2_f1 | FOSL2 | 219.26 | 203.38 | 687.72 | 129.78 | 567.57 | -437.79 |
| JUN_f1 | JUN | 205.93 | 194.99 | 644.00 | 124.85 | 546.91 | -422.06 |
| NFE2_DBD |  | 209.38 | 204.05 | 656.02 | 129.50 | 550.19 | -420.69 |
| FOSL1_f2 | FOSL1 | 280.46 | 284.84 | 751.53 | 196.90 | 609.78 | -412.87 |
| JDP2_DBD_1 |  | 206.41 | 201.98 | 661.68 | 134.81 | 540.03 | -405.23 |
| Jdp2_DBD_1 |  | 192.52 | 183.20 | 631.19 | 117.72 | 522.34 | -404.62 |
| JDP2_full_1 |  | 199.93 | 195.53 | 645.96 | 127.96 | 529.14 | -401.18 |
| BATF_si | BATF | 211.95 | 215.59 | 623.18 | 146.24 | 537.69 | -391.46 |
| MA0462 | BATF::JUN | 208.19 | 218.31 | 582.07 | 157.11 | 509.91 | -352.80 |
| FOSB_f1 | FOSB | 212.66 | 221.88 | 590.69 | 154.03 | 480.27 | -326.25 |
| MA0501 | NFE2::MAF | 375.07 | 458.63 | 845.55 | 370.03 | 664.23 | -294.20 |
| MAFK_si | MAFK | 357.09 | 417.72 | 775.57 | 335.41 | 567.23 | -231.81 |
| NFE2_f2 | NFE2 | 243.13 | 276.98 | 523.53 | 212.35 | 402.92 | -190.57 |
| MA0591 | Bach1::Mafk | 240.12 | 268.16 | 522.18 | 200.20 | 376.92 | -176.72 |
| MA0150 | Nfe2l2 | 210.80 | 247.10 | 466.44 | 193.67 | 365.43 | -171.77 |
| MA0419 | YAP7 | 204.18 | 259.83 | 477.94 | 223.19 | 392.28 | -169.09 |
| BACH1_si | BACH1 | 316.17 | 391.00 | 670.22 | 315.94 | 483.80 | -167.86 |
| NF2L2_si | NF2L2 | 231.73 | 272.41 | 467.22 | 216.85 | 350.53 | -133.68 |
| MA0272 | ARG81 | 149.67 | 175.57 | 304.56 | 134.78 | 240.33 | -105.55 |
| PITX2_si | PITX2 | 211.54 | 275.82 | 226.51 | 193.45 | 269.93 | -76.48 |
| TEAD4_DBD |  | 343.93 | 408.58 | 873.91 | 406.65 | 481.29 | -74.64 |
| OTX2_si | OTX2 | 379.42 | 497.81 | 531.81 | 401.03 | 475.49 | -74.46 |
| RUNX1_f1 | RUNX1 | 376.39 | 462.34 | 872.47 | 437.52 | 511.97 | -74.44 |
| IKZF1_f1 | IKZF1 | 367.25 | 481.65 | 467.92 | 362.21 | 435.18 | -72.97 |
| TEAD3_DBD_2 |  | 291.35 | 337.29 | 765.33 | 336.39 | 406.03 | -69.64 |
| TEAD1_full_1 |  | 332.40 | 395.82 | 824.34 | 394.84 | 460.62 | -65.78 |
| RUNX3_full |  | 234.44 | 280.78 | 574.40 | 257.14 | 321.38 | -64.24 |
| FOS_si | FOS | 310.39 | 391.12 | 554.54 | 317.07 | 378.80 | -61.73 |
| RUNX3_DBD_2 |  | 213.44 | 256.47 | 532.24 | 232.11 | 291.43 | -59.32 |
| RUNX2_DBD_3 |  | 186.32 | 223.60 | 464.74 | 195.71 | 254.22 | -58.51 |
| Nr2f6_DBD_1 |  | 253.54 | 329.68 | 323.15 | 251.14 | 308.15 | -57.01 |
| MAFG_full |  | 279.24 | 348.65 | 597.01 | 316.85 | 372.07 | -55.21 |
| MA0242 | run::Bgb | 203.61 | 249.39 | 477.24 | 219.58 | 273.84 | -54.26 |
| MAFF_DBD |  | 193.55 | 236.08 | 415.22 | 206.64 | 260.75 | -54.11 |
| MAFG_si | MAFG | 169.07 | 232.00 | 338.22 | 203.97 | 255.65 | -51.68 |
| NF2L1_f1 | NF2L1 | 169.07 | 232.00 | 338.22 | 203.97 | 255.65 | -51.68 |
| RARA_f1 | RARA | 338.15 | 438.75 | 468.18 | 337.08 | 385.54 | -48.46 |
| TEAD3_DBD_1 |  | 257.66 | 322.75 | 599.71 | 324.11 | 362.22 | -38.11 |
| RARG_DBD_1 |  | 211.19 | 274.91 | 271.24 | 214.68 | 251.08 | -36.40 |

Table 7.17 Table of motif frequencies for motifs enriched in mesenchymal accessible loci. Values indicate the number of motifs per accessible megabase alongside the relative change in the specified condition (rightmost column).