GENETIC AND EPIGENETIC CHANGES IN CNS CANCERS

by

Thoraia M Shinawi

A thesis submitted to the University of Birmingham for the degree
of DOCTOR OF PHILOSOPHY

Medical and Molecular Genetics College of Medicine & Dentistry

School of Clinical and Experimental Medicine

The University of Birmingham

July 2015

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

Cancer is a result of multiple genetic and epigenetic alterations that are acquired during tumour formation and progression. The development of cost-effective high-throughput technologies has positively contributed to the field of cancer research to identify cancer-associated genes, in addition to diagnostic and prognostic markers.

Using the most recent Illumina Infinium HumanMethyaltion450K BeadChip, this work has identified 23 genes that were significantly and differentially hypermethylated in short-term survivor (STS) glioblastoma patients and showed an association with poor prognosis. Interestingly, *NR2F2* showed hypermethylation across multiple CpG loci and was significantly correlated with poor survival. In addition, the work has identified a subset of long-term survivors (LTS) that showed high DNA hypermethylation across many CpG loci indicating the presence of CIMP$^+$.

The other methylation study has identified 7 genes (*EOMES*, *WDR69*, *MIR125B1*, *DZIP1, SOX1, PHOX2B* and *PRDM13*) that were differentially hypermethylated in breast cancer brain metastasis (BCBM) tumours compared to non-metastatic breast tumours. These genes also showed hypermethylation in the paired breast tumours and BCBM samples, signifying the importance of the 7 genes in the development of brain metastasis in the early stages. Loss of expression of *DZIP1*, *PRDM13* and *PHOX2B* was significantly associated with poor patient survival, indicating their potential as prognostic markers for BCBM patients.

Whole exome sequencing (WES) analysis of sporadic non-*NF2* vestibular schwannoma was disappointing and was not able to identify somatic mutations in the novel genes involved in schwannoma formation. However, the identification of novel deleterious mutations in the *NF2* gene confirmed the validation of WES to identify cancer-associated variants.

This study has used the most recent and developed high-throughput technologies to identify candidate genes that were genetically or epigenetically altered in a range of CNS tumours.

Acknowledgment

Table of Contents

List of Figures

List of Tables

Chapter One: Introduction

1.1 Cancer

Cancer is defined as the abnormal growth of cells caused by alterations in gene expression leading to an imbalance of cell proliferations and cell apoptosis programmes, which results in the growth of tumour cells that can invade local tissues or spread to distant organs of the body (Ruddon 2007).

There are >100 types of cancer which are named by their site of origin. Most cancer types can be categorised into: (i) carcinoma, which is the most common form of all cancers that originates from the skin or tissues surrounding the internal organs. Examples of this type of cancer include adenocarcinoma and squamous cell carcinomas, (ii) sarcomas, which arise from bone, cartilage, blood vessels and connective tissues such as osteosarcomas and ewing's sarcoma, (iii) leukaemia, which is cancer of the blood forming tissues such as bone marrow that causes the production of abnormal cells in the blood, (iv) lymphoma, cancers of the immune system's cells, and (v) central nervous system cancer, which arises from the brain and spinal cord such as glioblastoma and meningioma. However, not all tumour cells are cancerous; they can be either benign tumours or malignant tumours. Benign tumours can be surgically removed with no relapse or spreading to nearby organs, while malignant tumours are capable of invading local tissues and spread to other organs of the body (metastasis).

1.1.1 Cancer global incidence and mortality

Cancers are among the leading causes of morbidity and mortality worldwide. In 2012, around 14.1 million new cancer cases were diagnosed in the world, there were 8.2 million cancer deaths and around 32 million individuals living with cancer (Ferlay et al. 2014) (Figure 1.1). The 5 most common forms of cancer diagnosed around the globe are lung cancer, prostate cancer, colorectal cancer, stomach cancer and liver cancer in males (Figure 1.2). Among females, breast cancer, colorectal cancer, cervix/uterus cancer, lung cancer and stomach cancer are the most common sites (Ferlay et al. 2014) (Figure 1.3). It is expected that the number of annual new cases will increase by 70% in the next two decades (World Cancer Report 2014). The World Health Organization (WHO) has recently listed the five leading behavioural risks for cancer: high BMI, unhealthy diet, absence of physical activity, smoking tobacco and alcohol consumption, which are responsible for 30% of cancer related deaths globally. Smoking tobacco is the biggest risk factor for lung cancer deaths (71%) worldwide and 20% of other global cancers. In developing countries, viral infections, e.g. HCV and HPV, are responsible for 20% of cancer related deaths ("WHO | Cancer"). Other factors may include air pollution, exposure to radiation, exposure to carcinogenic chemicals and risk of inherited cancer.

In 2013, WHO and the International Agency for Research on Cancer (IARC) revealed the Global Action Plan for the Prevention and Control of Non-communicable Diseases 2013 - 2030 which aims to relatively reduce the number of pre-mature mortality cases from cancer, diabetes, cardiovascular and respiratory diseases by 25% globally ("WHO | Global Action Plan for the Prevention and Control of NCDs 2013-2030"). Within the action plan, nine voluntary aims were set to be achieved by the end of 2025 and some

were relevant to cancer prevention such as the aim to reduce alcohol consumption by at least 10%, a 30% reduction in the current rate of worldwide tobacco consumption, increase physical activity among the global nations, and 80% availability of essential medicines and basic hospital facilities required for the treatment of non-communicable diseases ("WHO | Global Action Plan for the Prevention and Control of NCDs 2013-2030").
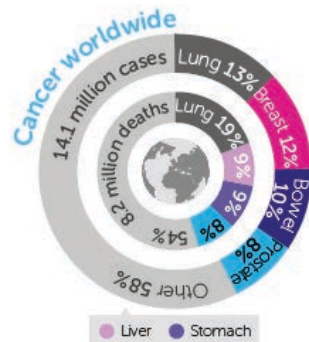
Figure 1.1 Global cancer incidence and mortality

This figure has been taken from Cancer Research UK's website (www.cancerresearchuk.org) which shows global cancer statistics for cancer diagnosis and cancer death in 2012.
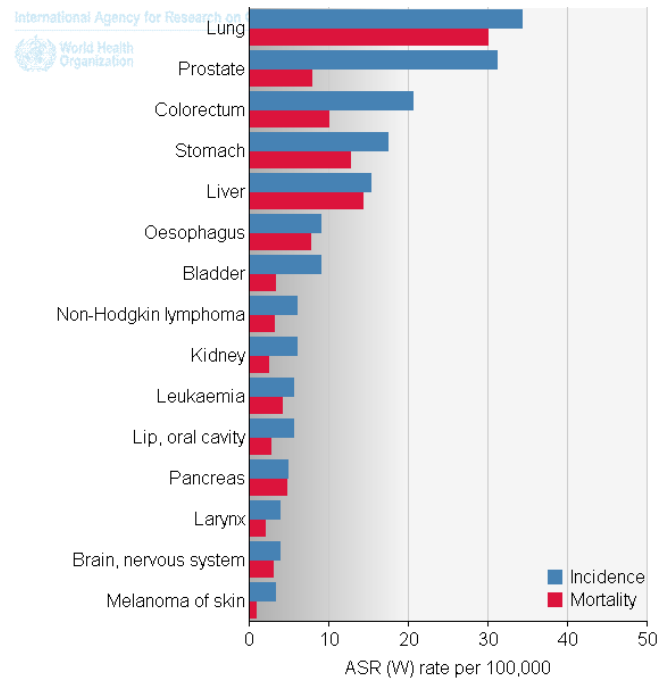
Figure 1.2 Estimated global incidence and mortality rates in male cases

This figure shows the estimated incidence (blue) and mortality (red) rate of the most common types of cancers in males. ASR = age standard rate. The figure is taken from the World Health Organization, GLOBOCAN 2012 website (http://globocan.iarc.fr).
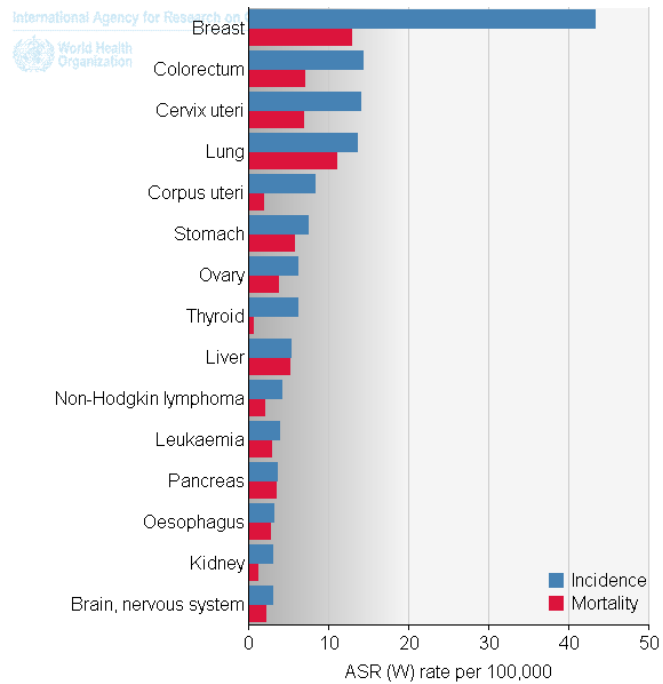
Figure 1.3 Estimated global incidence and mortality rates in female cases

This figure shows the estimated incidence (blue) and mortality (red) rate of the most common types of cancers in females. ASR = age standard rate. The figure is taken from the World Health Organization, GLOBOCAN 2012 website (http://globocan.iarc.fr).

1.2 Cancer genetics

For many decades, researchers have tried to answer the question "what are the cellular alterations that may cause cancer". To date, no definitive answer has been found and there is no single cause for these alterations.

The probability for developing cancer increases with age which led to the hypothesis of multiple cellular hits. The theory that cancer is a result of the accumulation of multiple DNA mutations was initially proposed by Carl O. Nordling in 1953 (Nordling 1953). Nordling proposed that the incident of cancer increases by six to the power of age and requires 6 consecutive hits (Nordling 1953). Later in 1971, Knudson formulated the theory by performing multiple statistical analysis of 48 sporadic and familial retinoblastoma cases (Knudson 1971). Knudson observed that cancer could be caused by at least two DNA alterations which occur at certain rates. In familial retinoblastoma cases, one hit was inherited and the second hit was acquired later; while in sporadic retinoblastoma, both hits were somatic ("Milestone 9: Nature Milestones in Cancer"; Knudson 1971). In later years after the Knudson two-hits hypothesis, many related concepts have emerged till the current view of "the development of cancer"; that all cancers are developed due to multiple genetic and epigenetic alterations which are acquired during the progression of tumour development ("Milestone 9: Nature Milestones in Cancer"). These alterations affect the process of carcinogenesis either by the activation of oncogenes or the deactivation of tumour suppressor genes (TSGs).

1.2.1 Oncogenes

Proto-oncogenes are normal genes which when mutated become oncogenes that accelerate cell proliferation and division uncontrollably. The first discovered oncogene was *Src*, which was as an oncogene in a chicken retrovirus. Later, the American scientists J. Michael Bishop and Harold E. Varmus found that oncogenes were the activated form of proto-oncogenes and that they can be found in all organisms including humans (Bishop 1985). Bishop and Varmus were jointly awarded the Nobel Prize in physiology and medicine in 1989 for their discovery of the retroviral oncogenes.

The main function of proto-oncogene proteins is the regulation of cell differentiation and proliferation. Any changes in a proto-oncogene's sequence or an increase in its protein production leads to uncontrolled cell proliferation and later tumour formation. Proto-oncogenes can be activated to oncogenes by different mechanisms; (i) point mutations, which can arise spontaneously or be acquired by exposure to carcinogenic agents or radiation. Mutations within a proto-oncogene sequence can change the protein structure and cause an increase in protein production. A famous example is *RAS* proto-oncogene, which is found to be mutated in >20% of human cancers including pancreas, colon, lung, breast, skin, kidney and leukaemia (Rajalingam et al. 2007). (ii) Chromosomal translocation which leads to high expression and uncontrolled production of the active protein. For instance, the Philadelphia chromosome translocation of chromosome 22 (which contains *BCR* gene) and chromosome 9 (which contains *ABL1* gene) in leukaemia. Translocation leads to uncontrolled expression of the BCR-ABL fused gene and causes cells to grow and proliferate uncontrollably (Talpaz et al. 2006), and (iii) gene amplification, resulting in an overproduction of the onco-protein such as amplification of

KRAS and *BRAF* in colorectal cancer (Little et al. 2011) and *H-RAS* in thyroid cancer (Várkondi et al. 2005).

1.2.2 Tumour suppressor genes

Tumour suppressor genes (TSG) mainly function as the regulator of cell cycle and promoting cell apoptosis by; (i) repression of genes inducing the cell cycle, (ii) increasing the expression of DNA repair genes, and (iii) initiating cell apoptosis for abnormal cells or cells with damaged DNA. The first discovery of a tumour suppressor gene was in 1987 when Lee et al. cloned the retinoblastoma (*RB1*) tumour suppressor gene (W. H. Lee et al. 1987). Later in 1971, Knudson proposed the "two-hits hypothesis" based on the observation of genetic mutations in the RB1 gene (section 1.2) (Knudson 1971).

In normal cells, tumour suppressor genes work in parallel with proto-oncogenes to control cell proliferation. Tumour suppressor proteins restrain cell growth, while proto-oncogenes code for proteins that promote cell growth. DNA alterations of either type of gene lead to disruption of the cell cycle and uncontrolled cell growth. A number of well-known tumour suppressor genes have been identified such as *TP53* that encodes the p53 protein. p53 plays many cellular roles by initiating apoptosis and inhibiting angiogenesis. More than 50% of all cancers contain mutations or deletions of the *TP53* gene which led research studies to focus on developing gene therapy to restore p53 function in the cells (Ventura et al. 2007). Other examples include mutations in *BRCA1* and *BRCA2* genes which are associated with breast cancer and ovarian cancer (King, Marks, and Mandell 2003), and mutations in the *NF1* gene in neurofibroma (Sawada et al. 1996). Tumour suppressor genes should not be confused with metastasis suppressor genes (MSGs).

9

MSGs have the ability to inhibit the development of metastasis in distant organs without controlling tumour growth at the primary site (HORAK et al. 2008).

With the development of cost-effective NGS technologies, many novel tumour suppressor genes have been identified and are frequently found to be mutated in many cancers such as in the genes of SWI/SNF signalling pathway complex, which were found to be mutated in many types of cancers (Shain and Pollack 2013; Smith et al. 2012; Rousseau et al. 2011).

1.2.3 Pathways in cancer

Cancer is a result of the accumulation of cellular alterations. These alterations may occur at the level of the chromosomes, such as gain/loss of regions or chromosomal rearrangement, or the level of the genes, such as mutations, deletion and amplification, or epigenetic alterations, such as histone modification and DNA hypermethylation. Despite the type of alteration, the end result affects either the structure or the function of the protein encoded by the gene. Therefore, many studies now focus on screening tumours to identify cancer-associated alterations. However, it has been proposed that proteins encoded by mutated genes are actually clustered in certain pathways, in certain tumour types, and are affecting the same functional processes (Baudot, de la Torre, and Valencia 2010). Therefore, the pathway analysis approach is now widely applied due to the availability of the large-scale high-throughput technologies, including targeted NGS panels (such as TruSight sequencing panels from Illumina).

Many of the identified signalling pathways are involved in cell adhesion, apoptosis, DNA repair, angiogenesis, cell cycle and other critical cellular functions. Alteration of these

signalling pathways results in the formation of cancer cells that have the ability to spread uncontrollably and invade other tissues. For instance, the epidermal growth factor receptor (EGFR) and HER2 signalling receptors play a role in cell proliferation and survival, which are frequently overexpressed in cancer cells (Bianco et al. 2006). Growth factor receptors of the insulin pathway are also commonly mutated in cancer, including K-RAS, B-RAF and insulin-like growth factor (I-GF). Pathways involved in angiogenesis have also been widely studied to understand the process of angiogenesis and to find therapeutic cancer treatments targeting this process.

Some studies have found overlapping genes between different types of cancer by using bioinformatics analysis and mapping pathway genes. For example, studies have shown that mutations in K-RAS and EGFR are mutually exclusive in non-small-cell lung cancer and are associated with poor patient prognosis (Eberhard et al. 2005).

The study of gene mutations involved in cancer pathways may reveal new cancer genes and investigation of the molecular mechanisms behind specific cancer types will help with the identification of diagnostic and prognostic biomarkers and developing targeted therapies.

1.2.4 Genome-wide association studies (GWAS)

Genome-wide association studies are used to identify single nucleotide polymorphisms (SNPs) or single nucleotide variants (SNVs) that may cause diseases such as cancer. This method has been widely used since the development of the HapMap (Haplotype Map) project in 2002. The principle of the HapMap project was to sequence the genomes of hundreds of individuals from different populations to produce a map of human genomes

showing human genetic variations. The project started with populations of European, African, Chinese and Japanese descent. Later, the project was expanded to include 11 global populations to identify genome-wide SNPs and copy number polymorphisms (CNPs) (*Nature* 2010).

In a typical GWAS design, a group of individuals that have the disease are matched with normal individuals (control) to reduce confounding variants between the groups. When a variant is found to be significantly more frequent in the population with the disease, this variant is called a disease-causing (or disease-associated) variant.

With the numerous GWAS published studies, many databases have archived GWAS data of SNPs and CNVs from different diseases and different populations that can easily be accessed by researchers. The Single Nucleotide Polymorphism database (dbSNP) of NCBI, the 1000 Genomes Project database, and the Human Gene Mutation Database (HGMD) are well known SNP databases among different populations. Other databases are also available for certain diseases such as Exome Variant Server (EVS) for SNPs contributing to heart, lung and blood diseases, and the Catalogue Of Somatic Mutations In Cancer (COSMIC) database for variants associated with cancers.

1.2.4.1 Next generation sequencing (NGS) technology

The exponential development of cancer research requires advanced research tools that are cost effective and able to study dozens of tumours in depth and within a shorter period of time. In recent years, many high-throughput technologies have been developed to facilitate cancer research. For instance, next generation sequencing, through whole genome sequencing (WGS) and whole exome sequencing (WES), are promising

diagnostic techniques used in clinics for the identification of DNA alterations, genetic recombination and mutations, and are prognostic tools that guide clinicians to look for specific characteristics in patients. RNA sequencing (RNA-Seq) has recently revolutionized the study of gene expression by measuring the relative expression changes and identifying other transcripts and their isoforms, in addition to the identification of fusion genes (Wang, Gerstein, and Snyder 2009). Bisulfite sequencing and methylation microarrays are all examples of the high-throughput genomic technologies that provide further insights into cancer biology.

NGS technology finds genetic variations in human populations, in addition to identifying rare variants involved in diseases and copy number variants (CNVs). In cancer WGS, tumour samples are comprehensively sequenced to identify somatic variants in coding and non-coding regions. On the other hand, cancer WES finds cancer variants within the protein coding regions (exons). These technologies perform massive parallel sequencing of DNA fragments from one tumour sample in less than one day (Grada and Weinbrecht 2013).

Many platforms are available that provide low cost, high-throughput sequencing services. The most popular platforms are the Life Technologies Ion Torrent Personal Genome Machine (PGM) and the Illumina MiSeq (Figure 1.4). Both platforms use the same methodology for DNA template preparation; the DNA library is prepared by fragmenting the DNA and then adapters (synthetic primers of known sequences) are added to the end of each fragmented DNA. Then, DNA libraries are prepared for amplification and clone sequencing. In Ion Torrent PGM, emulsion PCR is used to amplify each library fragment onto microbeads. While Illumina MiSeq uses bridge amplification onto flow cells (Figure

1.4) (Berglund, Kiialainen and Syvänen 2011; Quail et al. 2012). Each amplified library fragment is sequenced through cycles to synthesize new DNA fragments by incorporating nucleotides in the growing DNA strand. Each PGM and MiSeq platform uses a different mechanism for detecting the synthesized sequences; the PGM platform detects pH changes caused by the production of hydrogen ion during the process of nucleotide incorporation into the growing DNA strand (Quail et al. 2012), whereas the Illumina MiSeq platform detects fluorescence light generated by fluorescence labeled nucleotide incorporation (Grada and Weinbrecht 2013; Quail et al. 2012) (Figure 1.4). Following the completion of sequencing, the generated raw data are processed through analysis pipelines to remove adapter sequencing and low quality reads, align sequencing data to a reference genome and then to analyse the final sequences (Grada and Weinbrecht 2013).
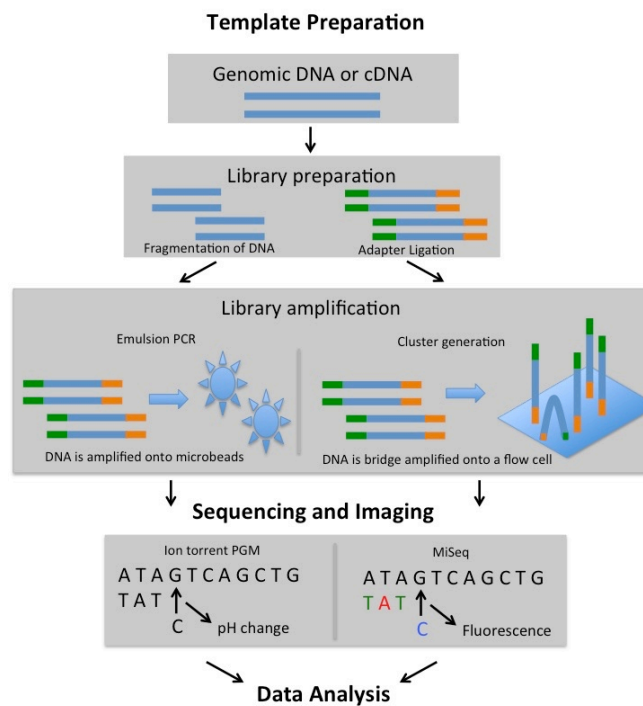
Figure 1.4 Comparison between Ion Torrent PGM and Illumina MiSeq next generation sequencing methodologies

This figure was adapted from (Grada and Weinbrecht 2013).

Researchers have to understand the purpose and the strategy of the experiment prior to selecting the WGS or WES method for sequencing. WES has been extensively used in cancer research to investigate somatic mutations in gene-coding regions. WES is faster and cheaper than WGS since the exome comprises 1% of the entire genome (Bick and Dimmock 2011). However, both technologies have a few limitations, some of which can be resolved or controlled. For instance, the generated data is massive which requires the time consuming analysis and application of additional bioinformatics tools to reduce the list of candidate variants such as tools of genomic prioritization, for example CANDID (dsgweb.wustl.edu/hutz/) and Endeavour (tomcat.esat.kuleuven.be/endeavour/), effect of the mutation on protein function using tools such as Polyphen and SIFT, and algorithmic applications for the identification of variants such as VarScan. In addition, different pipelines and multiple statistical analyses may be applied to reduce the probability for false calls and technical errors. In cancer studies, sequencing tumour samples and matching normal samples is required to overcome the heterogeneity of cancer cells. Furthermore, the identified candidate variants require validation in the lab.

However, NGS is a powerful tool that is more efficient and easier to use for the identification of mutations compared to other methodologies. NGS has shown successful results such as the identification of *NOTCH1* and *MyD88* mutations in CLL patients, which leads to new subgroups for therapeutic targets being determined (Puente et al. 2011), the discovery of *IDH1* mutation in glioblastoma (Parsons et al. 2008a), and germ-line mutation in *PALB2* in hereditary pancreatic cancer (Jones et al. 2009). Therefore,

NGS has now emerged in clinical practice and is used as a diagnostic method for certain diseases.

1.3 Cancer epigenetics

Epigenetic alterations in cancer genome either affect DNA methylation or histone modification. This thesis studies DNA methylation changes in certain types of brain tumours. However, this introduction will provide a brief description of major epigenetic terminologies and mechanisms in cancer.

1.3.1 Histone modification in cancer

Histone modifications affect overall chromatin structure and control the binding of other molecules to the DNA, either positively or negatively. Chromatin structure is important in regulating gene expression by modifying the configuration of the chromatin. Histone modification, including methylation, acetylation, ubiquitination, phosphorylation, biotinylation and sumoylation can affect and change gene transcription (Figure 1.5) (Dalgliesh et al. 2010). Histone acetylation of lysine residue changes the configuration of histone by neutralizing the positive charge of lysine which may consequently reduce the attraction bond between the histone and the negatively charged DNA, leading to a more accessible chromatin for transcription. Acetylation and deacetylation are regulated by the activity of the enzymes histone lysine acetyltransferases (HAT family) and histone deacetylases (HDACs). Here HDAC removes the acetyl group from lysine residue and restores the positive charge and HAT adds the acetyl group. Both HAT and HDAC enzymes are required to function in equilibrium to regulate gene expression and

transcription. Dysregulation of the proteins was reported to cause gene silencing and cancer cell transformation. HAT was firstly identified as a histone modification enzyme in cancer and has the capability to transform the viral oncoprotein E1A (Bannister and Kouzarides 1996). Expression aberrations of HATs proteins were also reported in many types of cancer (Avvakumov and Côté 2007).

The HDAC family contains 18 enzymes, which are divided into 4 classes depending on their sequence homology. Class I (HDAC1-3, and HDAC8), class II (HDAC4-7 and HDAC9-10) and class IV (HDAC11) require Zn^{+2} metal as a cofactor, while class III (sirtuin 1-7) is NAD^+ dependant. In contrast to HATs proteins, HDACs proteins regulate histone and non-histone targets, such as in leukaemia where HDACs interact with BCL6 to repress the protein activity by acetylation (Bereshchenko, Gu, and Dalla-Favera 2002).

Phosphorylation usually occurs in serine, threonine and tyrosine residues of the histone. Phosphorylation causes an alteration of the protein's charge, which affects the ionic properties and the overall configuration of the chromatin. Histone phosphorylation can affect either transcript regulation or chromatin condensation, however, H3S10 is associated with both functions (Baek 2011).

Histone methylation in lysine residue has also been described widely in cancer. Histone methylation affects gene expression by changing the structure of the chromatin rather than changing the charge of the residue. For instance, the methylation of H3K36 and H3K70 allows chromatin to open up which increases gene expression. On the other hand, methylation of H3K9 and H4K20 is associated with closed and repressive chromatin (Dawson and Kouzarides 2012). Several histone methyltransferases have been involved in cancer. The most important enzymes are EZH2 (H3K27 histone methyltransferase) and

its overexpression was associated with poor prognosis in breast and prostate cancer (Cao and Zhang 2004), and SMYD3 (H3K4 methyltransferase) in colon and liver cancer (Hamamoto et al. 2004).
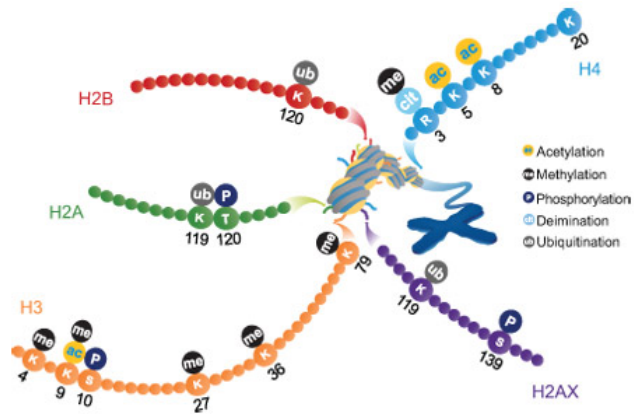
Figure 1.5 Types of histone modifications, methylation, acetylation, ubiquitination, phosphorylation, and biotinylation that affect gene transcription.

All these histone modifications can affect the chromatine structure and molecules binding to the DNA, either positively or negatively. This figure was taken from (http://www.abcam.com).

1.3.2 DNA methylation in cancer

Methylation of the cytosine residue at 5' carbon (5mC) usually occurs within cytosine-guanine (CpG) dinucleotides. Most studies that have observed epigenetic alterations in cancer have reported methylation changes within the CpG islands. CpG islands are regions with concentrated CpG sites and are distributed in >70% of gene promoters (Malzkorn et al. 2011). The methylation of CpG islands is associated with transcriptional alterations, which is commonly observed during tumour transformation (Baylin and Jones 2011). Studies using advanced genome-wide technologies such as NGS have found that 5-10% of unmethylated CpG islands become hypermethylated in various types of cancer genomes. In addition, DNA hypermethylation does not only affect the gene expression of coding genes but also affects the expression of non-coding RNAs, some of which have target genes that are also involved in tumorigenesis (Baylin and Jones 2011). More importantly, genome-wide methylation studies have identified DNA methylation alterations within CpG shores, which are located 2kb upstream from CpG islands, and CpG shelves, which are located upstream from CpG shores, that were strongly associated with aberrant gene expression (Irizarry et al. 2009).

DNA methylation is regulated by 3 DNA methyltranferases (DNMTs) that modify cytosine methylation in CpG dinucleotides. DNMT1 is responsible for hemimethylated DNA maintenance and methylates newly synthesized CpG dinucleotides following DNA replication. On the other hand, DNMT3a and DNMT3b are also capable of maintaining hemimenthylated DNA but their primary role is de novo methyltransfereases targeting unmethylated CpG dinucleotides. The role of DNA methylation depends on the location

of the CpGs in the genome. For instance, DNA hypermethylation in the promoter regions represses transcription and causes gene silencing. While methylation of the repetitive regions increases genome instability (Malzkorn et al. 2011).

Researchers have been focusing on TSGs in cancers. TSGs are frequently silenced in cancer by DNA hypermethylation. Global hypomethylation has also been observed in cancer and associated with repetitive DNA elements (Melanie Ehrlich 2009). The first reported epigenetic alteration in cancer was DNA hypomethylation in 1983 when Gama-sosa and his group identified global DNA hypomethylation in benign/malignant tumours in comparison to normal tissues (Gama-Sosa et al. 1983). Promoter hypomethylation of proto-oncogenes may activate gene transcription such as promoter demethylation of the proto-oncogene, *HOX11*, in leukaemia (Watt, Kumar, and Kees 2000). Hypomethylation of highly repetitive DNA sequences are the main reason for global DNA hypomethylation in cancer which causes genomic instability (Figure 1.6). Cancer global hypomethylation was observed in satellite DNAs such as tandem centromeric and pericentromeric satellite repeats, and the interspersed Alu and LINE-1 elements in different types of cancer including ovarian cancer, prostate metastasis, leukaemia, liver and colon cancers (Bedford and van Helden 1987; Feinberg et al. 1988; Lin et al. 2001; M. Ehrlich et al. 2006; Hoffmann and Schulz 2005; Rodriguez et al. 2008).

Cancer-associated DNA hypermethylation is usually found in CpG islands. A large methylation study of 98 primary cancers, including breast, colorectal, brain and leukaemia, using restriction landmark genomic scanning of CpG islands in the genome, showed that an average of 600 out of 45,000 CpG islands were highly methylated in tumour samples but not in normal control samples (Costello et al. 2000).

Hypermethylation of the promoter region CpG island(s) of TSGs reduces the expression of proteins that are involved in various regulatory processes such as cell apoptosis, DNA repair and anti-angiogenesis (Malzkorn et al. 2011) (Figure 1.6). However, recent studies indicate that regions outside CpG islands, such as the CpG shore (2kb from CpG island) and shelf (2kb from CpG shore), play an important regulatory role in cell proliferation and differentiation during tumorigenesis. The significance of the role of isolated CpGs within open sea regions remains unclear.

DNA hypermethylation has been identified in several pathways; the frequently mutated gene in cancer, *p53*, has also been reported as inactivated by methylation-mediated repression of the TSG, *p14ARF*. *p14ARF* normally inhibits the MDM2 protein that promotes p53 degradation (Manel Esteller, Tortola, et al. 2000). The cell cycle inhibitor p16INK4a is commonly hypermethylated in various types of cancer (Gonzalez-Zulueta et al. 1995). Also, there is hypermethylation of DNA repair pathway genes such as the mismatch repair gene *hMLH1* in gastric and colon cancer (Fleisher et al. 1999; Herman et al. 1998, 1), *MGMT* promoter in colorectal and brain cancer (M. Esteller et al. 1999; Mitsutoshi Nakamura et al. 2001), the mitotic checkpoint *CHFR (Mizuno et al. 2002)*, and the DNA repair gene *BRCA1* in breast and ovarian cancer (Manel Esteller, Silva, et al. 2000).

Not all hypermethylation changes are an indication of disease, as methylation changes in some genes can be a prognostic factor or predictive of a treatment response factor.

Hypermethylation of *p16INK4a* has been associated with poor prognosis for lung and colorectal cancer patients, while CpG island hypermethylation of *hMLH1* is associated with good response to anti-cancer drugs. In addition, silencing of *MGMT* repair gene by

hypermethylation has been considered as an indication, in glioma and lymphoma cancer patients, of a patient's sensitivity to chemotherapy (Manel Esteller et al. 2001; Plumb et al. 2000; Manel Esteller, Garcia-Foncillas, et al. 2000).
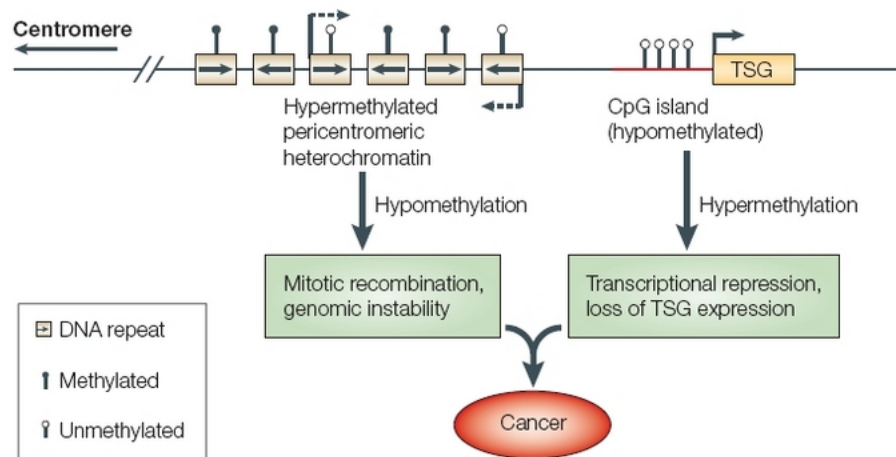
Figure 1.6 DNA methylation and cancer

The above diagram represents a small region of genomic DNA in a normal cell and the contribution of DNA hyper/hypo methylation in cancer formation. DNA Hypomethylation at the pericentromeric regions causes genomic instability, while DNA hypermethylation of TSGs promoter region leads to loss of gene expression. This figure was taken from Sarraf and Stancheva (2004).

1.3.2.1 5-hydroxymethylcytosine

The 5mC can be further modified by adding hydroxyl group to generate 5-hydroxymethylcytosine (5hmC), an oxidation process mediated by the TET (ten-eleven-translocation) family (Pfeifer, Kadam and Jin 2013). 5hmC was found in 40% of brain tissues, indicating an important biological role (Kriaucionis and Heintz 2009). Yet, the main function of 5hmC is still poorly understood and still under research.

Researchers have raised the question of how to differentiate 5hmC from 5mC. Laboratory methods based on the bisulfite treatment of DNA cannot distinguish between 5hmC from 5mC since both cytosines are not converted by sodium bisulfite and are recognised as methylated cytosine. Methylation techniques based on restriction enzymes are usually inhibited by 5hmC and immunoprecipitation that uses antibodies against 5mC does not detect 5hmC (Jin, Kadam, and Pfeifer 2010). Specific antibodies against 5hmC have been commercially designed for 5hmC detection, in addition to other enzyme-based methods that modify the hydroxyl group of 5hmC by using T4 b-glycosyltransferase, which adds the glucose molecule to prevent DNA cleavage at the MspI restriction site (Malzkorn et al. 2011).

1.3.2.2 CpG Island Methylator Phenotype in cancer

CpG island hypermethylation is a main feature of cancer, especially when hypermethylation occurs within the promoter region of the tumour suppressor gene causing gene silencing. Many studies have reported aberrant DNA methylation that occurs in different types of cancer. However, a study in 1999 identified subgroups of colorectal cancer that demonstrated concordant tumour specific hypermethylation and

referred to such tumours as exhibiting a CpG island methylator phenotype (CIMP) (M. Toyota et al. 1999). Tumours exhibiting CIMP were characterised by a high methylation index and *KRAS* mutations in comparison to non-CIMP tumours (M. Toyota et al. 2000). Subsequently, additional studies have expanded these findings and were able to identify additional subgroups of colorectal cancers; non-CIMP tumours with *TP53* mutations, CIMP tumours with mutations of *BRAF* gene, and CIMP tumours with *KRAS* mutations (Ogino et al. 2006; Shen et al. 2007). Additionally, in Noushmehr's study, CIMP was identified in a subgroup of glioma tumours (G-CIMP) using 7 signature genes (*ANKRD43*, *HFE*, *MAL*, *LGALS3*, *FAS*, *RHO* and *DOCK5*) and the phenotype was tightly associated with *IDH* mutations (Noushmehr et al. 2010).

CIMP has been observed in other cancer types such as breast cancer (Fang et al. 2011), lung cancer (Marsit et al. 2006), ovarian cancer (Strathdee et al. 2001) and leukaemia (Roman-Gomez et al. 2006). These findings indicate that CIMP is perhaps not restricted to certain types of cancer but involves different cancer-specific genes for each type.

1.3.3 Genome methylation analysis

Most of the methods used to detect DNA methylation changes are based on sodium bisulfite treatment of DNA, which converts the unmethylated cytosine base to an uracil base, while methylated cytosine remains protected. After DNA conversion, the modified DNA becomes a starting material in all subsequent techniques (Figure 1.7). Methylation specific PCR (MSP) is widely used for detecting methylation changes in DNA. MSP is a fast and very sensitive method, which has been selected as the standard method for *MGMT* DNA hypermethylation detection in diagnostic labs. Combined bisulfite

restriction analysis (CoBRA) is a semi-quantitative method that depends on a restriction enzyme site to differentiate between methylated and unmethylated DNA. MethyLight is a quantitative method based on fluorescent real-time PCR to quantify DNA methylation after bisulfite treatment. The golden standard technique for mapping the sequence of methylated DNA is bisulfite modified DNA sequencing, either by Sanger sequencing, which depends on the chain termination method, pyrosequencing, which is sequencing that uses the synthesis method, or NGS of whole genome bisulfite sequencing (WGBS).

Technologies based on epigenome-wide analysis have been widely used for comprehensive DNA methylation analysis. For instance, HumanMethylation BeadChip arrays from Illumina have provided high-throughput and cost effective comprehensive methylome analysis and have been used in several methylation studies on diseases and disorders. Illumina has developed three high-throughput profiling assays; the first and the simplest array is the GoldenGate BeadChip array which covers over 1500 CpG sites per sample. While the HumanMethylation27 BeadChip array can analyse more than 27,000 CpG sites per sample. The new generation HumanMethylation450 BeadChip array provides genome wide coverage and can assess the methylation status of more than 480,000 CpG sites per sample (Figure 1.8). The principle behind the BeadChip technology is the same for all three assays and is based on genotyping bisulfite converted DNA. The BeadChip technology has been adapted to evaluate DNA methylation through SNPs assessment of C and T bases after DNA conversion. In the GoldenGate BeadChip array and HumanMethylation27 BeadChip array, each CpG locus is assessed by Infinium I assay that measures the intensity of methylation by one methylated and one unmethylated probe separately (Figure 1.9). On the other hand, HumanMethylation450

BeadChip array uses both Infinium I and Infinium II assays. Infinium II design uses one probe for each single CpG and determines the methylation status after hybridization and extension (Figure 1.10).
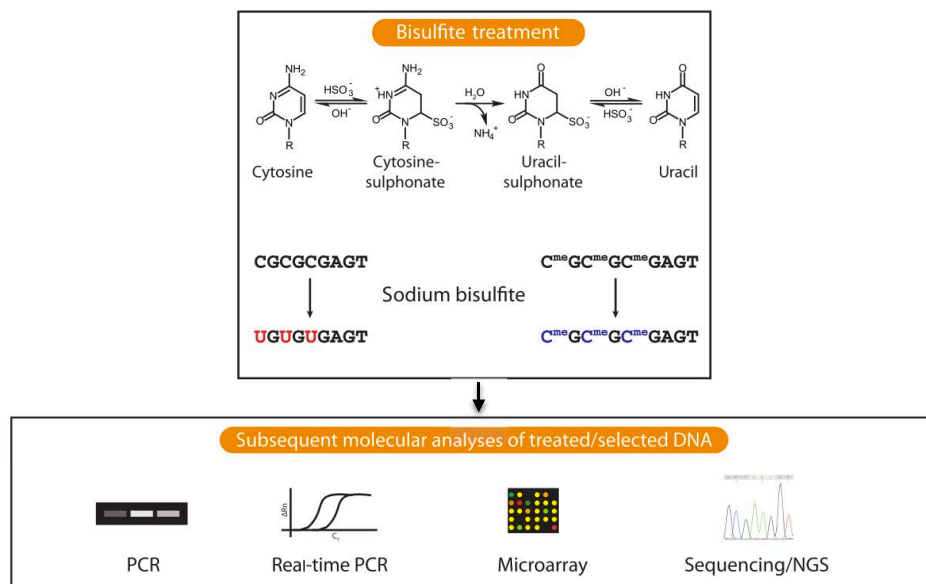
Figure 1.7 Common methods used for DNA methylation analysis

DNA modification by sodium bisulfite treatment is usually the first step in methylation analysis. Subsequent DNA methylation analysis methods such as MSP and CoBRA PCR, real-time Methylight, pyrosequencing, Sanger sequencing and genome-wide methylation arrays are based on distinguishing the converted C to T or unaltered C base, after DNA bisulfite treatment. This figure was taken from Malzkorn et al. (2011).
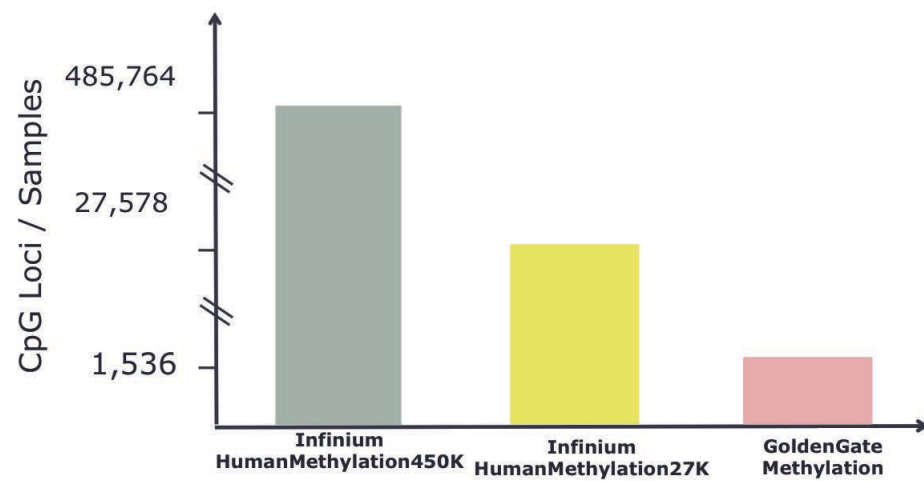
Figure 1.8 Three high-throughput profiling assays from Illumima

The graph shows the number of CpG loci covered per sample for each profiling assay.
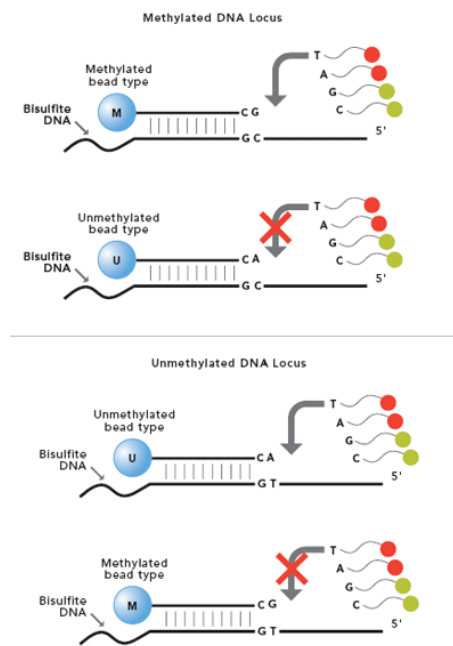
Figure 1.9 Schematic of Illumina Infinium I methylation assay

GoldenGate BeadChip array and HumanMethylation27 BeadChip array use Infinium I assay to measure the methylation status. Infinium I assay uses two separate probes per CpG locus; one for methylated locus and one for unmethylated locus. This figure was taken from www.illumina.com.
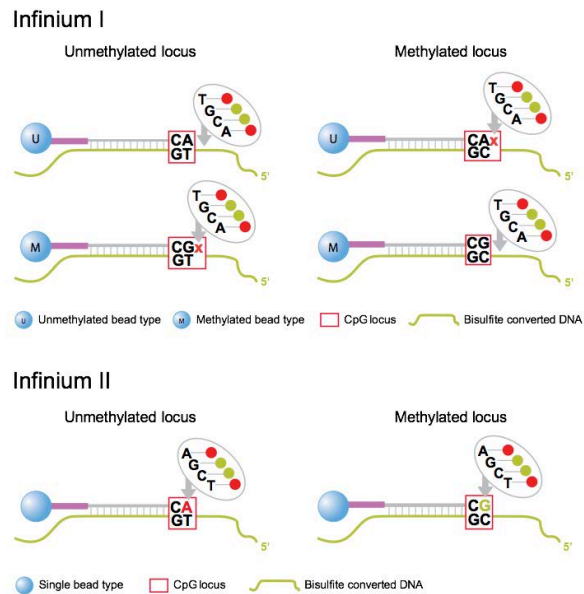
Figure 1.10 HumanMethylation450 BeadChip array uses both Infinium I and

Infinium II assays

Infinium I assay uses two probes, methylated and unmethylated, for each CpG locus. Infinium II assay uses one probe per CpG locus where the methylation status is determined after base hybridization and extension. This figure was taken from www.illumina.com.

The most recent HumanMethylation450 BeadChip array is more comprehensive and is able to assess the methylation status of >450,000 CpG sites distributed around the genome that are within the CpG island, CpG shores, CpG shelves and also single isolated cytosine bases. In addition, the array covers regions within the transcription start sites, the gene's 1st exon, gene body, 5' and 3' un-translated regions, and the coding gene's and miRNA's promoter regions.

The GoldenGate BeadChip array and HumanMethylation27 BeadChip array have been used widely in cancer studies and were successfully able to identify the methylation changes involved in tumorigenesis, such as the successful identification of CIMP phenotypes in renal cell carcinoma and glioma (McRonald et al. 2009; Noushmehr et al. 2010) and the identification of methylated genes in cancer pathways (E. M. Wolff et al. 2010). The HumanMethylation450 BeadChip array has recently been introduced into the research field and most initial studies started to evaluate and validate the new assay. However, the number of published papers and research studies using HumanMethylation450 BeadChip array is currently increasing.

1.4 Cancer genome online databases

With the massive amount of data generated by the Human Genome Project since 2003, researchers are able to study the genetic factors and molecular mechanisms that are involved in human diseases including cancer.

After sequencing the human genome by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), many international projects such as The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) were launched to catalogue and archive DNA changes involved in human cancer. These genome-based studies will facilitate biomedical researchers to identify prognostic and diagnostic markers, improve diagnostic tests and develop customised cancer therapies. In addition, they will provide a deeper understanding of the molecular mechanisms involved in carcinogenesis.

1.4.1 The Cancer Genomic Atlas (TCGA)

The Cancer Genome Atlas (TCGA) was initially launched in 2006 to catalogue most genomic abnormalities involved in human cancer ("The Cancer Genome Atlas" 2015). The TCGA project is managed by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) of the US National Institutes of Health (NIH). Glioblastoma multiforme (GBM) was the pilot project for TCGA which was published in *Nature* 2008 (McLendon et al. 2008). Based on gene expression profiling, copy number variations and sequencing analysis of 206 normal/tumour matched pair GBM cases, the study found that most of the cases carried frequent mutations in *ERBB2*, *NF1* and *TP53* in addition to other alterations in core brain cancer pathways (McLendon

35

et al. 2008). Since 2009, the TCGA project has been expanded to phase II and now to phase III with >11,000 samples and >20 tumour types. Most of the TCGA data are available and free to access, except for data that includes unprotected patient information. TCGA provides data from several platforms using high-throughput methods that analyse DNA mutations, chromosomal alterations and copy number variations; RNA alterations such as gene expression analysis and epigenetic aberrations including DNA methylation. All data in TCGA are generated and analysed by different centres and resources. The Biospecimen Core Resource (BCR) is responsible for the quality control assurance of tissues and samples as well as the quality control of DNA and RNA isolation from the samples ("The Cancer Genome Atlas" 2015). Each sample in TCGA must be examined and verified prior to being submitted to the TCGA data portal. General guidelines such as a minimum 200 mg of tissue weight, at least 80% tumour nuclei, and a matched normal (germ-line) pair sample (matched normal tissue, blood, or both) must be considered for sample approval ("The Cancer Genome Atlas" 2015). However, other quality control standards that vary between tissue types and cancer types are also considered, for instance, brain tumour samples must contain at least 80% tumour nuclei and less than 50% necrosis in addition to a secondary pathology assessment to validate the original diagnosis of the samples (McLendon et al. 2008). TCGA is funding three genome sequencing centres: The Genome Center at Washington University, the Baylor College of Medicine and the Broad Institute, to perform all the next generation sequencing NGS technologies. Also, TCGA is co-funding the Data Coordinating Centre (DCC) which is responsible for deposing data into the TCGA database and maintaining its website. This

work is a collaboration between DCC and SRA International for bioinformatics ("The Cancer Genome Atlas" 2015).

Many published research studies have analysed tumour samples from TCGA and were successful in identifying genetic or epigenetic abnormalities and potential causative or marker genes in different types of cancer. For instance, the hallmark TCGA report on the comprehensive characterisation of 276 colorectal tumour samples from TCGA using exome sequencing, gene expression and DNA methylation analysis (Network 2012). Also, the DNA methylation analysis of glioma samples by Noushmehr et al. which identified a distinct subset of glioma with high and concordant DNA hypermethylation across many CpG loci, indicating the existence of Glioma CpG Island Methylator Phynotype (G-CIMP) (Noushmehr et al. 2010).

By making TCGA data accessible to the public and all analytical tools available to researchers, TCGA will provide the opportunity to expand knowledge and improve cancer research, diagnosis, treatment and also cancer prevention.

1.4.2 Catalogue Of Somatic Mutations In Cancer (COSMIC)

The Catalogue of Somatic Mutations in Cancer (COSMIC) is a freely available online database for somatic mutations in human cancer which combines data from scientific literature, the Cancer Genome project (CGP) at the Wellcome Trust Sanger Institute, The Cancer Genomic Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) projects (S.A. Forbes et al. 2008; Simon A. Forbes et al. 2015). COSMIC was first launched in 2004 when somatic mutation data from 4 cancer genes, *HRAS*, *KRAS2*, *NRAS*, and *BRAF* was published (Bamford et al. 2004). Now the COSMIC database has

been expanded to include data from 28,735 genes/transcripts, >1 million tumour samples and >2 million mutations including coding mutations, non-coding mutations, gene rearrangements, abnormalities in copy number and more (Simon A. Forbes et al. 2015) (Table 1.1). Approximately half of the data in COSMIC are collected from published literature and the other half are from different cancer genomic projects. Data are analysed by a software pipeline and run through a specific quality control to be published later on the website with the reference. Evaluating each gene/mutation/tumour data prior to it being published is important to exclude false imperfect results.

The COSMIC website (cancer.sanger.ac.uk/) is user friendly and free to access which makes it possible to browse the data either in a gene-centric or tissue-centric fashion. Typically, the search starts by selecting the gene name or cancer type, then different main webpages show different information that can be navigated separately and in depth such as gene overview, sample overview, mutation overview and tissue overview (Figure 1.11). To navigate the data, diagrams of pie charts and histograms provide a breakdown of mutation types (non-sense, missense, frameshift, etc.), mutation counts and sample type (primary samples, cell-line or unknown).

With the increase in the development of cancer genome data, COSMIC allows the integration of all cancer genomic data directly into the Ensembl Genome Browser (www.ensembl.org). Data of gene mutations have been uploaded from COSMIC into the Ensembl database and incorporated in the Ensembl website as 'Somatic_SNV' annotations to differentiate these somatic variants from other common SNPs (Simon A. Forbes et al. 2015). The COSMIC mutations can be found in the Ensembl gene

transcripts and are provided with links to the COSMIC website for more detailed information about the variant. In addition, many other projects and databases are supporting and integrating COSMIC data, such as Intogen (http://www.intogen.org/home), UniPort (http://www.ebi.ac.uk/uniprot) and Pfam (http://pfam.sanger.ac.uk) projects. However, Ensembl was the first external project which added COSMIC data to their website and established "COSM" ID (Cosmic Somatic Mutation identifier) for COSMIC somatic SNVs (Simon A. Forbes et al. 2015).

Total contents in v70 of the COSMIC database (the August 2014 release)	
Genes (transcripts)	28,735
Tumour samples	1,029,547
Coding mutations	2,002,811
Curated publications	19,703
Fusion mutations	10,435
Genomic rearrangements	61,299
Whole genomes	12,542
Copy number aberrations	695,504
Gene expression variants	60,119,787

Table 1.1 Total contents in v70 of the COSMIC database (the August 2014 release)

This table was redrawn from Simon A. Forbes et al. (2015).
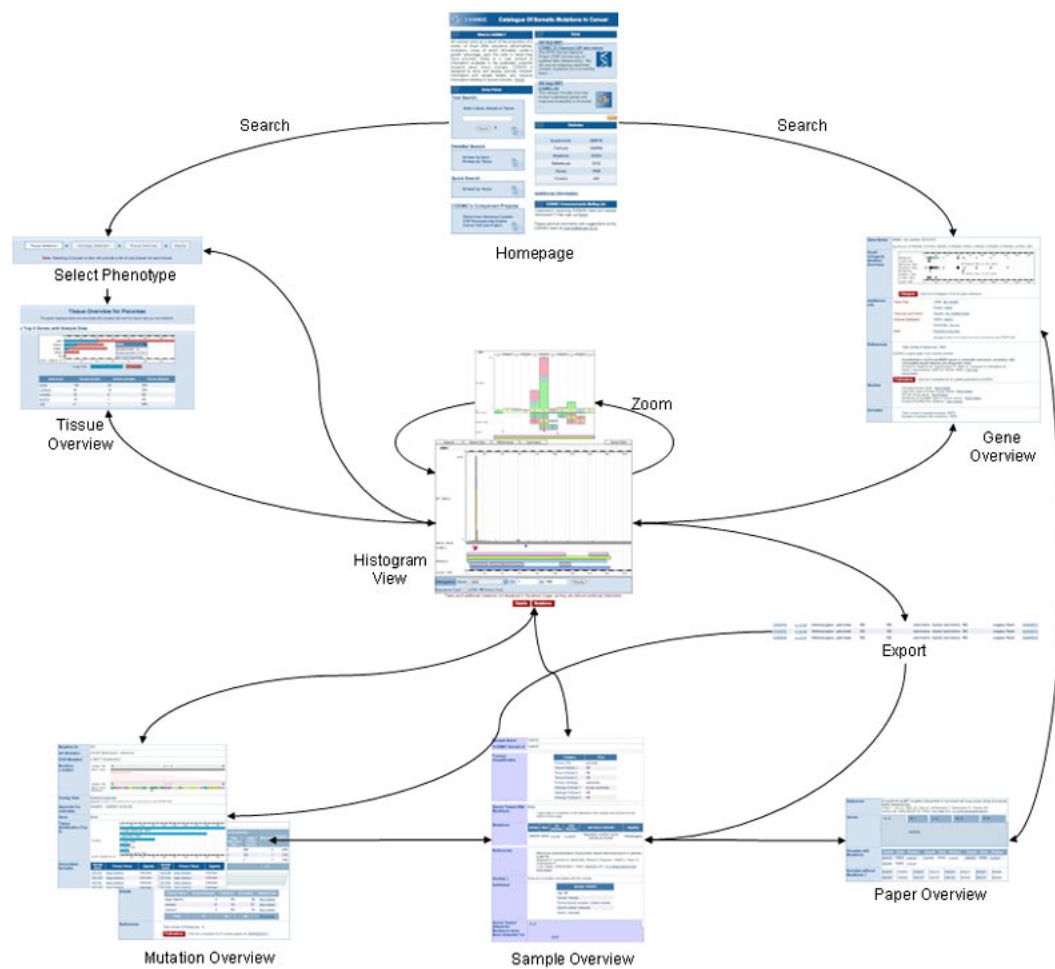
Figure 1.11 Main webpages to navigate COSMIC database

This figure was taken from Simon A Forbes et al. (2008).

1.5 CNS tumours

In 2012, it was estimated that 256,000 brain and other CNS tumour cases were diagnosed worldwide (Ferlay et al. 2015). CNS tumours are named after the cell of origin and they are classified into: (i) benign tumours; the most common types are meningioma, schwannoma and neuroma, (ii) malignant tumours; these include glioblastoma multiforme (GBM), astrocytoma, oligodendroglioma and ependymoma, (iii) metastatic brain tumours; where cancer starts in a different primary organ and later spreads to the brain. The most common primary sites for brain metastasis are breast, lung, melanoma, colon and kidney (Gavrilovic and Posner 2005).

1.5.1 Classification of CNS tumours

In 1979, the WHO announced a new classification and grading of tumours of the CNS to standardise the nomenclature of CNS tumours and communication around the world (Louis et al. 2007a). This classification was revised several times by different international groups and edited in 1993, then in 2000, and the final fourth edition was published in 2007 (Molnar 2011). The classification was based on the premise of tumour initiating cells (Figure 1.12), while the grading was according to their malignant potential (Figure 1.13). The CNS tumours grades ranged from grade I (benign) tumours to grade IV (most aggressive) tumours (Molnar 2011) (Table 1.2).
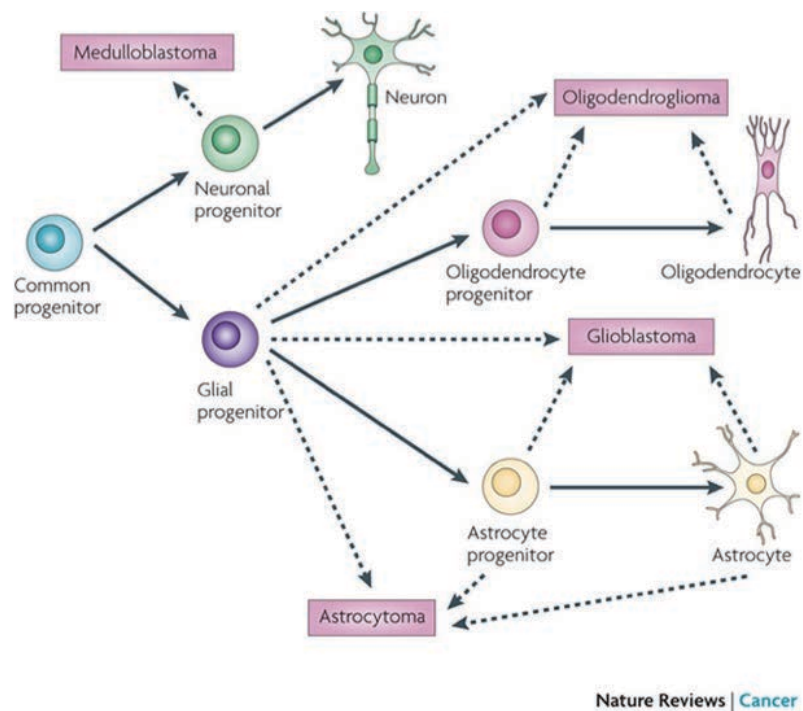
Figure 1.12 Central nervous system tumour initiating cells

Common progenitors are expected to produce neuronal and glial progenitors that differentiate into mature cells. The origin cells of medullablastoma and some of diffuse glioma are still unknown. The candidate cells of origin are indicated by dotted lines. This figure was taken from Huse and Holland (2010).

Grade I	Low proliferative potential and possibility of surgical cure	Well-differentiated
Grade II	Diffuse growth, tendency to recur, though only few mitoses	Moderately-differentiated
Grade III	Histological evidence of malignancy, destructive growth, brisk mitotic activity, anisonucleosis	Poorly-differentiated
Grade IV	Cytologically malignant, mitotically active, necrosis-prone	Anaplastic

Table 1.2 WHO four-tier grading scheme for CNS tumours

This table was redrawn from Molnar (2011).

The WHO's grading is based on a combination of criteria used to predict patients' therapy response and patients' outcome in addition to other criteria including genetic alterations and clinical information about the patient such as age, site of tumour, etc. The WHO's grading system is used as a parameter for patient prognosis, for instance, patients with WHO grade II tumours usually show a longer survival rate (>5 years), while the median survival rate for patients with WHO grade III tumours is 2-3 years. Tumours with WHO grade IV are typically fatal if untreated and patients can survive for up to 12 months (Louis et al. 2007a).

WHO grades of CNS tumours

	I	II	III	IV
Astrocytic tumours				
Subependymal giant cell astrocytoma	•			
Pilocytic astrocytoma	•			
Pilomyxoid astrocytoma		•		
Diffuse astrocytoma		•		
Pleomorphic xanthoastrocytoma		•		
Anaplastic astrocytoma			•	
Glioblastoma				•
Giant cell glioblastoma				•
Gliosarcoma				•
Oligodendroglial tumours				
Oligodendroglioma		•		
Anaplastic oligodendroglioma			•	
Oligoastrocytic tumours				
Oligoastrocytoma		•		
Anaplastic oligoastrocytoma			•	
Ependymal tumours				
Subependymoma	•			
Myxopapillary ependymoma	•			
Ependymoma		•		
Anaplastic ependymoma			•	
Choroid plexus tumours				
Choroid plexus papilloma	•			
Atypical choroid plexus papilloma		•		
Choroid plexus carcinoma			•	
Other neuroepithelial tumours				
Angiocentric glioma	•			
Chordoid glioma of the third ventricle		•		
Neuronal and mixed neuronal-glial tumours				
Gangliocytoma	•			
Ganglioglioma	•			
Anaplastic ganglioglioma			•	
Desmoplastic infantile astrocytoma and ganglioglioma	•			
Dysembryoplastic neuroepithelial tumour	•			

	I	II	III	IV
Central neurocytoma		•		
Extraventricular neurocytoma		•		
Cerebellar liponeurocytoma		•		
Paraganglioma of the spinal cord	•			
Papillary glioneuronal tumour	•			
Rosette-forming glioneuronal tumour of the fourth ventricle	•			
Pineal tumours				
Pineocytoma	•			
Pineal parenchymal tumour of intermediate differentiation		•	•	
Pineoblastoma				•
Papillary tumour of the pineal region		•	•	
Embryonal tumours				
Medulloblastoma				•
CNS primitive neuroectodermal tumour (PNET)				•
Atypical teratoid / rhabdoid tumour				•
Tumours of the cranial and paraspinal nerves				
Schwannoma	•			
Neurofibroma	•			
Perineurioma	•	•	•	
Malignant peripheral nerve sheath tumour (MPNST)		•	•	•
Meningeal tumours				
Meningioma	•			
Atypical meningioma		•		
Anaplastic / malignant meningioma			•	
Haemangiopericytoma		•		
Anaplastic haemangiopericytoma			•	
Haemangioblastoma	•			
Tumours of the sellar region				
Craniopharyngioma	•			
Granular cell tumour of the neurohypophysis	•			
Pituicytoma	•			
Spindle cell oncocytoma of the adenohypophysis	•			

Figure 1.13 WHO classification and grading of CNS tumours

This figure was taken from Louis et al. (2007a).

1.5.2 Vestibular Schwannoma

Vestibular Schwannoma is a benign tumour which is characterised by uncontrolled proliferation of schwann cells in an encapsulation way around the nerves. Schwannoma can develop anywhere from vestibular nerves to the nerves of sensory organs and can cause hearing loss, imbalance, tinnitus and death in the late stages (Kondziolka et al. 2012).

Vestibular schwannoma can be classified into 2 groups; sporadic unilateral or bilateral, which is associated with neurofibromatosis type 2 (NF2) and caused by mutations of *NF2* gene. However, patients with NF2 can develop other types of tumours such as meningioma and ependymomas (Torres-Martin et al. 2013).

1.5.2.1 Molecular biology of schwannoma

The genetic and epigenetic changes involved in schwannoma development have been studied in the past few years including *NF2* gene mutation, chromosome 22 LOH and DNA methylation.

1.5.2.2 NF2 gene

NF2 gene has been shown to have tumour suppressing properties, and encodes for the cytoskeletal protein merlin (moesin-ezrin-radixin-like protein). This gene is located in chromosome 22 (17 exons) and is mutated in neurofibromatosis type 2 (NF2) and sporadic schwannoma patients (Torres-Martin et al. 2013). Merlin has the ability to control cell growth and proliferation by inhibiting ubiquitin cascade (Celis-Aguilar et al. 2012). Merlin protein is involved in several signalling pathways such as the Hippo

pathway and PI3K, deregulation of ErbB2/ErbB3 and EGFR (Torres-Martin et al. 2013). Mutations of the *NF2* gene have been identified in NF2 as well as in sporadic vestibular schwannoma patients. According to Knudsen's two-hit hypothesis, sporadic vestibular schwannoma is caused by somatic biallelic *NF2* inactivation and later, *NF2* acquires additional somatic mutations due to impaired DNA repair mechanisms (D. G. R. Evans, Maher, and Baser 2005; D. Gareth R. Evans 2009). However, a few sporadic vestibular schwannoma cases showed no alterations in the *NF2* gene. Genetic alterations of other genes located on chromosome 22 have been identified in schwannoma cases such as *SMARCB1* in 50% of familial schwannoma and 10% of sporadic schwannoma (Rousseau et al. 2011; Smith et al. 2012), and *LZTR1* in schwnomatosis (Piotrowski et al. 2014). These findings indicated that chromosome 22 genes are essential and play an important role in schwannoma development.

1.5.2.3 DNA methylation

DNA methylation of vestibular schwannomas has not been fully studied. Some studies have analysed DNA hypermethylation in tumour associated genes that are frequently hypermethylated in other brain cancers such as *VHL, MGMT, TP16, CASP8, TIMP3, DAPK* and *THBS1*. Less than 10% of schwannoma cases displayed DNA hypermethylation (Lassaletta et al. 2006; Gonzalez-Gomez et al. 2003). In addition, *NF2* promoter hypermethylation is rare in vestibular schwannoma which suggests a different mechanism of *NF2* inactivation (Kullar et al. 2010; J. D. Lee et al. 2012; Gonzalez-Gomez et al. 2003).

1.5.2.4 Loss of heterozygosity

Loss of heterzygocity on chromosome 22 was widely identified in schwannomas as the most common form of second hit with a frequency >50% (Celis-Aguilar et al. 2012). LOH on chromosome 22 is commonly found in the long arm of the chromosome and in conjunction with *NF2* mutation and is correlated with the proliferation activity in schwannoma (Bian et al. 2005; Celis-Aguilar et al. 2012). Mainly, mitotic recombination is the cause of LOH which consists of deletion followed by the re-duplication of chromosome 22 that leads to the formation of 2 identical copies of the mutated gene and removal of the wild-type copy (R. K. Wolff et al. 1992; D. Gareth R. Evans 2009; Hadfield et al. 2010).

1.5.3 Glioblastoma multiforme

Glioblastoma is the most common and frequent type of primary brain tumour in adults. According to the World Health Organization's (WHO) classification criteria, glioma can be histologically classified as astrocytoma, oligoastocytoma, oligoendroglioma and pilocytic astrocytoma (Louis et al. 2007b; Yan et al. 2009a; Riemenschneider et al. 2010a). WHO grade I gliomas are benign and curable by surgery. On the other hand, grade II and III gliomas are malignant and more invasive forms. The majority of gliomas are glioblastoma multiform (GBM) (Astrocytoma grade IV) which is the most aggressive and lethal type of glioma (Wen and Kesari 2008; Chen, McKay and Parada 2012). Generally, GBM are classified into primary and secondary based on their clinical presentation. If the cancer has evolved from lower grade tumours, it is considered as

secondary GBM, while primary GBM is discovered at diagnosis (*de novo*) as advanced cancer (Parsons et al. 2008a).

Currently, the standard treatment for GBM patients is surgical removal of the tumours in addition to radiotherapy and chemotherapy with alkylating agent temozolomide (Parsons et al. 2008). Yet, the median survival time of GBM patients is 12 months (Barbus et al. 2011a). However, 3%-5% of glioma patients survive for more than 36 months (Krex et al. 2007; Liu et al. 2010) and these patients are called Long Term Survivors (LTS). These patients are mostly young (<50 years old) and have a good clinical performance. Generally, glioblastomas from LTS are characterised by promoter hypermethylation of the *MGMT* gene and mutations of the *IDH* gene (Barbus et al. 2011b).

1.5.4 Genetic alterations in glioma

Like other cancers, gliomas are developed due to genetic alterations. Recent publications have identified several genes that are involved in glioma development and progression. The presence or absence of these chromosomal and genetic changes depends on the progression of the tumour toward higher grades (Yan et al. 2009b). For instance, *IDH1* and *IDH2* mutations are more common in secondary glioblastoma, WHO grade II and III diffuse astrocytoma, WHO grade II oligodendroglioma and WHO grade II oligoastrocytic glioma (Losman and Kaelin 2013; Riemenschneider et al. 2010b). Mutation of *TP53* is mostly found in WHO grade II diffuse astrocytoma, while WHO grade II oligodendroglioma is usually characterised by 1p/19q deletion. Additional genetic changes can been found in these lower grade gliomas during the progression towards higher grades, such as mutations in *RB1*, methylation of *CDKN2A/B* and deletion of the

50

short arm of chromosome 9 during the progression from WHO grade II gliomas as well as loss of 10q and loss of expression of *DCC* through the progression from WHO grade III anaplastic astrocytoma to secondary glioblastoma (WHO grade IV). Primary glioblastomas (WHO grade IV), which are the most aggressive type, carry several genetic, epigenetic and loss/gain chromosomal alterations. These alterations include mutations in the genes involved in the receptor tyrosine kinase pathway (*EGFR, ERBB2, PTEN, PIK3R1* and *PI3KCA*), the p53 pathway (*TP53, MDM2* and *MDM4*) and the Rb1 (*CDKN2A/B, CDK4/6,* and *RB1*) as well as 19q and 20q gain, monosomy 10p and trisomy 7p. The most frequent alterations in WHO grade I pilocytic astrocytoma are mutations and fusion/duplication of BRAF in addition to gain on chromosomes 7q and 5q. Figure 1.14 summarises the most frequent genetic alterations in all glioma grades.

Figure 1.15 Summary of the most frequent genetic alterations in all glioma grades

This figure was taken from Riemenschneider et al. (2010b)

1.5.4.1 IDH mutations

IDHs are key enzymes in the Krebs cycle, which catalyse the oxidative carboxylation of isocitrate to α-ketoglutarate (α-KG) and reduce NAD+ (or NADP+) to NADH (or NADPH). α-KG is important for the demethylation of 5-methylcytosine (5mC) and histone that are essential for epigenetic regulations in the cells. There are 3 IDH genes that are expressed in various tissues including the brain. These are IDH1 which is found in cells' cytoplasm, IDH2 sequence is 70% identical to IDH1 (Xu et al. 2004) and is localized in cells' mitochondria and IDH3 is also found in the mitochondria and is structurally unrelated to IDH1 and IDH2 (Nichols et al. 1995).

The discovery of *IDH* gene mutations in brain tumours is an example of NGS success in the identification of disease-associated SNVs. To date, the identified mutations in gliomas are found in *IDH1* or *IDH2* genes. No IDH3 mutations have been discovered in gliomas (Krell et al. 2011).

When *IDH* is mutated, the conversion of isocitrate to α-KG drives the synthesis of 2-hydroxyglutarate (2-HG), which acts as an inhibitor of the α-KG-dependant demethylase process, leading to hypermethylation (Figure 1.15).

Isocitrate

IDH1^{wt}

NAD+

NADH, CO$_2$

α-Ketoglutarate

IDH1^{R132H}

NAD+

NADH, CO$_2$

2-Hydroxyglutarate

Promoter hypermethylation
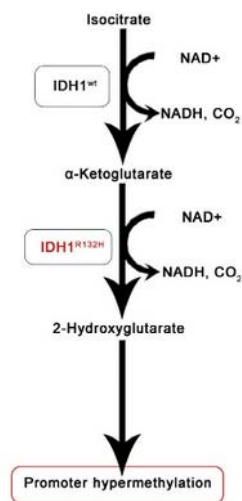
Figure 1.15 The enzymatic activity of the wild-type IDH and mutated IDH genes

When *IDH* is mutated, the conversion of isocitrate to α-KG drives the synthesis of 2-hydroxyglutarate (2-HG), which acts as an inhibitor of the α-KG-dependant demethylase process, leading to hypermethylation. The figure is taken from Tew and Fisher (2014).

IDH1 and *IDH2* mutations are somatic missense changes which affect the arginine residue at codon R132 in *IDH1* and codon R172 in *IDH2*, within the substrate binding site of the enzyme. *IDH1* mutations are widely found in glioma tumours and mostly change arginine to histidine (Mellai et al. 2013). The most commonly reported missense change in glioma tumours is p.R132H (92.4%) followed by the less common mutations; p.R132C (3.2%), p.R132G (2.1%), p.R132S (1.6%) and p.132L (0.6%) (Mellai et al. 2013). One rare mutation (p.R132V) was originally identified as a somatic *IDH1* mutation (Balss et al. 2008) until it was reported as a SNP according to the dbSNP database of NCBI. On the other hand, mutations in *IDH2* are less common (3-5%) in glioma tumours and are found in glioma cases without *IDH1* mutation. The missense mutation p.R172K accounts for 60% of the cases, while p.R172M, p.R172W and p.R172S are less common (27%, 11% and 2% respectively). Single studies have reported other new IDH mutations such as the missense mutation (p.R100Q) of *IDH1* gene in grade III oligodendrogliomas and grade II astrocytoma (Pusch et al. 2011), the missense mutations (p.R49C and pG.97D) of *IDH1* in paediatric glioblastoma (Paugh et al. 2010) and a single nonsense mutation of *IDH2* gene identified in a brain biopsy sample (Horbinski et al. 2010) (Figure 1.16).
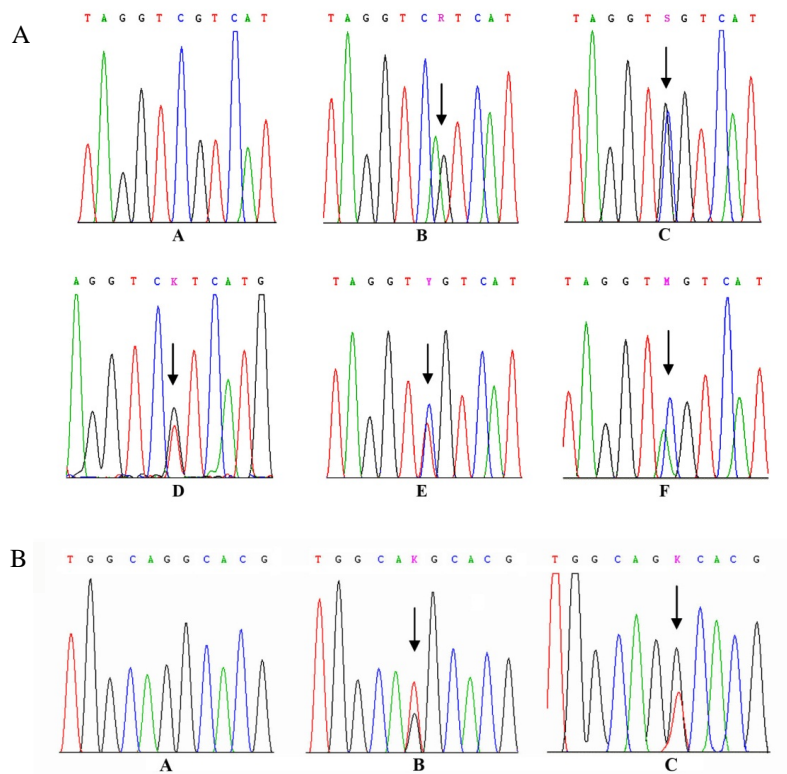
Figure 1.16 Electropherograms of the most common IDH1 and IDH2 mutations

A) Electropherograms of representative mutations in the IDH1 gene. A – DH1 wild-type; B – c.G395A (p.R132H); C – c.C394G (p.R132G); D – c.G395T (p.R132L); E – c.C394T (p.R132C); F – c.C394A (p.R132S). B) Electropherograms of representative mutations in the IDH2 gene. A – IDH2 wild-type; B – c.G515T (p.R172M); C – c.G516T (p.R172S). This figure was taken from Mellai et al. (2013).

The majority of *IDH1* or *IDH2* mutations were identified in adult secondary glioblastoma (~80%), while mutations in primary glioblastoma are less frequent and account for approximately <10% (Losman and Kaelin 2013). Also, other types of adult brain tumours showed high frequency of *IDH1*/*IDH2* mutations, such as grade II diffuse astrocytoma (75%), grade III Anaplastic astrocytoma (58%) and grade III anaplastic oligodendroglioma (62%) (Yan et al. 2009b). In paediatric glioblastoma cases, *IDH1*/*IDH2* mutations were rare (<16%) but showed higher frequency in children >14 years old (35%) (Pollack et al. 2011).

Many studies have found an association between IDH mutations and glioma patient's survival. Patients with *IDH* mutated glioblastoma and other WHO grade III gliomas have shown significant longer overall survival and better progression-free survival rates (Sanson et al. 2009; Weller et al. 2009). The correlation between IDH mutations and survival in WHO lower grade gliomas is not clear. Many studies have reported the prognostic role of IDH mutations in WHO grade II gliomas (Sanson et al. 2009; Hartmann et al. 2011; Mukasa et al. 2012), however, other studies using a larger set of WHO lower grade glioma samples have shown no prognostic correlation of IDH mutations (Ahmadi et al. 2012; Juratli et al. 2012; Kim et al. 2010). Also, patients with WHO low grade glioma that do not harbour IDH mutations have shown poor prognosis (Metellus et al. 2010). The findings of the favourable outcome in glioma patients with IDH mutations raised the consideration of using IDH mutation analysis as a diagnostic marker in clinics.

1.5.5 Epigenetic alterations in glioma

1.5.5.1 DNA methylation

Several studies have reported changes in DNA hypermethylation and hypomethylation in glioma. For instance, global hypomethylation has been identified in primary glioblastoma in *Sat2* (satellite 2), *D4Z4* and interspersed *Alu* elements. The hypomethylation event in these genes was associated with the copy number changes of the adjacent euchromatin, suggesting that hypomethylation is responsible for genomic instability (Cadieux et al. 2006). Also, promoter demethylation of *MAGEA1* has been found to increase the activity of the gene which is involved in transcriptional regulation of the histone deactelyase HDAC1 (Cadieux et al. 2006). Interestingly, a recent study has reported that overexpression of *DNMT1* by histone modification and overexpression of *DNMT3B* by promoter hypomethylation in glioma resulted in the re-expression of a number of tumour suppressor genes by inhibiting these DNA methyltransferases (Rajendran et al. 2011).

On the other hand, DNA hypermethylation events are common in glioma which usually affects tumour suppressor genes and genes involved in different cellular processes. For instance, promoter hypermethylation of cell cycle regulatory genes *p14*, *p16* and *RB1*, the DNA repair gene *MGMT*, cell migration genes *TIMP3* and *EMB3* and the WNT signalling pathway inhibitor *DKK1* are more frequent in secondary glioblastomas compared to primary glioblastomas (Götze et al. 2010; M. Nakamura, Watanabe, et al. 2001; M. Nakamura, Yonekawa, et al. 2001; Mitsutoshi Nakamura et al. 2001; Kunitz et al. 2007; Mitsutoshi Nakamura et al. 2005). On the other hand, down regulation by promoter hypermethylation of *SFRP1*, *SFRP2*, the tumour suppressor gene *NDRG2* and the negative regulator of WNT signalling pathway *NKD2* are found in 40% of primary

58

glioblastomas (Götze et al. 2010; Tepel et al. 2008; Malzkorn et al. 2011). Other important regulatory genes such as the tumour suppressor genes *PTEN*, *RASSF1*, *CTMP* and *ECRG4* (Baeza et al. 2003; Horiguchi et al. 2003; Knobbe et al. 2004; Götze et al. 2009); cell apoptosis regulatory gene, *PYCARD* (Stone et al. 2004) and cell migration and invasion genes, *CDH1* and *SOCS3* (Lindemann et al. 2011; D'Urso et al. 2011) have also been reported to be hypermethylated in primary glioblastomas. Moreover, the phenomena of hypermethylation of several genes in certain types and grades of glioma are known as glioma CpG Island Methylator Phenotype (G-CIMP), which will be explained later in section 1.5.5.3.

1.5.5.2 Hypermethylation of MGMT

Hypermethylation of O^6-*methylguanine-DNA methyltransferase (MGMT)* is seen in 50% of glioma patients and has become a significant marker for prognosis (Krex et al. 2007). The association between *MGMT* promoter hypermethylation and glioma patients' response to chemotherapies, such as carmustine and temozolomide, radiotherapy or both has been reported in many studies (Mitsutoshi Nakamura et al. 2001; Hegi et al. 2005; Herrlinger et al. 2006). In Hegi et al.'s study on primary glioblastoma patients who were treated with radiotherapy combined with a temozolomide agent, the results showed that patients with *MGMT* promoter hypermethylation tumours have survived significantly longer than patients with unmethylated *MGMT* tumours.

(Hegi et al. 2005).

MGMT is a DNA repair enzyme which counteracts DNA damage caused by the temozolomide agent during chemotherapy by removing the alkyl group from the O^6 of

guanine residue to induce apoptosis and, thereby, poor response to treatment (Riemenschneider et al. 2010b). *MGMT* promoter hypermethylation decreases expression of the protein, which leads to the inability to counteract the DNA damage and results in a better response to chemotherapy and better prognosis in glioma patients (Hegi et al. 2005; Malzkorn et al. 2011).

Therefore, testing for *MGMT* promoter hypermethylation is increasingly performed in clinical trials for high grade glioma patients in addition to some routine diagnostic settings (Riemenschneider et al. 2010b). Nowadays, some trials are selective to perform studies with either *MGMT* hypermethylation or unmethylated tumours by analysing the methylation status of the *MGMT* promoter (Malzkorn et al. 2011). Furthermore, in anaplastic gliomas, *MGMT* promoter hypermethylation was associated with better prognosis, not only in temozolomide-treated patients but also in patients just receiving radiotherapy (Weller et al. 2010). *MGMT* methylation is less frequent in ependymal glioma, medullablastoma and anaplastic meningioma (Ebinger et al. 2004; Brokinkel et al. 2010; Buccoliero et al. 2008).

1.5.5.3 G-CIMP

Many genome-wide methylation studies have identified CpG island methylator phenotypes (CIMP) in cancers. Interestingly, CIMP was first identified in colorectal cancer in 1999 (M. Toyota et al. 1999) to define a distinct methylation profile subtype. CIMP is defined as an overall increase and concordant DNA hypermethylation in many CpG loci and has been identified in other types of cancer including AML (Figueroa et al. 2010) and clear cell renal carcinoma (Arai et al. 2012). Remarkably, in glioma, G-CIMP

is strongly associated with *IDH* mutation and is more frequent in lower grade glioma (Noushmehr et al. 2010). Later, the study by Turcan et al. found that the *IDH1* mutations are the lead cause for CIMP phenotype formation (Turcan et al. 2013). Patients with G-CIMP+ are characterised by younger age and better survival in comparison to G-CIMP- patients (Christensen et al. 2011; Turcan et al. 2012; Noushmehr et al. 2010). In rare events, CIMP can occur in IDH wild-type gliomas and the reason behind the formation of the phenotype is still unknown (Brennan et al. 2013; Hill et al. 2014).

Also, many studies have been conducted to discover the association between G-CIMP, *IDH* mutations and *MGMT* promoter hypermethylation. The G-CIMP glioma subgroup with *IDH* mutations is associated with longer survival and glioma patients with MGMT promoter hypermethylation are also characterised by better prognosis. However, no direct association has been reported between *MGMT* hypermethylation and G-CIMP status (Bady et al. 2012; van den Bent et al. 2011).

1.5.5.4 Histone modifications

Genome-wide studies have reported many alterations in genes involved in histone modification in gliomas. In a study by Parsons et al. that used large-scale sequencing analysis of glioblastoma, several genetic mutations have been detected in genes involved in histone demethylation (*JMJD1A*, *JMJD1B*) and histone methyltrasnferase (*SET7*, *SETDB2*, *MLL3*, and *MLL4*) in addition to *MBD1* gene (Methyl-CpG Binding Domain Protein 1) (Parsons et al. 2008b). Also, mutations in *MLL2* and *MLL3* were identified in a few cases of medulloblastoma (16%) (Parsons et al. 2011). In neuroblastoma, CpG island methylation was observed in *NSD1* (Nuclear Receptor Binding SET gene) that encodes

histone methyltransferase which leads to inactivation of the gene (Berdasco et al. 2009). Another study using immunohistochemistry expression analysis of histone H3 and H4 modifications in a large set of glioma cases identified 10 distinct prognostic subgroups based on the expression of the histone modifications H3K9Ac, H3K4diMe, H3K18Ac and H4K20triMe (Liu et al. 2010).

1.5.6 Brain metastasis from distant organ

Metastasis to the brain is the most common form of CNS tumour which usually indicates a patient's poor prognosis. The incidence of brain metastasis is increasing and outnumbers primary brain cancer by 10 times (Patchell 2003). Mostly, tumour cells reach the brain by blood circulation from different sites, most commonly from the breast, lung, melanoma and liver (Gavrilovic and Posner 2005) (Figure 1.17). Most breast cancer patients develop brain metastasis after metastasis formation in the lung, liver or bone (Issa et al. 2002).

Metastasis in the brain usually forms in parenchyma tissue or leptomeninges. Leptomeningeal metastasis develops in the meninges (membranes surrounding the brain) which is the most common site for metastasis from a breast tumour (Palmieri et al. 2007). The brain is a special organ surrounded by the skull and protected by the blood-brain-barrier (BBB). The BBB is a diffusion barrier that strictly selects and regulates the access of cells and other molecules to the brain. However, the BBB is responsible for poor drug delivery to the brain tumour site and limits drug efficacy, and as a consequence, drugs must be specifically designed to be able to cross the barrier (Palmieri et al. 2007). In

addition, even with the surgical removal of the tumour and radiotherapy, patients with brain metastasis only survive 6-9 months.

In 1989, the hypothesis of "seed-and-soil" was proposed by Stephen Paget to suggest that metastasis to the brain does not randomly occur (Paget 1989) but tumour cells (the seeds) must be attracted to the new microenvironment of the metastatic organ (the soil). The hypothesis proposes that tumours are composed of heterogeneous sub-populations of cells that can disseminate, multiply, and selectively invade the brain which can be later colonized at the site depending on the interaction between tumour cells "seeds" and host microenvironment "soil" (Rahmathulla, Toms, and Weil 2012). Metastasis formation involves two stages: (i) cellular migrations, where cells start to invade the surrounding tissues, disseminate and adhere to new tissues and (ii) colonization of the tumour cells in the new site. The ability of tumour cells to go through all the previous processes depends on their ability to disturb the E-cadherin-Catenin molecular complex function and go through epithelial mesenchymal transition (Svokos, Salhia, and Toms 2014). Underlying these stages are various of genetic, epigenetic changes responsible for metastasis formation and development known as "metastasis cascade" (Rahmathulla, Toms, and Weil 2012). Therefore, further molecular biology and genetic studies have been conducted recently to increase knowledge and understanding of the underlying processes involved in tumour metastasis to distant organs.
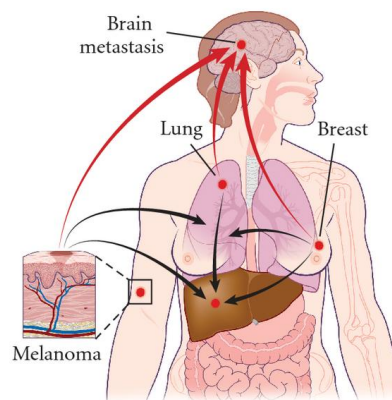
Figure 1.17 Illustration of the most common primary sites of brain metastasis

Tumour cells from primary sites like the breast, lung and skin (melanoma) metastasise to the brain via blood circulation (red arrows) or to adjacent sites such as the liver and lung (black arrows). This figure was taken from Rahmathulla, Toms and Weil (2012).

1.6 Aim of the research

Now a day, with the availability of the high-throughput technologies, such as NGS and microarrays, cancer researches has improved our understanding of cancer biology and the molecular mechanisms behind cancer development and progression. The aim of this thesis is to identify genetic and epigenetic changes in CNS tumours. The main focus will be on GBM tumours, brain metastasis from breast cancer and sporadic vestibular schwannoma, using the following approaches.

- Using Illumina Infinium HumanMethylation450 BeadChip array to;

  (i) Characterise the methylation profile of STS and LTS glioblastoma tumours and to find the differences in methylation between the two groups. Also, to identify epigenetic markers associated with poor prognosis.

  (ii) Identify potential epigenetic markers for brain metastasis formation from breast cancer, find the association of promoter hypermethylation of the genes with patient survival and compare the methylation event of these genes in brain metastasis tumours from different primary organs, such as lung cancer and melanoma.

- Using WES technology, a NGS technology that covers the coding regions of the genome, and different bioinformatic tools and international SNP and SNV databases on non-NF2 sporadic vestibular schwannoma tumours and their matching normal samples to identify somatic causative variants or variants involved in the development of schwannomas.

Chapter Two: Materials and Methods

2.1 DNA and RNA Samples

All samples used in the study were anonymised. The collection and use of the samples was ethically approved by the review boards of the institutions and local research committees. All studies were conducted according to the principles expressed in the Declaration of Helsinki.

2.1.1 DNA from Patients' Samples

2.1.1.1 GBM DNA samples

DNA from 35 glioblastoma patients (grade IV) and from 4 non-cancerous brain samples were kindly provided by our collaborator, Prof. Dietmar Krex (Klinik und Poliklinik fur Neurochirurgie, Universitatsklinikum Carl Gustav Carus, Technische Universitat Dresden, Germany).

2.1.1.2 Brain metastasis DNA samples

Brain metastasis DNA from breast tumours (n=15), melanoma (n=20) and lung cancer (n=20) were used in this thesis and provided by our collaborator, Prof. Dietmar Krex (Klinik und Poliklinik fur Neurochirurgie, Universitatsklinikum Carl Gustav Carus, Technische Universitat Dresden, Germany).

2.1.1.3 Breast tumours and corresponding brain metastasis DNA samples

DNA from 11 breast tumours and corresponding brain metastasis samples (matched pairs) were generously provided by Dr. Mark Morris (School of Biology, Chemistry and Forensic Science, University of Wolverhampton). RNA from brain metastasis samples was also provided by Dr. Mark Morris.

66

2.1.1.4 Schwannoma DNA samples

DNA from 59 schwannoma tumour samples and their corresponding normal blood DNA samples, plus 2 peripheral nerve DNA samples (normal, non-cancerous), were provided by our collaborator, Dr. Juan A. Rey (Unidad de Investigacion, IdiPAZ Hospital, Universitario La Paz Paseo de la Castellana, Madrid, Spain).

2.1.2 Cell lines

2.1.2.1 Glioma cell lines

Glioma cell lines (A172, U343, HS683, H4 and T17) were used in this study and were provided by Prof. Dietmar Krex (Klinik und Poliklinik fur Neurochirurgie, Universitatsklinikum Carl Gustav Carus, Technische Universitat Dresden, Germany).

2.2 Nucleic acid extraction

2.2.1 DNA extraction

DNA was extracted from samples using the Qiagen DNeasy Blood and Tissue Kit (Cat. No.: 69506) and instructions were followed according to the "Total DNA extraction from animal blood or cells" protocol. Briefly, pellets were re-suspended in 200µl buffer AL, then 20µl proteinase K was added and mixed by vortexing. Samples were then incubated for 10 minutes at 56°C. After that, samples were washed using 200µl absolute ethanol and vortexed until the solution was homogenised. Using DNeasy Mini spin-column, samples were centrifuged at room temperature for 1 minute at 20,000 x g (14,000 rpm) and the flow-through solution was discarded. After that, 500µl buffer AW1 was added to the column and centrifuged for 1 minute at 20,000 x g (14,000 rpm) at room temperature

and the flow-through solution was discarded. The spin-column was washed again using 200µl buffer AW2 and centrifuged for 3 minutes at 20,000 x g (14,000 rpm) at room temperature. This final flow-through was discarded along with the collection tube and was replaced by a new 1.5ml microcentrifuge tube. DNA was eluted from the spin-column's membrane by adding 200µl buffer AE to the membrane and incubated at room temperature for 1 minute followed by centrifugation for 1 minute at 20,000 x g (14,000 rpm). Then, DNA was stored in the freezer at -20°C.

2.2.2 RNA extraction

RNA was extracted from cell lines using the EZ-RNA Total RNA Isolation kit (Biological Industries, Cat. No.: 20-400-100) according to the manufacturer's instructions. Briefly, 0.5ml denaturing solution (guanidine thiocyanate) was added to each sample and mixed by repeated pipetting and then incubated for 5 minutes at room temperature. Then 0.5ml extraction solution (phenol and chloroform) was added and mixed vigorously followed by another incubation for 10 minutes at room temperature, then centrifugation at 12,000 x g at 4°C for 15 minutes. After that, the RNA precipitation step was carried out by transferring the upper phase to another tube. It was mixed with 0.5ml isopropanol and stored for 10 minutes at room temperature then centrifuged for 8 minutes at 12,000 x g (4°C). The supernatant was then discarded and RNA was washed using 75% (v/v) ethanol. RNA was then centrifuged at 7,500 x g for 5 minutes (4°C) and the supernatant was carefully removed. RNA was allowed to air-dry prior to the addition of 100µl 0.5% SDS solution and was incubated for 15 minutes at 55°C for RNA solubilisation. The samples were then stored at -80°C in the freezer until they were used.

2.2.3 Quantification of DNA and RNA concentration using Nanodrop Spectrophotometer

DNA and RNA were quantified using the NanoDrop 1000 Spectrophotometer (NanoDrop Technology, USA). The blanking procedure was performed using 2µl of the same elution solution used to elute DNA/RNA during preparation (buffer AE or RNase free water). After that, 1-2µl of nucleic acid was added to the NanoDrop spectrophotometer and absorbance was read at 260nm and 280nm. The output readings of the NanoDrop consisted of nucleic acid concentration in ng/µl and nucleic acid purity. Purity was measured using the ratio of absorbance 260/280. DNA with a purity ratio of ≈1.8 was considered pure, while a ratio of ≈2.0 was generally accepted as pure for RNA.

2.3 Bisulfite modification of DNA

DNA conversion was performed according to the Qiagen EpiTect Bisulfite Kit (Qiagen, Cat.No: 59104) as follows; aliquots of Bisulfite Mix were dissolved by adding 800µl RNase-free water to each aliquot and vortexed until completely dissolved. Bisulfite reactions were prepared in 200µl PCR tubes by the addition of 500ng-1µg DNA and RNase-free water to make a total volume of 20µl. This was followed by the addition of 85µl dissolved Bisulfite mix and 35µl DNA protect buffer. Then, the samples were mixed by pipetting until the solution turned into blue. After that, the bisulfite DNA conversion reaction was carried out using the thermal cycler according to the following conditions (Table 2.1).

Step	Duration	Temperature
Denaturation	5 minutes	95°C
Incubation	25 minutes	60°C
Denaturation	5 minutes	95°C
Incubation	1 hour 25 minutes	60°C
Denaturation	5 minutes	95°C
Incubation	2 hours 55 minutes	60°C
Hold	Indefinite	20°C

Table 2.1 Thermal cycler conditions for DNA bisulfite conversion

The above conditions are according to the Qiagen EpiTect Bisulfite kit.

After completion of the bisulfite conversion reaction, samples were briefly centrifuged and transferred to new 1.5ml microfuge tubes. A freshly prepared Buffer BL containing 10µg/ml carrier RNA was added to each sample and mixed by vortex. The samples were then transferred to an Epitect spin-column and spun down for 1 minute at 20,000 x g (14,000 rpm). The flow-through solution was discarded and the spin-column membrane was washed with 500µl buffer BW. Samples were centrifuged again at maximum speed for 1 minute and the flow-through was discarded. After that, the de-sulfonation step was carried out by adding 500µl buffer BD to each spin-column and the samples were incubated at room temperature for 15 minutes. The samples were then centrifuged for 1 minute at maximum speed and flow-through was discarded again. The samples were washed twice with 500µl buffer BW and centrifuged at 20,000 x g (14,000 rpm) for 1 minute. After the second wash, the spin-column of each sample was placed into a new 1.5ml microfuge tube and 20µl of buffer EB was added to the centre of the membrane to elute the purified DNA. Samples were incubated for 2 minutes prior to the final centrifugation step to increase the yield of DNA in the eluate. Bisulfite modified DNA was stored at -20°C.

2.4 Analysis of DNA methylation

Several methods were used in this thesis to analyse DNA methylation status.

2.4.1 Methylation/Unmethylated specific PCR

Methylation specific PCR (MSP) and unmethylated specific PCR (USP) reactions were performed at the same time for the same gene. PCR reaction contained 1x of 10X

reaction buffer with 20mM $MgCl_2$ (500mM Tris/HCL, 100mM KCl, 50mM ammonium sulphate, 20mM $MgCl_2$, pH8.3/25°C), 1x 5X GC-RICH solution, 2.5mM PCR grade Nucleotide mix, 0.8μM of forward primer, 0.8μM of reverse primer and 1.0u FastStart Taq DNA polymerase (Roche, Cat. No. 12158264001). Deionized distilled water was used to make the reaction up to 50μl. In each PCR reaction, 3μl of bisulfite modified DNA was added as a template and then PCR was performed using the standard PCR programme listed in Table 2.2. Primer sequences and annealing temperatures are shown in Appendix 7.1.2.

Temperature	Duration	Number of cycles
95 °C	5 minutes	1 cycle
95 °C	1 minute	
Annealing	1 minute	35 cycles
72 °C	2 minutes	
72 °C	10 minutes	1 cycle
10 °C	Indefinite	1 cycle

Table 2.2 Standard MSP/USP PCR programme

Table showing the PCR conditions of a standard PCR programme used for MSP and USP.

2.4.2 Agarose gel electrophoresis and UV light visualisation

In this study, 2% (w/v) agarose gel (Bioline, Cat. No. E8751) was used to visualise PCR products. Agarose gel was made by mixing 4grams of agarose and 200ml of 1X Tris-Borate-EDTA (TBE) that was diluted from 10X TBE (National Diagnostics, Cat. No. EC-860). The mix was heated in the microwave until it had completely dissolved. After cooling down, 2µl of 10mg/ml ethidium bromide (EtBr) was added to the gel and then poured in the tray to set at room temperature. Prior to loading, the samples were mixed with 5% (v/v) loading buffer (50% glycerol, 48.5% distilled water, 1% EDTA and 0.5% Orange G) and loaded into the gel electrophoresis to be migrated at ≈180 volts for 15 minutes to be visualised later under UV light.

2.4.3 Combined Bisulfite Restriction Analysis (CoBRA)

2.4.3.1 Primer design

All of the CoBRA primers were designed manually in this study unless they were pre-designed and listed in other published studies. The UCSC genome browser was used to identify the region of interest and primers were designed around this area. For each gene or probe, primers were designed for semi-nested CoBRA PCR (set of 3 primers) unless the PCR product was poor or the covered region was >500bp, in this case fully-nested CoBRA PCR primers were designed. To design the CoBRA primers, several aspects were considered (whenever possible) to ensure the maximum effectiveness of the primers; (i) the 3′ end of the primer was designed to end on the CAC or ACC of the genomic site to ensure high specificity to bisulfite modified DNA in addition to a string of Cs within the primer, (ii) to avoid biasing the PCR towards methylated or

unmethylated DNA, if CpG dinucleotides were within the primer sequence, a Y (C or T) was incorporated to replace a C in the forward primer and an R (G or A) to replace a G in the reverse primer, (iii) a restriction enzyme cutting site with the primer sequences was strongly avoided. Primers were designed to be around 30bp long (±5bp) and to produce 300-450bp PCR products that contain at least one restriction site within the region. Annealing temperature was calculated using the following equation and CoBRA PCR was performed at 2$^\text{o}$C below the calculated annealing temperature. All CoBRA primer sequences used in this study and their annealing temperature are listed in Appendix 7.1.1.

$$\left[\frac{41 * GC}{n}\right] + \left[64.9 - \left(\frac{600}{n}\right)\right]$$

GC = number of C and G nucleotides (Y and R are considered as 0.5)

n = total number of nucleotides

2.4.3.2 CoBRA PCR

All PCR reactions were performed using either semi- or fully-nested primer sets. Two rounds of PCR were carried out for each CoBRA assay using FastStart Taq DNA polymerase kits (Roche, Cat.No.12158264001). Each PCR reaction was made up to the total volume of 25µl (for primary reaction) or 50µl (for secondary reaction) and contained 1x of 10X reaction buffer with 20mM MgCl$_2$ (500mM Tris/HCL, 100mM KCl, 50mM ammonium sulphate, 20mM MgCl$_2$, pH8.3/25$^\text{o}$C), 1x 5X GC-RICH solution, 2.5mM PCR grade Nucleotide mix, 0.8µM of each of the forward and reverse primer,

0.5u (primary reaction) or 1.0u (secondary reaction) FastStart Taq DNA polymerase and distilled water. The bisulfite modified DNA template of 2µl (0.04µg) was added to the primary reaction as starting material and 5µl of the primary PCR product was used in the secondary reaction as starting material. CoBRA PCR's reaction conditions were either a standard PCR or a touchdown PCR programme, as detailed in Table 2.3.

A)

Temperature	Duration	Primary CoBRA reaction	Secondary CoBRA reaction
95 °C	5 minutes	1 cycle	1 cycle
95 °C	1 minute	35 cycles	40 cycles
Annealing	1 minute		
72 °C	2 minutes		
72 °C	10 minutes	1 cycle	1 cycle
10 °C	Indefinite	1 cycle	1 cycle

B)

Temperature	Duration	Primary CoBRA reaction	Secondary CoBRA reaction
95 °C	5 minutes	1 cycle	1 cycle
95 °C	45 seconds	5 cycles, decreasing 2°C in each cycle	5 cycles, decreasing 2°C in each cycle
Annealing	45 seconds		
72 °C	45 seconds		
95 °C	45 seconds	35 cycles	40 cycles
Annealing	45 seconds		
72 °C	45 seconds		
72 °C	10 minutes	1 cycle	1 cycle
10 °C	Indefinite	1 cycle	1 cycle

Table 2.3 Standard and touchdown CoBRA PCR programmes

A) Standard CoBRA PCR programme for primary and secondary reactions. B)
Touchdown CoBRA PCR programme for primary and secondary reactions.

2.4.3.3 Restriction of enzyme digest

After visualisation of the PCR products using gel electrophoresis (section 2.4.2) and depending on the intensity of the bands, a suitable volume of the PCR product (5µl-20µl) was digested with one of the restriction enzymes listed in Table 2.4.

The BstUI restriction enzyme (Fermentas, Cat. No. ER0921) is commonly used to analyse the methylation status of CpG islands due to the high frequency of its cutting site in the islands. The digestion reaction was carried out using 1u BstUI, 1x 10X buffer R, an appropriate amount of the PCR product and distilled water to make it up to the final volume of 25µl. Then, samples were incubated overnight at 37°C.

Samples were digested with the HpyCH4IV restriction enzyme (New England Biolabs, Cat. No. R0619S) and TaqαI (New England Biolabs, Cat. No. R0149L) and were incubated at 37 ^{o}C and 65 ^{o}C respectively for two hours. The digestion reaction was carried out with 20u of the enzyme, 0.01µl BSA, 1x 10X NEBuffer 1 (for HpyCH4IV) or NEBuffer 4 (for TaqαI), a suitable amount of the PCR product and made up with distilled water to the volume of 25µl.

Digested and undigested samples were loaded into gel electrophoresis (section 2.4.2) and run simultaneously to be visualised under the UV light.

Enzyme	Cutting site	Temperature	Incubation period
BstUI	CG^CG	37 °C	Overnight
TaqαI	TC^GA	65 °C	2 hours
HpyCH4IV	AC^GT	37 °C	2 hours

Table 2.4 Restriction enzymes used in this study

A table showing the restriction enzymes used for methylation analysis and their properties.

2.4.4 Bisulfite sequencing

2.4.4.1 Gel extraction

Prior to sequencing, the PCR products were initially purified from agarose gel using the Gel Extraction kit (Qiagen, Cat. No. 28704). Bands were excised carefully using a scalpel and transferred to a labelled 1.5ml microcentrifuge tube. According to the manufacturer's protocol, gel slices were weighed and 3x volumes of Buffer QG were added to 1 volume of gel. Tubes were placed in a block heater and incubated for 10 minutes at 50°C with vortexing every 3-4 minutes. After that, 1x gel volume of 100% Isopropyl alcohol was added to the samples and mixed vigorously. The solution was then transferred to a QIAquick spin-column and centrifuged for 1 minute at high speed (20,000 x g) to allow DNA to bind to the membrane. The flow-through was discarded and another 0.5ml Buffer QG was added to the column, then a second centrifugation step was applied as earlier. Again, the flow-through was discarded and 0.75ml Buffer PE was added to wash the column. The samples were centrifuged again and the flow-through was discarded. A QIAquick spin-column of each sample was then placed in a new 1.5ml microcentrifuge tube and 30µl of deionised distilled water was added to the centre of the membrane to elute DNA. The samples were incubated for 1 minute at room temperature prior to centrifugation at high speed for 1 minute. The purified PCR products were stored at -20°C until they were used.

2.4.4.2 Plasmid vector ligation

Purified PCR products were introduced into bacterial cells for cloning using the pGEM-T Easy vector system (Promega, Cat. No. A1360). The vector ligation reaction was prepared by gently mixing 0.5µl pGEM vector, 1µl T4 DNA ligase and 5µl 2X ligation buffer with 3.5µl purified PCR product. Samples were incubated overnight at 4°C to allow the ligation process to occur.

2.4.4.3 LB media preparation

Preparation of LB agar plates was carried out by mixing 35g/l LB powder agar (Sigma Aldrich, Cat. No. L2897) with distilled water before autoclaving. After cooling down the mixture, ampicillin drug (Sigma Aldrich, Cat. No. H0166) with a concentration of 100µg/ml was added and the solution was poured into petri dishes to set before being stored at 4°C.

2.4.4.4 Transformation of bacterial cells

Initially, α-select silver efficiency chemically component cells (Bioline, Cat. No. BIO-85026) were carefully defrosted on ice and 40µl aliquoted into microcentrifuge tubes. In each aliquot, 4µl of pGEM-PCR product ligation was added and gently mixed without damaging the cells. Each tube was incubated on ice for 20 minutes and then heat shocked at 42°C for 30 seconds. Then, 500µl SOC media (Invitrogen; Cat. No. 15544034) was added and cells were incubated for 1 hour at 37°C while being shaken. During the incubation period, 20µl X-gal (50mg/ml) (Bioline; Cat. No. BIO37035) was added to

each media plate and later, 200µl of each cultured cell was spread on the plate and incubated for 18 hours at 37°C.

2.4.4.5 Single colony PCR

After the incubation period, only white colonies that contained the insert were selected and transferred to 20µl deionised distilled water and heated to 95°C for 5 minutes to allow cell membrane rupture. Only the white colonies were selected due to the disruption of the β- galactosidase gene within the plasmid and therefore, white colonies were expected to have the insert. On the other hand, blue colonies were produced due to the undisrupted plasmids which allow the β- galactosidase enzyme to utilise the X-gal producing blue product. After the incubation period, 7-10µl of the solution was used as a template for a PCR reaction that was made up to the total volume of 30µl which contained 1x of 10X reaction buffer with 20mM $MgCl_2$ (500mM Tris/HCL, 100mM KCl, 50mM ammonium sulphate, 20mM $MgCl_2$, pH8.3/25°C), 1x 5X GC-RICH solution, 2.5mM PCR grade Nucleotide mix, 0.8µM of each of the forward and reverse primer, 0.5u FastStart Taq DNA polymerase and distilled water. The forward and reverse primer sequences are: F primer: 5'-TAATACGACTCACTATAGGG-3; and R primer: 5'-ACACTATAGAATACTCAAGC-3'.

2.4.4.6 ExoSAP for PCR product purification

The PCR products were cleaned up prior to the sequencing reaction to ensure the removal of dNTPs and primers excess. A reaction was made up with 1U of FastAP, 0.01U of 1x 10x FastAP buffer and ExoI each, and 10µl PCR product. The samples were then

incubated in a thermal cycler for 30 minutes at 37°C and then for 20 minutes at 80°C to inactivate the ExoI enzyme.

2.4.4.7 Sequencing reaction

The sequencing reaction was performed using the Big Dye Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems; Cat. No. 4336917). For each sequencing reaction, 5µl of the cleaned PCR product (see section 2.4.4.6) was mixed with 20pmol of forward (or reverse) primer, 0.5-0.75µl Big Dye, 2µl 5X sequencing buffer and nuclease free water. Then, the sequencing reaction was carried out using the programme in Table 2.5 and was later stored in the dark at -20°C until it was ready for alcohol precipitation.

Temperature	Duration	Number of cycles
94 °C	4 minutes	1 cycle
94 °C	25 seconds	35 cycles
50 °C	25 seconds	
60 °C	4 minutes	

Table 2.5 Conditions used for the cycle sequencing reaction in the thermal cycler

2.4.4.8 Alcohol precipitation

Ethanol precipitation was carried out using 3.5µl precipitation buffer containing 1.5M NaOAc, 1.5M EDTA and 100µl of absolute ethanol, mixed with each sample prior to spinning down for 30 minutes at 2,000 x g (4°C). Then, the supernatant was removed carefully and the samples were spun upside down briefly prior to the washing step with 200µl 70% (v/v) ethanol. Samples were centrifuged again at 2,000 x g (4°C) for 30 minutes and the supernatant was removed as before and then allowed to air dry in the dark. After that, pellets were resuspended in 10µl HiDi Formamide solution (Applied Biosystems; Cat.No: 4311320) and DNA was allowed to denature at 95°C for 5 minutes prior to running on the Applied Biosystem DNA Analyzer.

2.5 Gene expression analysis

2.5.1 cDNA synthesis

In all expression studies, cDNA synthesis was performed using 1µg total RNA. The reaction was set up using 1µl of 100µM random hexamers (Fermentas; Cat. No. S0142), 1µl 10mM dNTP mixture, 1µg RNA and RNase free water to make up the reaction to 14µl. Samples were incubated at 65°C for 5 minutes in the thermal cycler and then placed on ice for a minute to add the mixture of 4µl 5X First-Strand buffer (250mM Tris-HCL, 375mM KCl, 15mM $MgCl_2$), 1µl 0.1M DTT (Dithiothreitol), 1µl RNaseOUT (40$_U$) and 1µl (200$_U$) Superscript III RT (Invitrogen; Cat.No: 18080-093). Then, the samples were again placed in the thermal cycler and incubated at 25°C for 5 minutes followed by 50°C

for 60 minutes. In the last step, the reaction was inactivated by incubation for 15 minutes at 70°C and the cDNA was stored in the freezer until needed.

2.5.2 Primer design for Reverse transcription PCR (RT-PCR)

Primers for expression analysis were designed using the Primer3 online tool (http://primer3.ut.ee/) according to the following criteria; (i) Primers were designed to cover a region of 250-400bp and did not match other regions in the human genome when using NCBI BLAST, (ii) Primer length does not exceed 25bp with Tm 50°C-65°C, (iii) GC content ranges between 50-60%, (iv) Primer does not have a long string of As and Ts or Cs and Gs plus the 3' end contains C or G. The primer sequences are listed in Appendix 7.1.3.

2.5.3 RT-PCR

RT-PCR reaction was prepared to a final volume of 25µl containing 2.5µl 10X buffer (500mM Tris-HCL, 100mM KCl, 50mM $(NH_4)_2SO_4$, 2mM $MgCl_2$, 1x 5X GC rich solution, 2.5mM dNTP mixture), 0.8µM each of the forward and reverse primers, 0.5U FastStart Taq DNA polymerase (Roche; Cat. No. 12158264001) and 1µl cDNA. The reaction was carried out using the touchdown programme in Table 2.6. RT-PCR for *GAPDH* was carried out simultaneously with the exception of using 28 cycles for *GAPDH*.

Temperature	Duration	Number of cycles
95 °C	5 minutes	1 cycle
95 °C	30 seconds	3 cycles
60 °C	30 seconds	
72 °C	30 seconds	
95 °C	30 seconds	3 cycles
58 °C	30 seconds	
72 °C	30 seconds	
95 °C	30 seconds	34 cycles
56 °C	30 seconds	
72 °C	30 seconds	
72 °C	10 minutes	1 cycle

Table 2.6 Touchdown RT-PCR reaction conditions for expression analysis

2.5.4 Quantitative real time expression PCR qRT-PCR

MIR125B1 expression was determined using the TaqMan Universal PCR Master Mix (Applied Biosystems, Cat. No. 4324018) and the TaqMan MicroRNA Reverse Transcription assay (Applied Biosystems, Cat. No. 4366596). Predesigned TaqMan probes, ordered from Applied Biosystems, for MIR125B1 and the endogenous control RNU24 were used to hybridise the previously synthesised cDNA. To determine cDNA concentration, the FAM dye protocol was selected to detect the amount of fluorescence. Standard DNA and sample cDNA were run in triplicates and the produced quantitative amplification was calculated from the standard curve. ΔCt was calculated after normalising to the endogenous control RNU24.

2.5.5 Tissue culture and 5-aza treatments

All tissue culture work was carried out by a laboratory technician, Dean Gentle, in a medical and molecular genetics laboratory (College of Medical and Dental Sciences, University of Birmingham, United Kingdom).

2.5.5.1 Growth media preparation

The glioma cell lines in this study were preserved in DMEM media (Sigma-Aldrich, Cat. No. D5671) containing 10% (v/v) foetal bovine serum (FBS) (PAA Laboratories; Cat. o: A15-151), 2mM L-glutamine (Sigma Aldrich, Cat. No. G7513), 100ᴜ/ml penicillin and 100µg/ml streptomycin (Sigma Aldrich; Cat. No. P4333).

2.5.5.2 Cell line resurrection

Frozen cell lines were defrosted by placing the tubes in a 37°C water bath and then transferred to a 10ml pre-warmed culture media. After that, the cells were spun down at room temperature for 4 minutes at 280 x g and the supernatant was carefully discarded. Then, cells were resuspended in 1ml media and transferred to a tissue culture flask prior to the addition of 7ml growth media. Cells were then maintained in the 37°C incubator with 5% CO_2.

2.5.5.3 Cell line passaging

Every two days, approximately half of the media were discarded from the flask and replaced with fresh pre-warmed media to feed the cells. When cells reached ~80% confluency, they were passaged into 1:3, 1:5 or 1:10 depending on the growth rate. Then, the media were removed and the cells were washed with 4ml PBS (Phosphate Buffered Saline) and 3ml 1x 0.25% Trypsin-EDTA (Invitrogen, Cat. No. 25200056). After that, the cells were placed in 37°C incubator until the cells displaced from the flask wall and were resuspended in 7ml culture media to be split into the desired volume. Media was added to make up a final volume of 15 ml prior to the cells being placed in the 37°C incubator.

2.5.5.4 Counting of cells

Prior to cell counting, the cells were first trypsinised as described in section 2.5.5.3 and then resuspended in 7ml media. Before the haemocytometer chamber was used, the chamber and its cover slip were disinfected thoroughly. Then, 70µl of the media

containing the cells were placed on each haemocytometer slide and counted under the microscope to calculate the average. Preferably, each ml of media contained $1x10^4$ cell counts.

2.5.5.5 Cell line treatment with 5-aza-2'deoxycitidine

Cell line treatment with the demethylation agent, 5-aza-2'deoxycitidine (5-azaDC), was carried out when cells reached 30%-50% confluency. This is due to the fact that demethylation by 5-azaDC only affects cells under DNA replication. Cells were maintained in media supplemented with 5µM 5-azaDC (Sigma-Adrich; Cat. No. A3656) and changed every day for five consecutive days. For each cell line, a non-treated control was grown simultaneously.

2.5.5.6 Cell pellets preparation

When cells reached 90% confluency or on the fifth day of 5-azaDC treatment, they were washed by PBS and then trypsinised to detach from the flask's wall. Then, the cells were resuspended in media and centrifuged at 1000 x g for 4 minutes. The supernatant was discarded carefully and then 1ml PBS was added to each pellet followed by centrifugation for 4 minutes at 1000 x g. The supernatant was removed and the pellets were placed in liquid nitrogen and then stored at -80°C until needed.

2.6 Infinium HumanMethylation450K BeadChip

The Infinium HumanMethylation450K BeadChip from Illumina and DNA bisulfite modification was performed by Cambridge Genomics Services, Department of

Pathology, Cambridge University. Raw data from the array were initially run in the GenomeStudio software and then processed using Lumi package from GenomeStudio software (Du, Kibbe, and Lin 2008) for normalization and colour correction between the Infinium I and Infinium II channels. The β-value was then calculated to estimate the methylation level for each probe using ratio of intensities between methylated and unmethylated alleles: β = Methylated allele intensity (M) / (Unmethylated allele intensity (U) + Methylated allele intensity (M) + 100). The β-value ranged between 0-1; where 0 represents no methylation (0%) and 1 represents full methylation (100%).

In Chapters 3 and 4, the filtration steps were applied on the array data to remove technically failed and unwanted probes. Probes showing the detection p-value, which represents the hybridisation quality over the background, >0.01 in a sample were excluded. In addition to the removal of probes located on X or Y chromosomes, probes for imprinted genes and probes with SNPs were also excluded.

2.7 Sanger sequencing

Sanger sequencing was used in this study for mutation detection.

2.7.1 Primer design

Primers were designed around the mutation site using the Primer3 website (http://primer3.ut.ee/) with this criteria; (i) GC content ranges between 50%-60%, (ii) Design primers with 19-25bp long and 3' end in G or C, (iii) Melting temperature of 60°C (50°C -65°C), and (iv) product size does not exceed 300bp. Then, primers were checked to see whether they contain SNPs within their sequences to avoid PCR

amplification bias. The primer sequences and annealing temperature are listed in Appendix 7.1.4.

2.7.2 MicroCLEAN

To clean up PCR products, MicroCLEAN (Microzone; Cat. No. 2MCL-10) was used by adding 2.4µl of the solution to an equal amount of the PCR product. Samples were spun down at maximum speed for 40 minutes.

Cycle sequencing reaction and ethanol precipitation were then carried out (as in sections 2.4.4.7 and 2.4.4.8) prior to the samples being loaded on the Applied Biosystem 3730 DNA Analyzer.

2.8 Whole exome sequencing

Whole exome sequencing (WES) for schwannoma samples was carried out by the NIHR GSTFT/KCL Comprehensive Biomedical Research Centre, King's College of London. Briefly, WES capture was carried out using in-solution hybridisation, followed by massive parallel sequencing with the SureSelect All Exon v5 Target Enrichment System (Agilent) and sequencing on an Illumina HiSeq200 with 100bp paired-end reads. Later, the reference genome (hg19) was aligned to the sequence reads using Novoalign (Novocraft Technologies SdnBhd). Errors such as duplicate reads and reads mapping to several locations were eliminated from further analysis. Then, the depth and breadth of the sequence coverage were calculated by custom scripts and the BedTools package. Substitutions, insertions and deletion mutations were identified and filtered within the SamTools software package and in-house software tools. Variants were annotated with

respect to genes and transcripts with the Annovar tool. Novel variants were identified by comparing them to dbSNP137 and the 1000 Genomes Project SNP calls (December 2010) and to variants identified in 900 control exomes (primarily of European origin).

The average coverage is dependent on a number of variables; these include the amount of sequencing undertaken, the success of the hybridisation, and the clonality of the library. Acceptable coverage is considered to be when >80% of the exome is covered by >20x.

Annotated, coding annotated and BAM files were downloaded from a webserver where annotated files contain all detected variants including those in non-coding regions and low quality calls. On the other hand, a coding annotated file only contains coding variants with good quality calls. BAM files were downloaded to be used in the IGV tool (section 2.9.5.2).

2.9 Bioinformatics

2.9.1 The Cancer Genome Atlas (TCGA)

The TCGA data portal (http://cancergenome.nih.gov) provides information from different platforms, such as DNA methylation, somatic mutations, miRNASeq, gene expression and clinical information about the patient samples. These data are freely available for download and analysis. In Chapters 3 and 4, DNA methylation data (HumanMethylation450) were downloaded for selected samples of glioblastoma and breast tumours respectively. In Chapter 5, cBioPortal for the cancer genome (http://www.cbioportal.org/public-portal/) was used to investigate the overall percentage of mutations in different cancer panels from TCGA.

2.9.2 Gene structural and functional information

2.9.2.1 UCSC Genome Bioinformatics

Gene information, CpG island information, SNP detection in primer sequence and FASTA sequences of genomic DNA and coding DNA were obtained from the UCSC Genome Bioinformatics' browser (http://genome-euro.ucsc.edu) using the latest human (Homo Sapiens) genome assembly (hg19).

2.9.2.2 The National Center for Biotechnology Information (NCBI)

The NCBI website was used to obtain information about an individual genome, such as gene function, publications and FASTA sequence and amino acid sequence for a specific transcript of a gene.

2.9.2.3 Ensemble genome browser

The Ensemble genome project (http://www.ensembl.org) is a freely available database that produces genomic data on humans and other species. The database was used in the study to analyse genes or SNPs (with rs ID), determine the number of exons of each gene, and to find the distribution of variants among populations.

2.9.2.4 1000 genomes

This website (http://www.1000genomes.org) contains a catalogue of human genetic variations from different populations. There are 14 populations which were divided into 5 main populations: African, Mixed American, East Asian, European and South Asian. In

the study in Chapter 5, the Iberian population in Spain (IBS) were mostly considered in the study and compared to ALL populations (1,092 samples). According to the website, the IBS population contains 14 samples (7 males and 7 females) which are lymphoblastoid cell lines derived DNA.

2.9.2.5 The Catalogue of Somatic Mutations In Cancer (COSMIC)

The COSMIC website (http://cancer.sanger.ac.uk) contains information about mutations, publications and samples in relation to human cancer. The website's database was used to investigate whether a specific mutation had been identified in any cancer tissue in addition to assessing the distribution of different mutations in a certain gene.

2.9.2.6 Exome variant server (EVS)

EVS is a SNP database (http://evs.gs.washington.edu) that contains variants generated by sequencing DNA from European American and African American populations. These populations share the same phenotype of having lung, heart and blood diseases.

2.9.3 The Data for Annotation, Visualisation and Integrated Discovery (DAVID)

The DAVID database (http://david.abcc.ncifcrf.gov) is a publically available database for analysing a large set of genes for pathways, functions and disease relations. It was used for prioritising gene lists in different studies in this thesis.

2.9.4 Prediction of functional effects of human SNPs

2.9.4.1 Polymorphism Phenotyping v2 (PolyPhen-2)

Polyphen-2 (http://genetics.bwh.harvard.edu/pph2) is a computational analysis tool for predicting the pathogenesis of missense variants on proteins. It uses a combination of two parameters; protein structure and protein function, to calculate the score of pathogenicity. Scores range from 0 (predicted to be benign) to 1 (predicted to be probably damaging).

2.9.4.2 Splice site prediction by Neural Network

This is an in silico method (http://www.fruitfly.org) to predict the effect of a splice-site mutation on gene splicing and proteins.

2.9.5 Software and tools for analysis

2.9.5.1 Ingenuity Pathway Analysis (IPA)

This software (www.ingenuity.com) was used in Chapter 3 to analyse a list of hypermethylated genes for gene ontology and pathways. IPA software makes it possible to analyse and find the relevance between a list of genes for signalling pathways and cell networks in addition to gene functions. Part of the IPA analysis was carried out by my colleague Victoria Hill.

2.9.5.2 The Integrative Genomics Viewer (IGV)

IGV (http://www.broadinstitute.org/igv/) is a visualising tool that allows the exploration and analysis of large exome sequencing data for SNP/SNV, number of reads, quality of calls, deletion or insertion of a region and many other aspects. It uses the most recent human genome (hg19) as a reference for genomic alignment. In Chapter 5, the IGV tool was used to validate SNV calls, reduce false positive calls and validate true findings.

2.9.5.3 VarScan2

VarScan2 tool is a somatic mutation and copy number alteration caller for exome sequencing and whole genome sequencing data. It analyses data from tumour/normal paired samples for the accurate detection of somatic mutations and alterations. VarScan2 uses the SAMtools pipeline which reads BAM files generated from exome sequencing for normal and tumour pairs. It then analyses the paired samples simultaneously to identify LOH, germ-line and somatic calls. By applying pre-setup criteria for call quality and coverage, SNV was present in >15% in tumour samples and <5% in matching normal samples were considered as a somatic call, while variants found in both samples were classified as a germ-line call. VarScan2 analysis was carried out by the bioinformatician Michael Simpson of Kings College London.

2.10 Statistical analysis

2.10.1 Student t-test

A student t-test was used in the study of LTS and STS methylation (Chapter 3). The test was performed using the function "TTEST" in Microsoft Excel software during the analysis. The formula selected for the calculation was two-tailed distribution and two-sample unequal variance. Significance was taken when P-value <0.05.

2.10.2 Fisher's exact test

This statistical test was performed to determine the significance of a methylated probe (p-value) between different groups. P-value was calculated using GraphPad software (http://graphpad.com) and selecting two-tailed test. P-value <0.05 was considered to be significant in all studies.

2.10.3 False discovery rate (FDR)

When analysing a large dataset, the chance of identifying false positive results increases, therefore, FDR correction was applied after the calculation of P-value from the student t-test. P-values showed significance (<0.05) from student t-test were ranked from smallest to largest and then each value was multiplied by its rank. Like other tests, P-value <0.05 was considered significant.

2.10.4 Kaplan-Meier survival analysis

In Chapter 3, the univariate Kaplan-Meier survival test was carried out to estimate patient survival when a particular gene is methylated. Calculations of the test and survival graphs

were produced using SPSS statistics software from IBM. As with other tests, P-value was considered significant when <0.05. In Chapter 4, Kaplan-Meier survival curves were produced using the PrognoScan online tool (http://www.prognoscan.org/). The online tool finds the relationship between gene expression and patient clinical prognosis by calculating overall-survival (OS) and disease-free-survival using freely available cancer microarray databases from the Gene Expression Omnibus (GEO) and other individual websites.

Chapter Three: DNA Methylation Profiling of Long-Term and Short-Term Glioblastoma Survivors

3.1 Introduction

According to the WHO classification, glioma is classified into 4 grades. Glioblastoma multiforme (GBM) is grade IV astrocytoma, which can be classified into primary GBM, when discovered de novo at diagnosis, or secondary GBM rises from a lower grade glioma. GBM is the most aggressive form of glioma in adults. Surgery and a combination of chemotherapy and radiotherapy are the standard treatment methods for GBM patients. However, the median survival rate for GBM patients is less than 1 year (Parsons et al. 2008). Researchers have found that a small fraction (3-5%) of GBM patients showed a higher survival rate of up to 3 years and are so-called long-term survivor (LTS) GBM patients (Krex et al. 2007). These patients are younger in age, have a better clinical performance and have *MGMT* hypermethylated tumours (Krex et al. 2007; Hartmann et al. 2013).

The development of Illumina methylation arrays from the GoldenGate DNA Methylation BeadArray to the Infinium HumanMethylation450 BeadChip allowed cancer researchers to expand and evolve to understand the epigenetic role in cancer development. The HumanMethylation450K array covers more than 485,000 CpG loci, of which 96% are located in CpG islands, CpG shores and CpG shelves Also, the array covers almost 99% of RefSeq genes and is able to annotate the position of the CpG probe according to the functional genomic distribution (5'UTR, 3'UTR, 1st exon and gene body) and to the CpG content and neighbourhood context (CpG island, shore, shelf or open-sea).

100

This study utilises the HumanMethylation450K array to study the methylation profile of the LTS gliomas and short-term survivors (STS) glioma and to identify potential epigenetic markers for poor prognosis in STS patients.

3.2 Aim of the study

- Methylation profiling of short-term survival (STS) and long-term survival (LTS) gliomas by using the genome-wide Infinium HumanMethylation450K BeadChip from Illumina.

- To identify a list of hypermethylated genes and potential candidates in STS that may be markers for poor survival.

3.3 Results

This study used the most advanced methylation detection array available at the time of the study, the HumanMethylation450K BeadChip array, on a panel of 35 grade IV glioblastoma tumours consisting of 19 STS and 16 LTS glioblastoma DNA samples, in addition to 4 normal non-cancerous brain DNA samples. Appendix 7.2 contains clinical information about the tumour samples.

The latest Infinium methylation array from Illumina, the HumanMethylation450K BeadChip, evaluates the methylation levels of more than 485,000 CpG sites across promoter regions (within 200bp of the transcription start site, 1500bp of the transcription start site, the first exon or 5'UTR regions), 3'UTRs and regions of the gene body. It covers CpG loci located in 96% of CpG islands, in addition to sites outside the island region such as CpG shore, shelf and isolated CpG (open sea). The

HumanMethylation450K array differs from other methylation arrays by using two probe types, Infinium I and Infinium II, to detect methylation level. Methylation level is represented by a β value ranging between 0 (= 0% methylation) and 1 (= 100% methylation).

Here, before starting the analysis of GBM methylation data, raw data were first corrected and normalised using a correction pipeline, Lumi R package, to adjust the colour bias between the two different assays of Infinium I/II. Then, a series of data filtration processes were applied by removing the probes showing detection p-value >0.01 (which reflects poor hybridization), probes located on X or Y chromosomes, probes for imprinted genes and probes with SNPs. Hypermethylation in tumour samples was considered when CpG loci demonstrated high methylation (β value >0.5) in >30% of the samples and β value <0.25 in any of the four normal samples. Hypomethylation was determined when the probe β value was <0.25 in 30% of tumour samples and methylated in any of the four normal samples (β value >0.5). At the end of the filtration process, 129,545 and 197,297 probes were tumour specific hypermethylated and hypomethylated, respectively (Figure 3.1).

HYPERmethylation Analysis
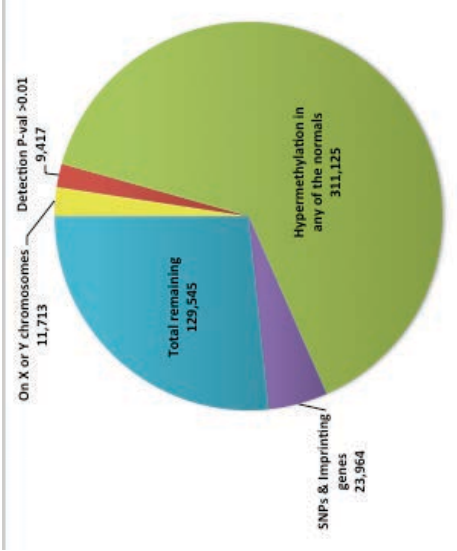
HYPOmethylation Analysis
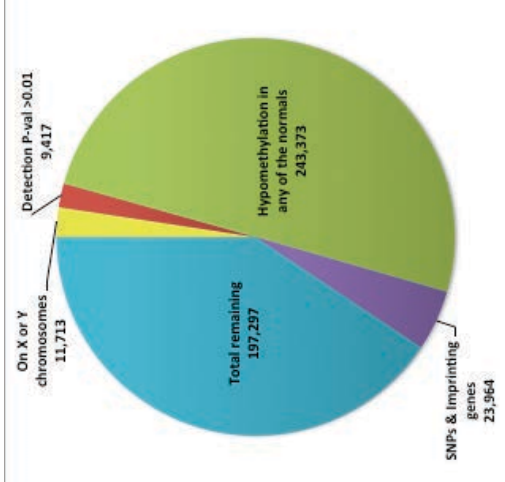
Figure 3.1 Filtration processes applied to array's data prior to hypermethylation and hypomethylation analysis

Loci with a high detection p-value (9,417 loci), loci located on X and Y chromosomes (11,713 loci), and loci for SNPs & imprinted genes (23,964 loci) were removed from the analysis. Also, tumour specific hyper/hypo methylation was ensured by excluding hypo and hyper methylated loci from the normal samples respectively. After the filtrations, the remaining loci were used for hyper (129,545 loci) and hypo (197,297 loci) methylation analysis in this study.

3.3.1 Array validation

At the time of this study, the HumanMethylation450K BeadChip array was a very new assay in the field. Validation of the individual β value of the array data using another method was required, thus, a number of selected genes known to be methylated in glioblastoma (*HOXA3*, *FZD9, RASSF1A*, and *SLIT2*) were used to evaluate the β values of the Infinium probes (Di Vinci et al. 2012, 3; Dallol et al. 2003; Hesson et al. 2004; Martinez et al. 2009). For each gene, selected CpG sites were examined in different samples with high and low β values using bisulfite sequencing. From this data, a probe methylation value was determined (which should be most comparable to the β-value generated from the array) and the Methylation Index (MI) of the whole selected region was calculated as a percentage (MI = percentage of methylated CpGs out of the total number of CpGs in the region). For *HOXA3*, 3 adjacent probes (cg15982700, cg16748008 and cg16406967) were examined in 3 different samples (2 with high β values and 1 with low β values) using bisulfite sequencing. Sequencing of the highly methylated samples showed that the probe methylation level (0.8/0.8/1 and 1/0.8/1, respectively) was high enough to be considered methylated, matching the methylation status of nearly all of the array β values (0.4/0.55/0.7 and 0.53/0.52/0.7, respectively) Reassuringly, high methylation index (MI) values for the two methylated samples (87.5% and 88.88%, respectively) were also detected and matched the probe methylation value well (Figure 3.2). Importantly, probes with lower β values (0.17/0.04/0.03) showed low probe methylation values (0.18/0/0, respectively) and low MI (3.7%). In addition to

confirming the criteria used to determine methylation events, these findings demonstrated that the level of methylation of a single CpG site is generally representative for the methylation of the surrounding CpG island region. The same observation was applicable for the other probes of the remaining genes (*FZD9, RASSF1A* and *SLIT2*). The probe values and β values were similar and MIs were representative of the overall methylation of the region (Figures 3.3, 3.4 and 3.5).

Generally, validation of the array's probe by bisulfite sequencing showed that probes with a β value ≥0.4 indicated a high level of methylation at the level of the CpG site and the level of the CpG island, while probes with much lower β values demonstrated very low to no methylation in the probe position and the surrounding CpG island. These results supported the stringent criteria selected in this study, where a β value of <0.25 represented low methylation and a β value of >0.5 represented hypermethylation.
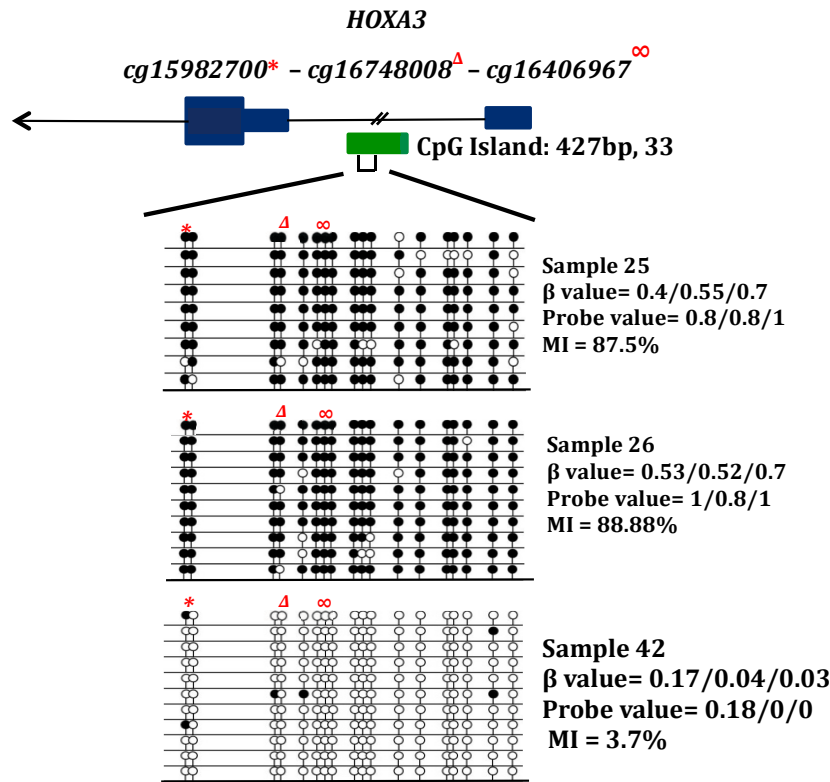
Figure 3.2 Validation of array's β value for HOXA3

Illustration of clone sequencing results for the region within the *HOXA3* CpG island, where the probes (cg15982700, cg16748008 and cg16406967) are located. A schematic diagram of part of the gene (blue) and CpG island (green) is shown. Sequencing results are shown for 3 samples with the corresponding β value, probe value and methylation index (MI). For each case, ~9 clones (each representing a single allele) were selected for bisulfite sequencing and then analysed using CpG Viewer software. Black lollipops represent methylated CpGs (retained as cytosine following bisulfite modification), while white lollipops represent unmethylated CpGs (converted to thymine following bisulfite modification).
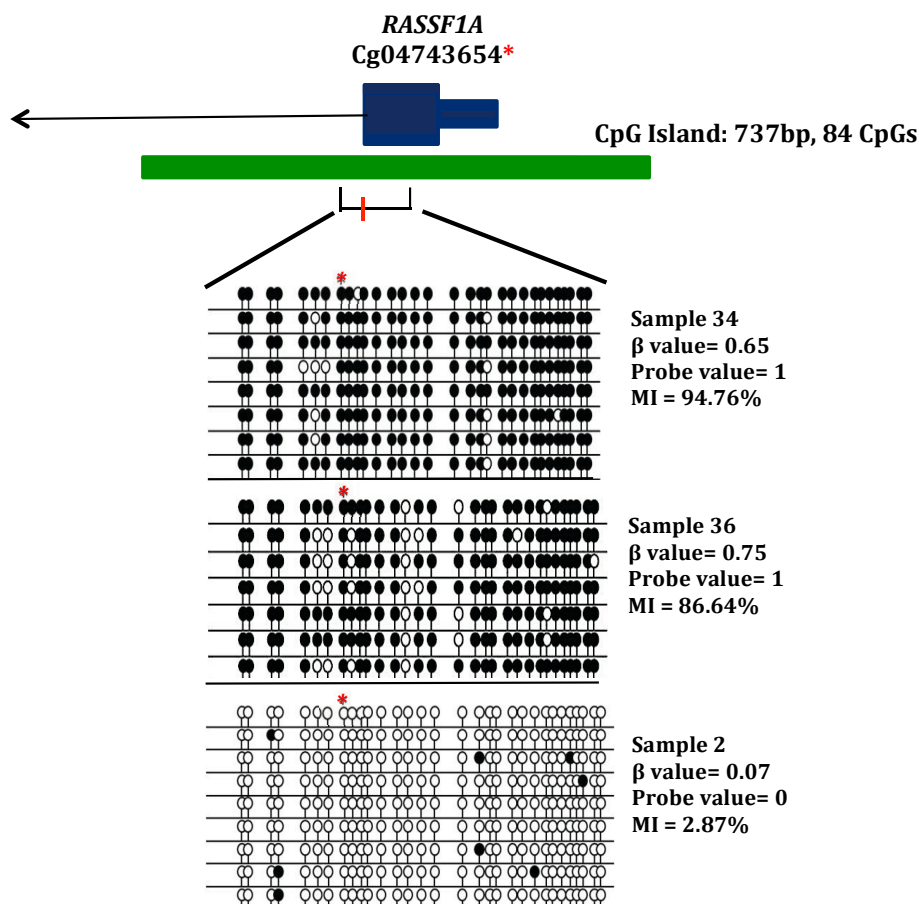
Figure 3.3 Validation of array's β value for RASSF1A

Illustration of clone sequencing results for the region within the *RASSF1A* CpG island that contains the cg04743654 probe. A schematic diagram of part of the gene (blue) and CpG island (green) is shown. Sequencing results are shown for 3 samples with the corresponding β value, probe value and methylation index (MI). For each case, ~9 clones (each representing a single allele) were selected for bisulfite sequencing and then analysed using CpG Viewer software. Black lollipops represent methylated CpGs (retained as cytosine following bisulfite modification), while white lollipops represent unmethylated CpGs (converted to thymine following bisulfite modification).

FZD9
cg12300292* – cg00866572$^\Delta$ – cg08649099$^\infty$

CpG Island: 2099bp, 206 CpGs

Sample 30
β value= 0.72/0.58/0.35
Probe value= 1/1/0.4
MI = 74.6%

Sample 28
β value= 0.05/0.09/0.01
Probe value= 0/0.08/0
MI =8.75 %

Sample 47
β value= 0.11/0.32/0.01
Probe value= 0/0.1/0
MI = 1.25%
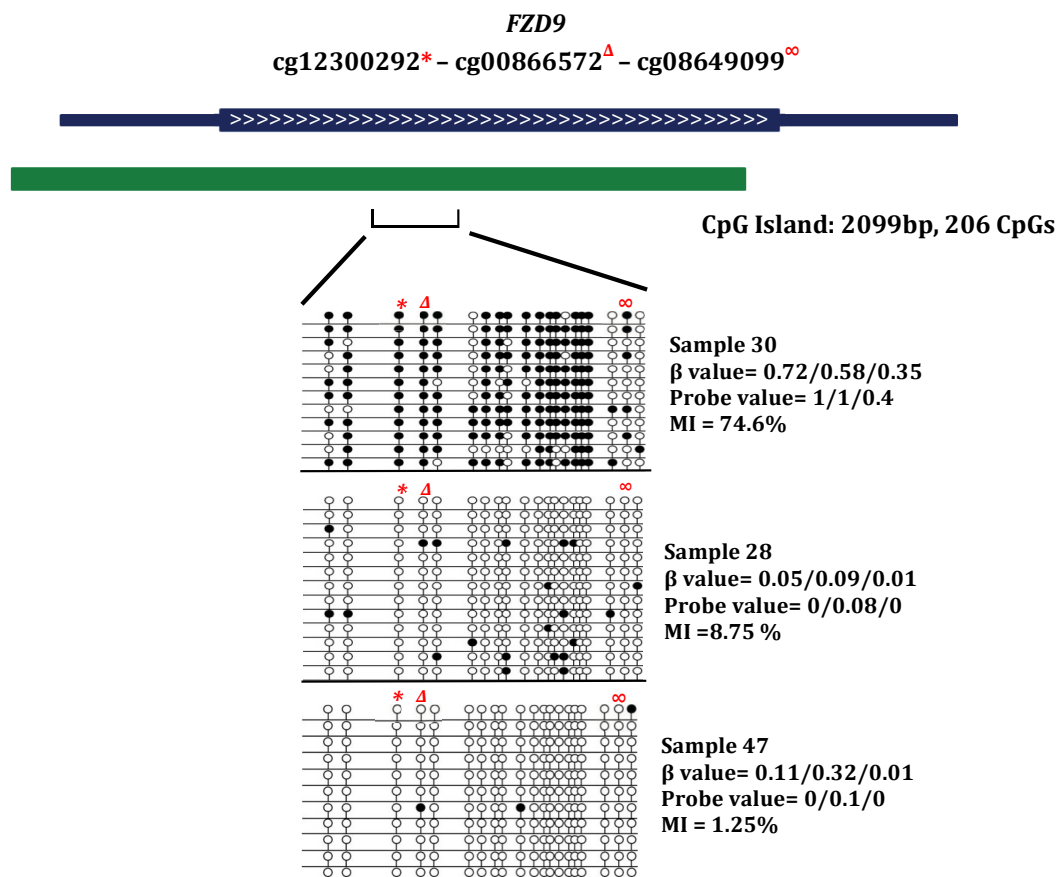
Figure 3.4 Validation of array's β value for FZD9

Illustration of clone sequencing results for the region within the *FZD9* CpG island, where the probes (cg12300292, cg00866572 and cg08649099) are located. A schematic diagram of part of the gene (blue) and CpG island (green) is shown. Sequencing results are shown for 3 samples with the corresponding β value, probe value and methylation index (MI). For each case, ~9 clones (each representing a single allele) were selected for bisulfite sequencing and then analysed using CpG Viewer software. Black lollipops represent methylated CpGs (retained as cytosine following bisulfite modification), while white lollipops represent unmethylated CpGs (converted to thymine following bisulfite modification).
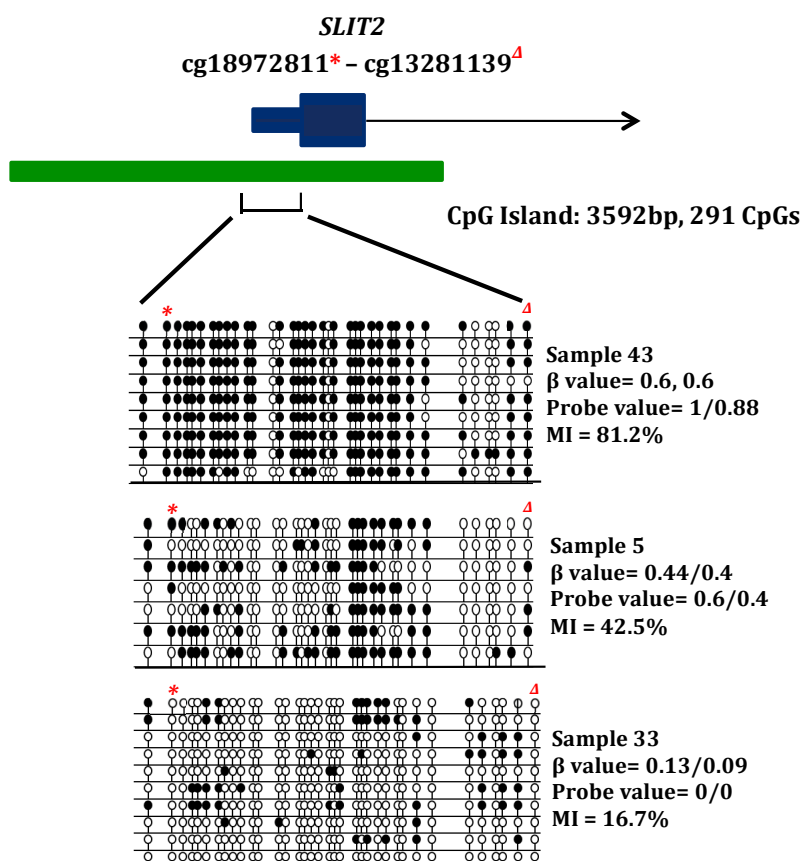
Figure 3.5 Validation of array's β value for SLIT2

Illustration of clone sequencing results for the region within the *SLIT2* CpG island, where the probes (cg18972811 and cg13281139) are located. A schematic diagram of part of the gene (blue) and CpG island (green) is shown. Sequencing results are shown for 3 samples with the corresponding β value, probe value and methylation index (MI). For each case, ~9 clones (each representing a single allele) were selected for bisulfite sequencing and then analysed using CpG Viewer software. Black lollipops represent methylated CpGs (retained as cytosine following bisulfite modification), while white lollipops represent unmethylated CpGs (converted to thymine following bisulfite modification).

3.3.2 CIMP in a subset of LTS grade IV glioblastoma

3.3.2.1 Hierarchical cluster analysis

To gain a general insight into the methylation pattern of the glioblastoma tumour samples used in this study, unsupervised hierarchical clustering of the 500 most variable loci was performed on the 19 STS and 16 LTS glioblastoma samples. The cluster analysis separated glioblastoma samples into 2 distinct groups; the smallest of the cluster groups consisted of 5 LTS tumour samples, which showed extensive hypermethylation across the probes (average β value = 0.66). Such a high level of generalised methylation is indicative of a CpG island hypermethylator phenotype (CIMP), which has been documented in many cancer types, including glioblastoma (Noushmehr et al. 2010). These samples have therefore been classified as CIMP[+] (Figure 3.6). The largest cluster group consisted of all STS samples plus 11 samples of non-CIMP LTS cases. Probes in this group demonstrated low levels of methylation (average β value = 0.25). To gain insight into the methylation differences between STS and non-CIMP LTS glioblastoma, another unsupervised hierarchical clustering was conducted on the 500 most variable probes in all non-CIMP glioblastoma tumours (n=30). The clustering divided the samples into two main cluster groups (Figure 3.7). Cluster 1 was associated with low level methylation (average β value = 0.22) and it mainly consisted of STS samples (n=10, 71.4%), while LTS samples comprised 28.6% of the group. Cluster 2 was associated with higher levels of methylation (average β value = 0.44) and contained equal numbers of STS (n=7, 50%) and LTS (n=7, 50%) samples. Two STS outlier samples were found on the left hand side of the clustering. This observation suggested there were slight differences in the methylation profiles within the non-CIMP tumour samples.
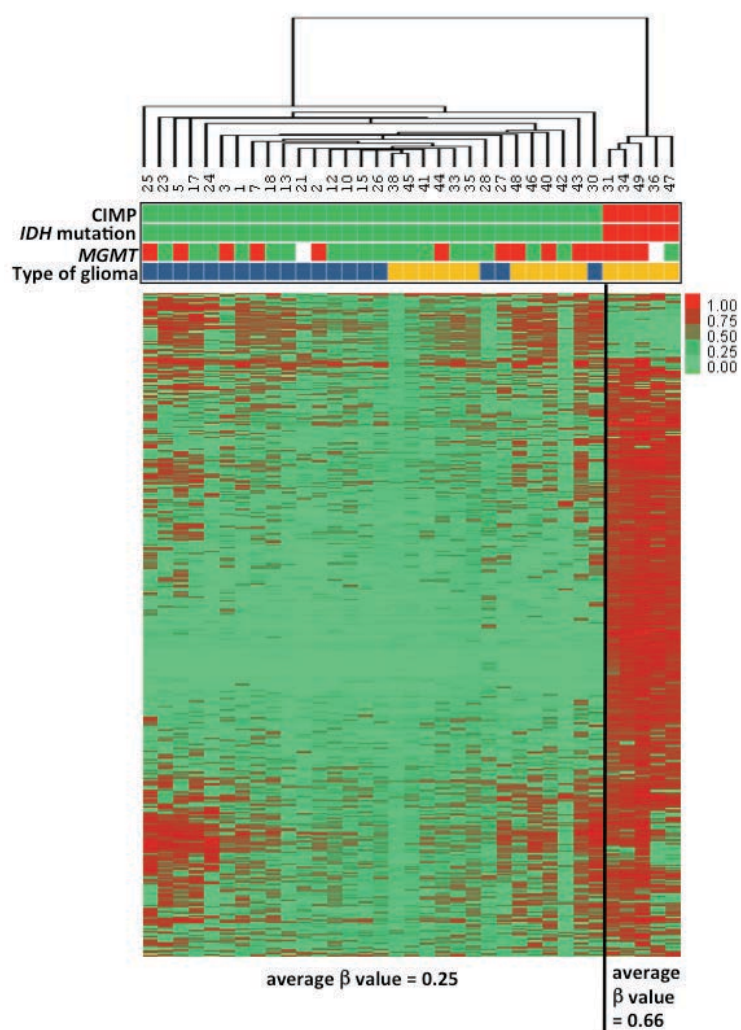
Figure 3.6 Unsupervised hierarchical cluster analysis of the 500 most variable loci

The clustering identified G-CIMP+ in 5 LTS tumours with a concentration of hypermethylated loci. The remaining G-CIMP- STS and LTS tumours were clustered together. CpG loci are colour coded from red (high methylation) to green (low methylation). CIMP status (red, CIMP+; green, CIMP-), IDH mutation status (red, mutated; green, unmutated), MGMT methylation status (red, methylated; green, unmethylated; white, undetermined), and type of glioma (blue, STS; yellow, LTS) is shown in the information bar below the figure.
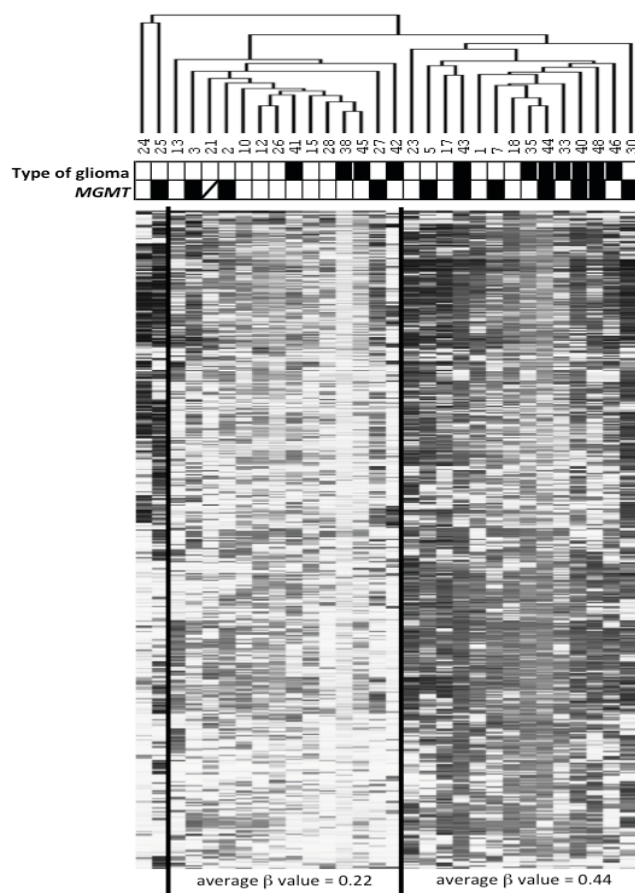
Figure 3.7 Unsupervised hierarchical clustering was conducted on the 500 most variable probes in non-CIMP glioblastoma

The clustering splits the samples into two main cluster groups. The first group is associated with a low level of methylation and it mainly contains STS samples (n=10, 71.4%), while LTS samples comprise 28.6% of the group. Also, the majority of the samples show no methylation of *MGMT* gene (76.9%), whereas the other group is associated with a high methylation level and contains equal numbers of STS (n=7, 50%) and LTS (n=7, 50%) samples. This group also shows 50% methylation of *MGMT* gene. Two outlier samples are found on the left hand side of the clustering. The black and white shading of the cluster represents high or low methylation respectively. The box above the clustering represents *MGMT* methylation (black, methylated; white, unmethylated) and type of glioma (white, STS; black, LTS).

3.3.2.2 *IDH*1/*IDH2* mutation and *MGMT* methylation

MGMT promoter hypermethylation and *IDH1/IDH2* mutation status have been described extensively in many glioblastoma studies and associated with better prognosis and treatment response. Therefore, methylation specific PCR (MSP) was performed to examine *MGMT* promoter methylation status in all glioblastoma samples. The results showed that methylation of the *MGMT* gene was similar between STS and LTS samples (38.9% and 46.7% respectively) regardless of the glioma CpG island methylator phenotype.

DNA sequencing was performed to determine the mutation status of 2 previously described hotspot mutations in the *IDH1* and *IDH2* genes in glioblastoma samples. The location surrounding the hotspot mutation in exon 4 (R132) was amplified and sequenced in the DNA of all STS and LTS samples. Only 5 cases demonstrated *IDH1* (R132) mutation and all 5 were LTS tumours with CIMP+ phenotype (Figure 3.8). Samples of wild-type for *IDH1* at the position R132 were subsequently analysed for *IDH2* mutation at R172 position, which is the equivalent to R132 in *IDH1*. No *IDH1* or *IDH2* mutations were found in the remaining STS and LTS glioblastoma samples. In this study, 60% of the *IDH* mutated cases of LTS showed hypermethylation of *MGMT*.
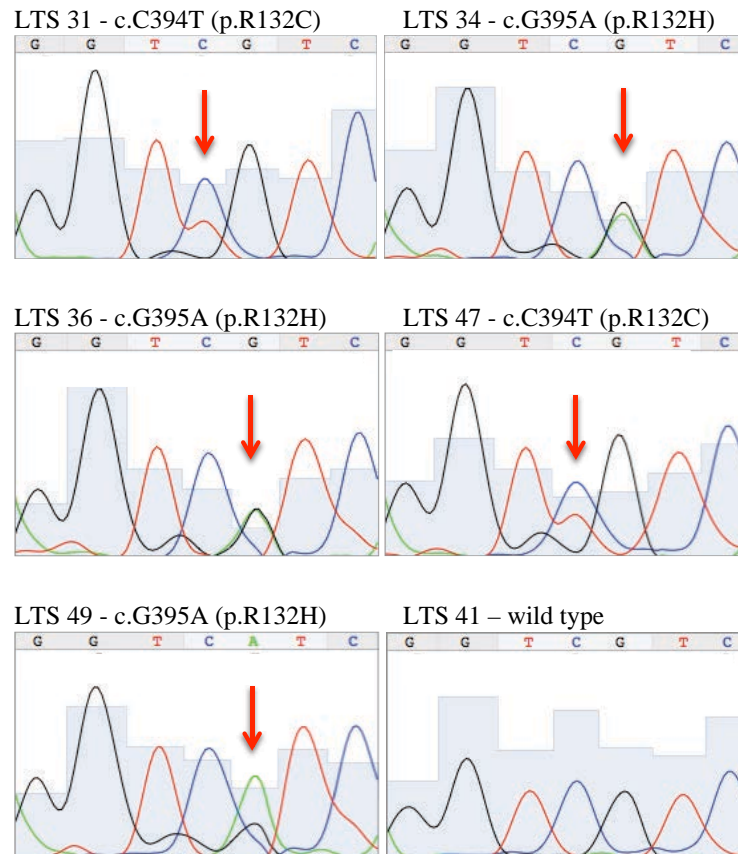
Figure 3.8 Identification of IDH1 mutations in G-CIMP+ LTS grade IV glioblastoma

Sanger sequencing confirms *IDH1* mutation in G-CIMP+ LTS glioblastoma tumours (LTS 31, LTS 34, LTS 36, LTS 47 and LTS 49) but not in G-CIMP- LTS glioblastoma (LTS 41). The identified *IDH1* mutation in all G-CIMP+ LTS glioblastoma tumours was located in R132 of exon 4. No *IDH2* mutation was identified in any of the samples.

3.3.2.3 Confirmation of G-CIMP$^+$

During the original study that identified CIMP within glioblastoma (Noushmehr et al. 2010), seven signature genes were identified (*ANKRD43*, *HFE*, *MAL*, *LGALS3*, *FAS*, *RHO*, *DOCK5*), whereby a specific methylation status indicated a hypermethylator phenotype within glioblastoma, referred to as G-CIMP$^+$. According to the criteria in the Noushmehr study, a combination of at least 5 of the 7 signature genes must show low methylation in *DOCK5* and/or high methylation in the other genes. To determine whether the CIMP classification from earlier clustering analysis was in agreement with previous G-CIMP descriptions, the probes for these 7 signature genes were assessed among the STS and LTS samples. When applying these criteria, only the 5 *IDH* mutated LTS samples showed hypomethylation of *DOCK5* and hypermethylation of at least 4 of the remaining genes, matching Noushmehr's criteria for G-CIMP$^+$ (Figure 3.9). None of the remaining *IDH*-wt glioblastoma samples fulfilled the criteria (max 4/7) and therefore exhibited G-CIMP$^-$ phenotype. This finding confirms the previous classification of CIMP status within STS and LTS samples and illustrates the close association between the CIMP phenotype and *IDH* mutation, which is in agreement with other studies (Turcan et al. 2013; Noushmehr et al. 2010).

Figure 3.9 Confirmation of CIMP status in a subset of LTS tumours

Using the criteria addressed by Noushmehr *et al.* based on 7 CIMP signature genes; *ANKRD43, HFE, MAL, LGALS3, RHOF, FAS* and *DOCK5* in STS and LTS glioma, a subset of LTS tumours with IDH mutation fulfilled the criteria. Green and red circles represent no methylation and methylation respectively. *IDH* mutation and *MGMT* methylation status are also given in the figure.

3.3.2.4 Identification of G-CIMP genes in LTS with *IDH*-mut

Use of the HumanMethylation450K BeadChip was able to identify a massive number of hypermethylated loci within the 5 *IDH* mut G-CIMP[+] compared to STS and non-CIMP LTS glioblastoma with wildtype *IDH* (11,293, 2,638 and 3,101 CpG loci, respectively). In order to use this data to identify additional G-CIMP[+] genes, the *IDH* mut/CIMP[+] and *IDH* wt/non-CIMP groups were subjected to further analysis. By applying a student's t-test and then selecting genes with >3 probes within the promoter region that exhibited an FDR corrected *P* value of <0.05 between the two groups of *IDH* mutated and *IDH* wildtype samples, 535 genes (2377 CpG loci) were identified as CIMP[+] genes (Appendix 7.3). DAVID bioinformatics functional analysis of the gene list indicated these genes were involved in various cell regulation processes such as regulation of cell migration (40 genes), cell-cell signalling (105 genes), cell-cell adhesion (47 genes), Wnt signalling pathway (22 genes), regulation of cell growth (55 genes) and apoptosis (89 genes). Analysis of the gene list using the Ingenuity Pathway Analysis (IPA) showed that 31% of the genes (166/535) were associated with different types of cancer and cancer developmental processes such as epithelial tumours (125 genes), solid tumours (122 genes), colon cancer (46 genes), metastasis (35 genes), tumour development (21 genes), growth of tumours (22 genes) and tumour angiogenesis (9 genes) (Appendix 7.4). Molecular and cellular function analysis of the CIMP genes divided the gene list into 5 different functional categories, where the majority of the genes were involved in cell proliferation and growth (166 genes, 31%). The other categories consisted of cellular movement (121 genes, 22.6%), cell morphology (126 genes, 23.6%), cellular

117

organization (105 genes, 19.7%) and cellular maintenance (143 genes, 26.7%), where 50 genes were common among the 5 categories (Table 3.1). Also, a large number of the hypermethylated CIMP[+] genes were found to be involved in glutamate receptor signalling (Figure 3.10) and cell morphology, death and survival (Figure 3.11). Interestingly, when looking at the G-CIMP[+] genes identified by Noushmehr *et al.'s* group, 31% of the identified CIMP[+] genes in the current study were common. Taking all this together, these 535 hypermethylated genes in *IDH*-mut LTS glioblastoma might be markers for the G-CIMP[+] phenotype.

Ingenuity Analysis	Category	No. of Genes	p-value Range	Gene Names (Cancer genes in bold)
Diseases and Disorder	Cancer	166/535 (31.0%)	1.30E-05 - 4.81E-03	See Table appendix 7.4
Molecular and Cellular Functions	Cellular Movement	121/535 (22.6%)	1.53E-09 - 4.78E-03	
Molecular and Cellular Functions	Cellular Growth and Proliferation	166/535 (31.0%)	8.36E-08 - 4.55E-03	Genes present in all functional categories: *ACVRL1, ARHGAP24, **BCL2**, CNTN2, **COL18A1, CRYAB, CXCR4**, DKK3, DPYSL5, **EFBN2, EPAS1, EPHA2**, ERBB3, **FAS**, FGR, FLOT1, **GDNF, GJA1**, GRIN1, **KLF5, LAMA3**, LGALS8, **LYN**, MAGI2, MAPT, **MMP14**, NBL1, **NEUROD1**, NR2E1, NTN1, ONECUT1, ONECUT2, P2RY2, **PAK1**, PARVB, PAX6, **PLEC, PRKCZ, PTEN, RAP1GAP**, RGNEF, RHOD, **SHH**, SST, SWAP70, **TGFB2, THBS1**, THBS4, TRH, VAX2, **WWTR1***
Molecular and Cellular Functions	Cell Morphology	126/535 (23.6%)	4.65E-07 - 4.54E-03	
Molecular and Cellular Functions	Cellular Assembly and Organization	105/535 (19.7%)	4.65E-07 - 4.54E-03	
Molecular and Cellular Functions	Cellular Function and Maintenance	143/535 (26.7%)	4.65E-07 - 4.54E-03	

Table 3.1 IPA Ingenuity analysis of the identified 535 CIMP+ genes

IPA analysis divided the gene list into 2 main functional categories: disease and disorder and molecular and cellular functions which were subdivided into 5 sub-categories: cancer associated, cellular movement, cellular proliferation, cell morphology, cellular organization and cellular function and maintenance. A list of 50 common genes between the 5 categories is shown on the right hand side of the table. Cancer genes are highlighted in bold.

Glutamate Signalling Receptor (7/69 molecules p=0.001)
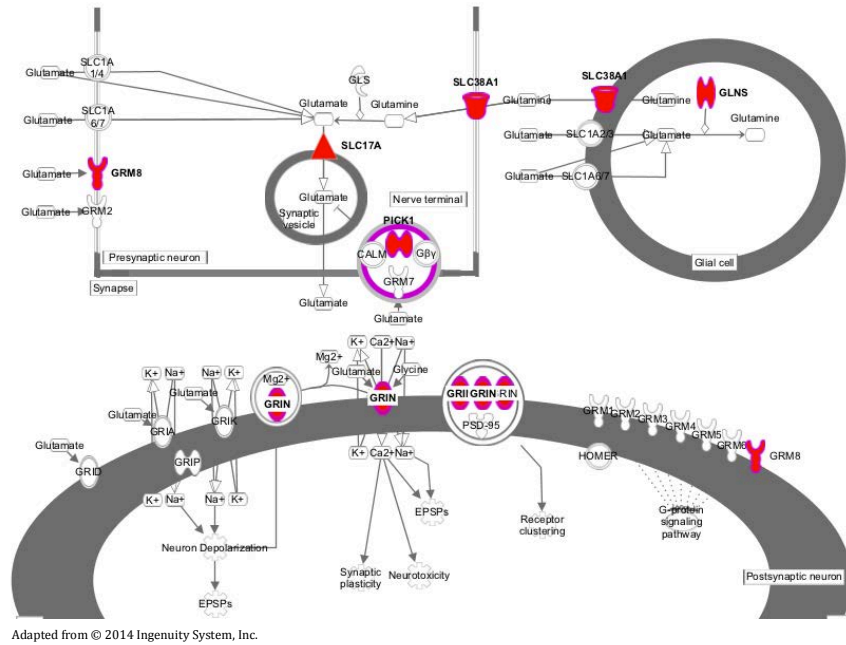
Figure 3.10 Canonical pathway generated from ingenuity analysis of the 535 CIMP associated genes list

Methylated genes in glutamate receptor signalling (*PICK1, GLNS, GRIN, GRM8, GLS, SLC38A1* and *SLC17A*) are shaded red.

120
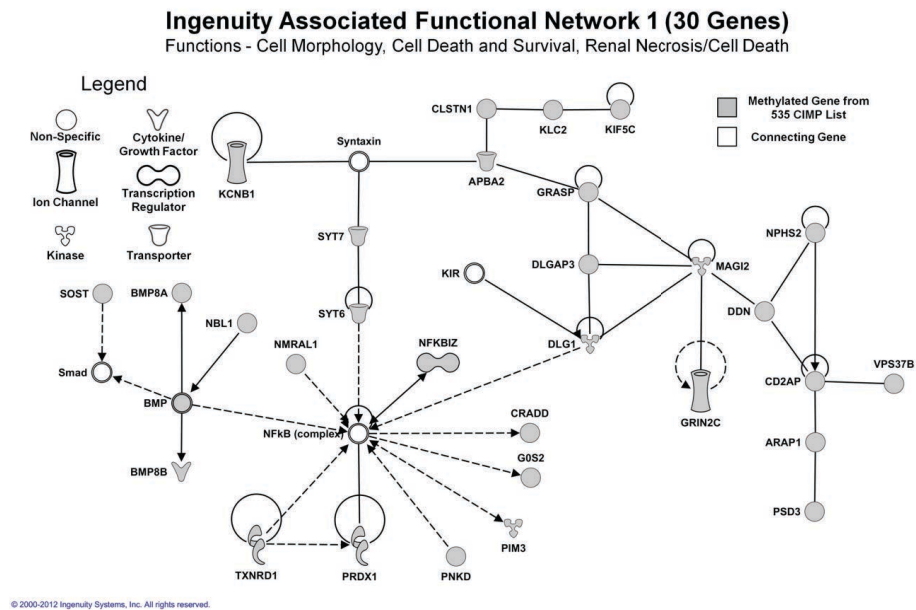
Figure 3.11 Top gene networks generated from ingenuity analysis of the 535 CIMP

Associated List

Methylated genes are shaded grey, while the necessary connecting genes are unshaded. This figure was kindly prepared by Victoria Hill (School of Clinical and Experimental Medicine, University of Birmingham).

3.3.3 Methylome of glioblastoma

Identification of *IDH* mutation in a subset of LTS glioblastoma samples led to splitting the study cohorts into 3 groups: STS, LTS with *IDH*-mut and LTS with *IDH*-wt. After filtering the methylation data and applying the hypermethylation and hypomethylation criteria (as discussed earlier at the beginning of the chapter), the number of hypomethylated CpG loci showing β value <0.25 was low in each of the 3 groups of glioblastoma (946, 690 and 1,402 CpG loci for STS, LTS with *IDH*-wt, and LTS with *IDH*-mut respectively) in comparison to hypermethylated loci (2,638, 3,101 and 11,293 CpG loci for STS, LTS with *IDH*-wt, and LTS with *IDH*-mut respectively). The large number of hypermethylated loci in LTS with *IDH*-mut was due to the presence of the CpG island methylator phenotype.

Despite the difference in the number of hyper/hypo methylated CpG loci, the distribution of these loci in the genome was very similar among the 3 groups. In STS glioblastoma, the majority of hypermethylated CpG loci (1,064 loci, 40%) were located in the gene promoter region, 863 loci (33%) in the intergenic regions, 636 loci (24%) in the gene body and 75 loci (3%) in 3'UTR (Figure 3.12). By looking at the distribution of CpG loci in relation to CpG island, the majority of the hypermethylated loci resided within CpG islands (1,747 loci, 66%) and the remaining were distributed in CpG shores (638 loci, 24%), CpG shelves (77 loci, 3%) and open sea (176 loci, 7%). On the other hand, the majority of hypomethylated loci in STS glioblastoma were located in the gene body and intergenic regions, 33% and 38% respectively, and 74% were found as isolated CpGs distributed in the genome (open sea), while 4% resided within the islands.

In LTS glioblastoma cases, hypermethylated CpG loci in LTS with *IDH1*-mut were mainly located in the promoter regions (5,232 loci, 46%) and CpG islands (7,402 loci, 66%) and shores (2,846, 25%), while hypomethylated CpGs were found in the open sea (1,026 loci, 73%), intergenic regions (523 loci, 37%) and very few (14 loci, 1%) in CpG islands. CpG loci in LTS with *IDH1*-wt showed a similar distribution as in STS cases, where 46% (1,414 loci) of the hypermethylated loci were located in the promoter regions and 72% in CpG islands (2,241 loci), while 73% and 4% of the hypomethylated loci resided in the open sea and CpG islands, 503 and 28 loci respectively (Fig 3.12).

A) Hypermethylation

B) Hypomethylation

i) Functional genomic distribution

Body
3'UTR
Intergenic
promoter

ii) CpG content and neighbourhood context

Island
Shore
Shelf
Open Sea
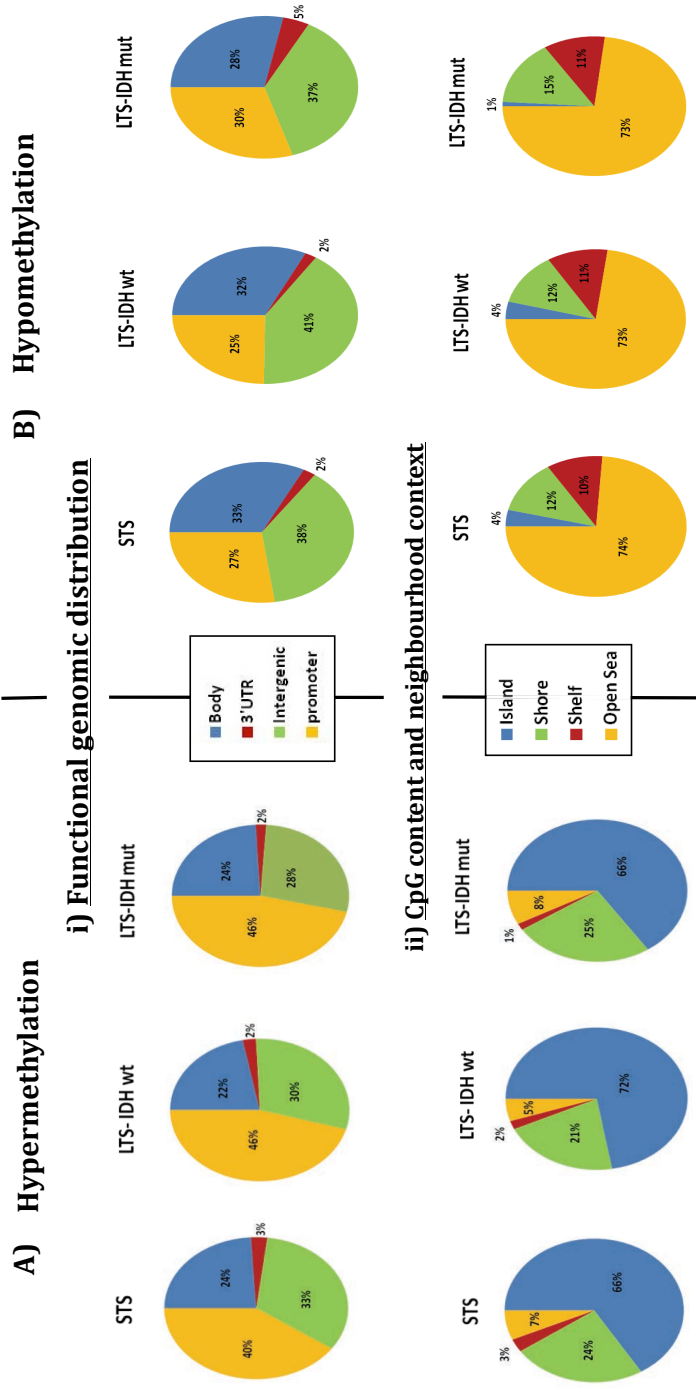
Figure 3.12 Methylome of the 3 groups of glioblastoma tumours: STS, LTS IDH-wt and LTS IDH-mut

Description of the methylome of 3 different groups in A) hypermethylated, B) hypomethylated CpG loci. The distribution was analysed according to i) functional genomic distribution; gene body, promoter region, intergenic or 3'UTR, and ii) CpG content and neighbourhood context; CpG island, shore, shelf or open sea.

124

3.3.4 In-silico validation using TCGA data

To verify the observed hypermethylation patterns in glioblastoma samples a comparison between the top hypermethylated loci in STS glioblastoma and glioblastoma samples from The Cancer Genome Atlas (TCGA) data portal was carried out. All glioblastoma samples from the TCGA website (n=18) that had both (a) data available from the HumanMethylation450K array and (b) a survival time of < 1 year from the initial date of diagnosis were used for this comparison (n=18). Of 2,638 hypermethylated CpG loci in STS glioblastoma, the top 80 hypermethylated loci that showed a β value of >0.5 in >30% of the total number of tumour samples were selected and analysed in glioblastoma samples downloaded from the TCGA website. As expected, all CpG loci were confirmed to be hypermethylated in TCGA glioblastoma samples (Figure 3.13). Analysis of LTS tumours in the TCGA dataset was not carried out because methylation data from the HumanMethylation450K array was not available for patients with a longer survival time.
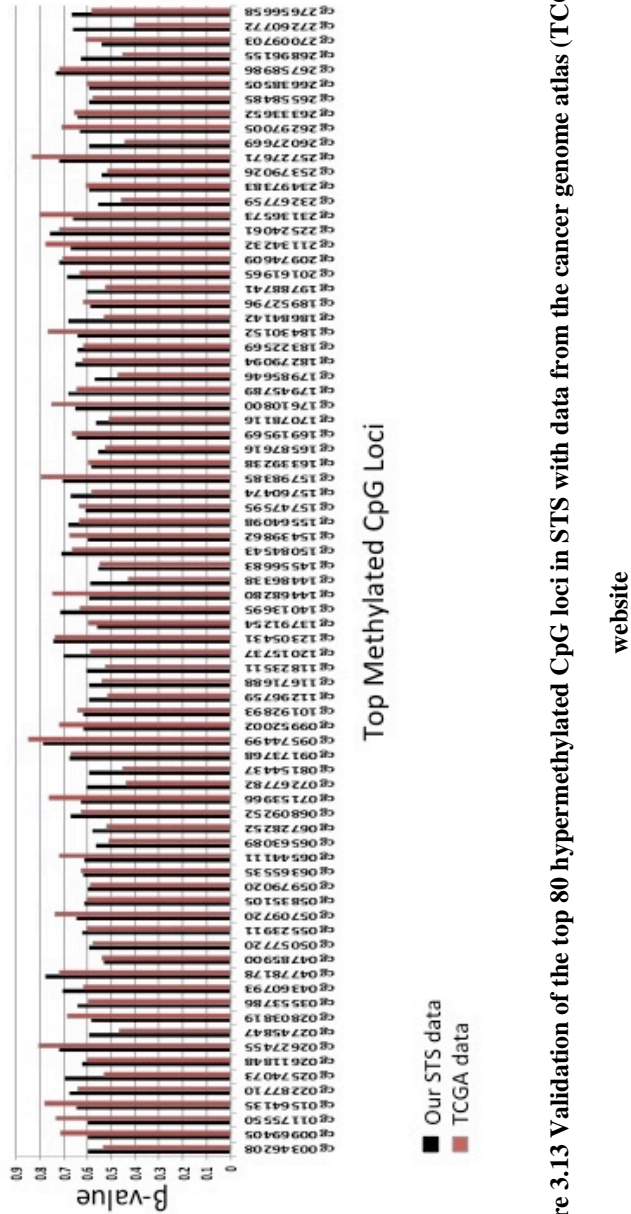
Figure 3.13 Validation of the top 80 hypermethylated CpG loci in STS with data from the cancer genome atlas (TCGA)

website

(Black, present study; red, TCGA data).

3.3.5 Differentially methylated genes in STS glioblastoma

3.3.5.1 Hypermethylated loci in STS vs. LTS glioblastomas

Differential analysis between STS and LTS glioblastoma was carried out to identify epigenetic markers for poor prognosis. To identify the hypermethylated genes in STS vs. LTS, the criterion of a FDR corrected P value below 0.05 and at least a median β value difference of 0.2 between STS and LTS cases was applied. Furthermore, to ensure hypermethylation and loss of expression association, only CpG loci located within the promoter region (TSS200, TSS1500, 5'UTR and 1st exon) were selected. This analysis generated a list of 23 genes (32 CpG loci) that showed significant hypermethylation in STS glioblastoma compared to LTS glioma (Table 3.2). Among the list, there were several genes previously known to be involved in cancer (*HOX* genes) or known to be methylated in glioma (*CBLN1*, *CBLN4* and *ISM1*). IPA analysis of the gene list revealed that 43.5% (10/23) of the genes were associated with cancer, 30.4% (7/23) were involved in cell morphology and 8.7% (2/23) were related to cell death and survival respectively (Table 3.3). Univariate Kaplan-Meier survival analysis was carried out for *NR2F2* and *INA*. DNA methylation was found to be associated with poor patient prognosis for both *NR2F2* and *INA*, P value = 0.034 and 0.036 respectively (Figure 3.14A). We also looked at gene expression and patient survival using the free online tool PrognoScan. PrognoScan uses publically available cancer microarray expression datasets to find the relationship between gene expression and patient prognosis and calculates the significance of the difference in outcome as corrected P values. The overall survival (OS) of the 3 genes, *CBLN1, HOXD13* and *DKK2,* was analysed in PrognoScan by selecting

cohorts of glioblastoma. The loss of expression of *HOXD13* and *DKK2* in the GSE7696 dataset was significantly associated with poor overall survival (corrected *P* value = 0.03, and 0.01, respectively). For *CBLN1*, overall survival was also analysed in a different glioblastoma cohort (MGH-glioma) and showed a significant association between loss of gene expression and poor outcome (corrected *P* value = 0.02) (Figure 3.14B). The association of DNA methylation/loss of expression of these genes with poor prognosis suggests that the identified genes may be potential markers for disease outcome and also potential candidates for further analysis.

SYMBOL	GENE NAME	TARGET ID	adj. *P*-value	RELATION TO CpG ISLAND	CHR
DKK2	dickkopf 2	cg00594011	0.02	Island	4
INA	internexin neuronal intermediate filament protein, alpha	cg00824018	0.02	Island	10
NR2F2	nuclear receptor subfamily 2, group F, member 2	cg01153166	0.02	Island	15
		cg04943986		Island	
		cg08370996		Island	
		cg09157727		Island	
		cg05155965	0.03	Island	
		cg06926934		Island	
		cg04330371		Island	
		cg05505872		Island	
		cg23042706	0.023	Island	
GRM6	glutamate receptor, metabotropic 6	cg12483476	0.02	Island	5
FAM190A	family with sequence similarity 190 A	cg17397028	0.02	Island	4
KHDRBS2	KH domain containing, RNA binding, signal transduction associated 2	cg26715952	0.02	Island	6
		cg22014661	0.03	Island	
HOXD8	homeobox D8	cg03858756	0.022	Island	2
CBLN4	cerebellin 4 precursor	cg20779964	0.022	Island	20
RAB6C	member RAS oncogene family	cg09676860	0.028	Island	2
HOXD13	homeobox D13	cg10418524	0.023	Island	2
CBLN1	cerebellin 1 precursor	cg16573782	0.023	Island	16
ISM1	isthmin 1 homolog	cg20081364	0.03	Island	20
HOXC4	homeobox C4	cg01473837	0.03	S_Shore	12
MSC	musculin	cg14546394	0.03	S_Shore	8
DPP10	dipeptidyl-peptidase 10	cg21678377	0.03	Island	2
TFAP2A	transcription factor AP-2 alpha	cg10129408	0.03	N_Shore	6
PDX1	pancreatic and duodenal homeobox 1	cg25685262	0.03	N_Shore	13
NEFM	neurofilament, medium polypeptide	cg04118306	0.035	Island	8
TBX18	T-box 18	cg04515996	0.04	Island	6
TPBG	trophoblast glycoprotein	cg08478195	0.04	Island	6
Hist3h2a	histone cluster 3, H2a	cg06034933	0.046	Island	1
PRR16	proline rich 16	cg27107171	0.046	Island	5
PRKCB	protein kinase C, beta	cg08406370	0.049	N_Shore	16

Table 3.2 List of 23 genes of the significantly differentially hypermethylated genes in

STS tumours

Gene name, probe ID, location in relation to CpG island and chromosome number are shown in the table along with adjusted p-value (p<0.05).

Ingenuity Analysis	Category	No. of Genes	p-value Range	Gene Names (Cancer genes in bold)
Diseases and Disorder	Cancer	10/23 (43.5%)	9.58E-05 - 4.93E-02	*INA, CBLN1, HOXC4, HOXD8, HOXD13, NR2F2, PDX1, PRKCB, TFAP2A, TPBG*
Molecular and Cellular Functions	Cellular Assembly and Organization	3/23 (13.0%)	5.30E-05 - 4.83E-02	*INA, NEFM, TPGB*
Molecular and Cellular Functions	Cellular Function and Maintenance	6/23 (26.1%)	5.30E-05 - 4.83E-02	*INA, NEFM, PDX1, PRKCB, TBX18, TPGB*
Molecular and Cellular Functions	Cell Death and Survival	2/23 (8.7%)	1.41E-03 - 4.70E-02	*PRKCB, TFAP2A*
Molecular and Cellular Functions	Cell Morphology	7/23 (30.4%)	1.41E-03 - 4.29E-02	*CBLN1, NEFM, DKK2, PDX1, PRKCB, TRAP2A, TPGB*
Molecular and Cellular Functions	Cellular Development	6/23 (26.1%)	1.41E-03 - 4.56E-02	*NEFM, PDX1, PRKCB, TBX18, TFAP2A, TPGB*

Table 3.3 Ingenuity analysis of the STS vs. LTS genes demonstrating the associated cancer genes and top 5 molecular and cellular functions
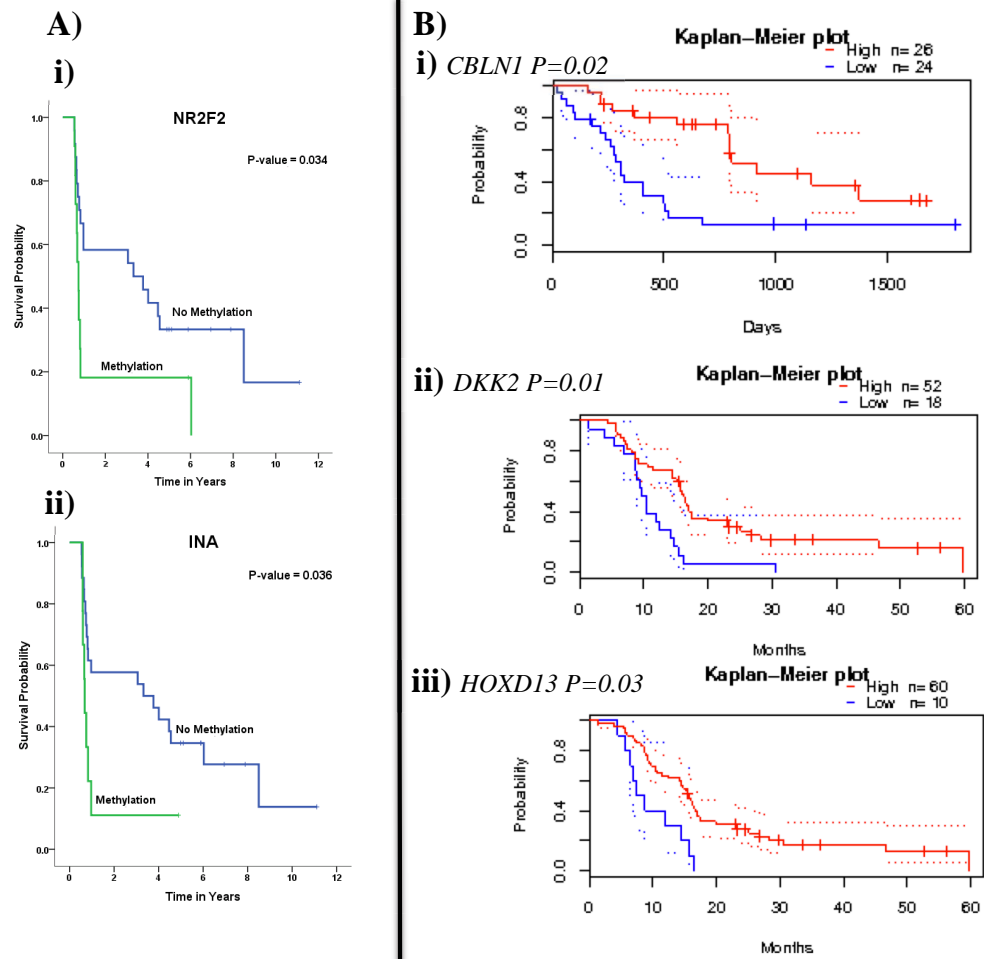
Figure 3.14 Kaplan Meier survival curves in association with A) gene methylation, B) gene expression

A) Curves show a strong association between CpG loci hypermethylation of *NR2F2* (i) and *INA* (ii) and poor patient survival. Both loci showed significance and *P* values are shown. B) Kaplan Meier survival plots show the association of gene expression and poor survival in *CBLN1* (i), *DKK2* (ii) and *HOXD13* (iii). Corrected *P* values are also shown in the figure.

3.3.5.2 Expression analysis using 5-azaDC treated cell lines

To confirm the possible biological relevance of DNA promoter hypermethylation, expression analysis was carried out in glioma cell lines pre- and post-treatment with the demethylating agent, 5-aza-2-deoxycytidine (5-azaDC) for 3 genes from Table 3.4. Hypermethylation of the top gene in the list, *DKK2,* was associated with loss of expression in glioma cell lines and showed re-expression in cell lines treated with 5-azaDC, thus functional validation of *DKK2* DNA methylation (Figure 3.15). The other 2 genes analysed, *TFAP2A* and *NR2F2,* also showed up-regulation in cell lines post 5-azaDC treatment.

3.3.5.3 Further confirmation of the methylation array data for *DKK2*

DKK2 CpG probe cg00594011 was located in the promoter region (1st exon) and CpG island of the gene. Bisulfite clone sequencing was carried out on two hypermethylated STS samples and two unmethylated LTS samples to confirm the array data for *DKK2*. Both hypermethylated STS samples, β value = 0.68 and 0.7, showed high probe values (0.88 and 1, respectively) and a high methylation index (MI = 56.5% and 60.3%, respectively). On the other hand, the two unmethylated LTS samples showed a low β value (0.07 and 0.1) and a low probe value (0.0 and 0.1, respectively) with a low methylation index of the region (MI = 3.9% and 10.6%, respectively) (Figure 3.16).
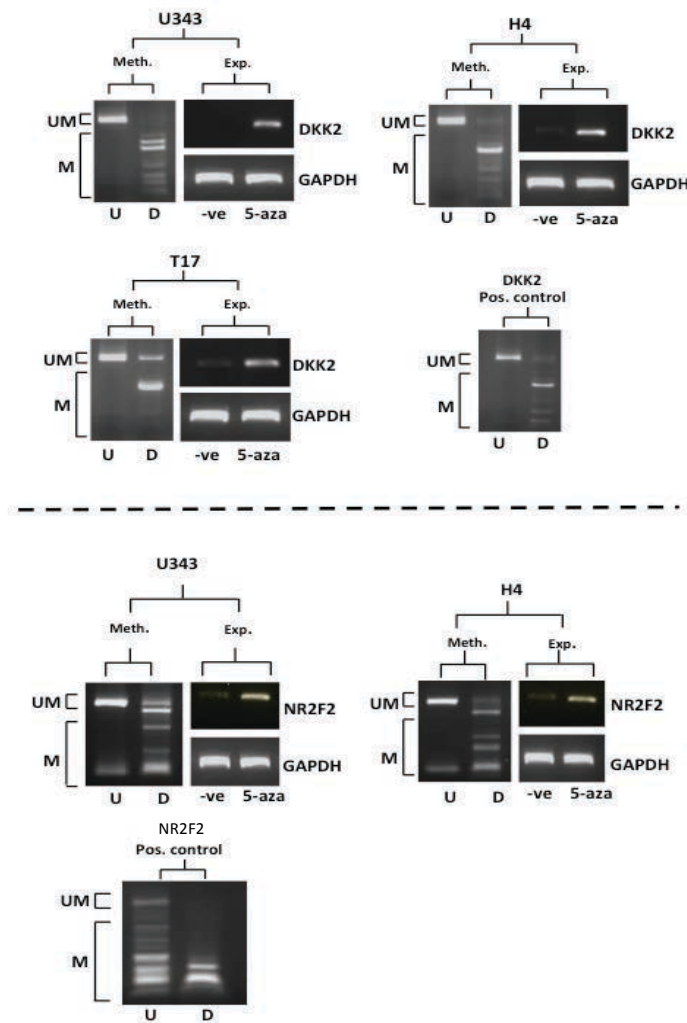
Figure 3.15 Expression analysis for DKK2 and NR2F2 in 5-azaDC treated cell lines

CoBRA results (Meth.) for *DKK2* and *NR2F2* are shown for selected glioma cell lines along with expression (Exp.) with (5-aza) and without (-ve) treatment. U, Undigested; D, Digested; M, methylated; UM, unmethylated.
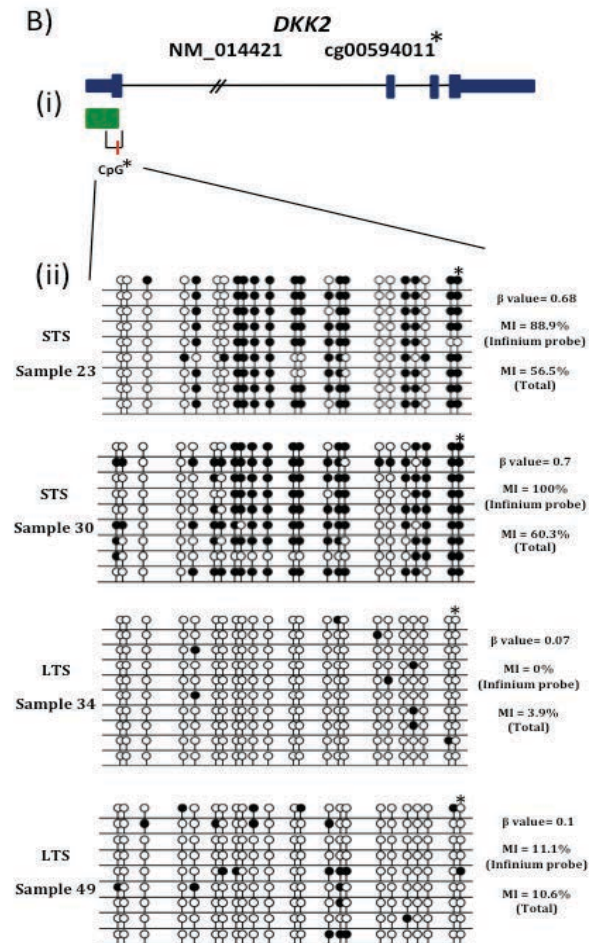
Figure 3.16 Confirmation of DKK2 methylation by clone sequencing

A schematic diagram of the gene (blue) showing the location of the CpG island (green) (i), Clone sequencing of the selected region, (ii) for each sample MI was calculated for the infinium CpG loci and for the whole analysed region.

3.4 Discussion

Glioma is the most common form of primary brain tumour. According to the World Health Organization (WHO), gliomas are classified into four grades. Grade IV tumours, glioblastomas, are the most aggressive form and have limited treatment options with a median survival time of 12 months after diagnosis (Barbus et al. 2011). However, a small subset (3%-5%) of glioblastoma patients show a longer survival time of >3 years. These patients are called long-term survivors (LTS) and they are characterised by a good Karnofsky performance (KPS), younger age, tumours with *MGMT* hypermethylation and almost no *EGFR* amplification (Krex et al. 2007; Hartmann et al. 2013).

The regulation of gene expression by DNA methylation is often altered in cancer. Generally, DNA hypermethylation at the promoter region represses the transcription of tumour suppressor genes, resulting in decreased gene expression, while methylation of the gene body is associated with oncogenes. Aberrant DNA methylation has been reported to be involved in glioma development. Some examples include DNA hypermethylation of the genes involved in apoptosis (*PYCARD*), cell migration genes (*TIMP3, CDH1* and *PCDH*) and DNA repair genes (*MGMT*) among others (Malzkorn et al. 2011).

In the past few years, many high throughput technologies have become available for genome-wide epigenetic studies including methylation arrays and next generation sequencing technology. Illumina methylation platforms for genome-wide DNA methylation analysis have been evolving since the release of the first methylation array, the 1,536 CpG Illumina GoldenGate DNA Methylation BeadArray followed by the Infinium HumanMethylation27 BeadChip, which was widely used in many studies

including those on cancer, aging and non-cancerous diseases. The most recent and largest platform from Illumina is the Infinium HumanMethylation450 BeadChip, which is able to assess the methylation level of more than 485,000 CpG sites. The HumanMethylation450K array is unique as it uses both Infinium I and Infinium II technologies where the Infinium I assay uses two separate bead types for each CpG probe, one for methylated C detection and one for the unmethylated or (T) detection, using one colour channel. On the other hand, the Infinium II assay uses one bead type to detect the methylated and unmethylated loci at the same time using two colour readouts (green/red) (Malzkorn et al. 2011). Then, the intensity of the signals is measured and calculated to generate the β value, which is the percentage of methylation represented in a range from 0 to 1 (corresponding to 0% and 100% methylation respectively) (Dedeurwaerder et al. 2011).

At the gene level, the HumanMethylation450K array covers ~99% of RefSeq genes and CpG loci distributed across the promoter region (200bp and 1500bp from transcription start site, 5'UTR and 1st exon), 3'UTR and gene body. One of the advantages of the technology is that it covers >96% of the CpG islands and areas adjacent to the islands (CpG shore; 2Kb away from the island and CpG shelf; 2Kb away from CpG shore) as well as independent CpG sites outside the islands (open sea). In this study, the latest genome wide HumanMethylation450K microarray was utilised to evaluate the methylation profiles of 19 short-term survivor and 16 long-term survivor glioblastoma patients and to gain an insight into promoter DNA methylation changes that may contribute to, or be indicative of, poor prognosis in STS glioblastoma patients.

3.4.1 Pre-processing the HumanMethylation450K array data

3.4.1.1 Divergence of Infinium I and Infinium II probes

After obtaining the array's raw data and in order to ensure an accurate analysis of DNA methylation array data in this study, stringent quality control analyses were applied at an early stage to exclude any possible faults or weaknesses that may result in invalid decisions, which may affect downstream analysis. Bibikova et al. noticed the technical differences between β values generated by Infinium I and Infinium II probes (Bibikova et al. 2011) where the distribution and value range in Infinium II were narrow, suggesting that Infinium II is less accurate due to using two different chemistry methods. Another evaluation study of the Infinium HumanMethylation450K array using HCT116 colon cancer cell line and its unmethylated derivative (DKO) cell line was carried out, where data from Infinium I and Infinium II were analysed separately to compare the performance of the two assays. β values from Infinium type II showed a smaller range compared to the type I probes which affected the values near to 0 and near to 1 β values (Dedeurwaerder et al. 2011). In this study it was suggested to treat the 450K array data as two different arrays and that the results from the two types of probe should be analysed separately. According to Illumina, the difference between the two chemistries should not affect the accuracy of the data, as the analysis should be performed on an individual site across all samples rather than comparing between different probes within a single sample. To overcome this limitation of the 450K array, many pipelines are available to re-scale and normalise β values generated from type II probes such as R packages: Illumina Methylation Analyzer (IMA) (Wang et al. 2012), Subset-quantile Within Array Normalization (SWAN) (Maksimovic, Gordon and Oshlack 2012) and LUMI (Du, Kibbe

and Lin 2008). To ensure an accurate analysis of the current study, data correction and normalization was used to correct the colour bias between Infinium I/II probes. To achieve this, a pipeline to process Illumina array, LUMI, from GenomeStudio software (Du, Kibbe, and Lin 2008) was applied. The IMA pipeline is able to correct the data including missing β values and then perform a quantile normalization to exclude the technical variations across the probes without major alterations in the data.

3.4.1.2 Detection p-value

Generally, the detection p-value is the intensity of a CpG probe compared to the intensity of an internal negative control. Several problems can be generated during the hybridization process, which may lead to faulty results. Problems such as incorrect signal reads, reads with low intensity, defects in the probe design, weak hybridization or mutations within the sample's DNA, are all translated as high detection p-values (Dedeurwaerder et al. 2013). Removing probes with a high detection p-value can reduce false-positive results, therefore in this study, more than 9,400 probes, showing a detection p-value of >0.01, were excluded from the analysis.

3.4.1.3 Imprinted genes, probes on X/Y chromosomes and probes over SNPs

Many sex-associated studies have discussed the methylation difference in X and Y chromosomes. Methylation levels can be affected by chromosome X inactivation or by having one X chromosome copy in males (Chen et al. 2013). To overcome this issue of gender bias, all probes located on X or Y chromosomes, or probes of imprinted genes were excluded from further analysis.

Polymorphic CpGs were also addressed as a problem in many studies, since the methylation array can detect naturally existing C/T SNPs in any sample. By matching the position of (C) of the Infinium array in the genome to the SNPs database in the 1000 Genomes Project database, Illumina has identified that 13.8% of total probes (9.4% of Infinium I probes and 15.5% of Infinium II probes) are polymorphic CpGs (Chen et al. 2013; Dedeurwaerder et al. 2013). For that reason, CpG with SNP were removed from the current study.

After applying this filtration process along with the individual validation of the β value, the analysis of the hypo/hyper methylation profile of STS and LTS tumours was conducted with confidence.

3.4.2 CIMP in a subset of LTS glioma

The term CpG island methylator phenotype, CIMP, has widely emerged as a distinct sub-classification in many types of tumour such as breast cancer, colorectal cancer and glioblastoma (Fang et al. 2011; Toyota et al. 1999; Noushmehr et al. 2010). The methylator phenotype was first observed in 1999 by Toyota et al. as massive hypermethylation of a specific group of loci in a subset of colorectal tumours (Toyota et al. 1999). Later, CIMP was extensively studied in colorectal cancer and other cancers using new technologies such as MethyLight and pyrosequencing. In glioma, CIMP was identified in a distinct subgroup and was tightly associated with *IDH* mutation (Noushmehr et al. 2010). Generally, IDH is required in the oxidative decarboxylation process of α-ketoglutarate (section 1.5.4.1). The arginine 132 residue (R132) in IDH1, and its analogue arginine 172 (R172) in IDH2, is functionally important for isocitrate

binding and converting the enzyme to α-ketoglutarate. However, mutated IDH1 drives the synthesis of 2-hydroxyglutarate which inhibits α-ketoglutarate dependant histone demethylases and leads to promoter hypermethylation (Figueroa et al. 2010; Xu et al. 2011).

Glioma specific *IDH1* mutations (R132) were found in >95% of IDH mutated glioma cases and they always occur as somatic, missenses mutations (Krell et al. 2011; Turcan et al. 2012). The mechanism of how the mutant IDH form leads to G-CIMP in glioblastoma is still unclear and raises questions. Turcan et al. demonstrated that *IDH1* mutation is the main cause of CIMP$^+$ in glioblastoma and thus DNA hypermethylation (Turcan et al. 2013). However, a few cases reported that CIMP could occur in wild-type *IDH* glioblastoma (Brennan et al. 2013; Hill et al. 2014). Interestingly, glioblastoma patients with the mutant *IDH* gene and that are positive for CIMP have shown a better prognosis and survival (Christensen et al. 2011; Turcan et al. 2012).

Promoter hypermethylation of *MGMT* is also widely reported in glioblastoma and has been associated with better response to treatment with alkyating agents. *MGMT* hypermethylation was first reported as a dominant observation in glioma cases (M. Esteller et al. 1999) and was later found to be associated with better treatment response (Manel Esteller et al. 2000). Many studies have found an association between CIMP status and *IDH* mutation and *MGMT* promoter hypermethylation. G-CIMP subgroups are always associated with better survival and are strictly linked with *IDH* mutation. *MGMT* hypermethylation was also associated with better prognosis, but was more of a general and prognostically favourable genome wide methylation profile with no close relation to

CIMP status (Bady et al. 2012; van den Bent et al. 2011). In the current study, cluster analysis showed massive hypermethylation in 5 LTS cases across a large number of CpG loci exhibiting the G-CIMP$^+$ phenotype. Only these 5 cases were mutated for *IDH1* (5/5, 100%), of which 60% (3/5) also demonstrated *MGMT* hypermethylation, which agrees with many recent studies (Noushmehr et al. 2010; Turcan et al. 2012), and fulfilled the criteria of G-CIMP$^+$ addressed in Noushmehr et al.'s paper. Our study identified 535 G-CIMP genes that were involved in many essential cell regulation processes and that had been reported previously in many cancer studies, suggesting biomarkers for the CIMP$^+$ phenotype in glioblastoma. Among the identified list of G-CIMP$^+$ genes, the cytokine signalling 3 gene (*SOCS3*), which was identified by 4 probes located within the promoter CpG island (Appendix 7.3), was confirmed to be statistically significantly hypermethylated in a group of LTS glioblastoma tumours as shown in a very recent study by Feng et al. (Feng et al. 2014). They were able to confirm the hypermethylation in an independent cohort using pyrosequencing and also in a TCGA glioblastoma cohort, which showed a significant association between DNA hypermethylation and longer survival, suggesting a novel marker for G-CIMP$^+$. This study agrees with our finding and is an indicator of the validity of the identified G-CIMP$^+$ gene list in our study.

3.4.3 Genes associated with poor prognosis in STS patients

Many genetic and epigenetic alterations have been associated with non-CIMP (or CIMP$^-$) glioblastoma formation and identified as predictors of poor prognosis. Patients with mutations in *LIG4, PTEN* and *BTBD2*, *EGFR* amplification and hypomethylation

changes in *LINE-1* had shown poor survival with a high risk of death at an early age (Malzkorn et al. 2011; Liu et al. 2010, 4). In addition, many studies have shown the effect of CpG island promoter hypermethylation in glioma malignancy, including the well-known *MGMT* role in glioma patients' treatment.

In this study, the investigation aimed to find differentially hypermethylated genes that may be responsible for poor prognosis in STS patients. The differential analysis was carried out between STS vs. LTS tumours. The above analysis led to the identification of 32 loci (representing 23 genes). These loci were located in CpG islands, shores or shelves and were associated with cancer, according to the IPA ingenuity analysis. Most of the genes within the list were represented by a single probe with the exception of *NR2F2* (nuclear receptor gene), which was represented by 9 CpG loci located in the CpG island and showed a strong correlation between gene hypermethylation and poor survival. Similarly, the disruption of *NR2F2* expression was previously reported in ovarian cancer and was associated with short survival (Hawkins et al. 2013). Another gene amongst our list of differentially methylated genes, the alpha internexin gene (INA) was recently investigated in pancreatic tumours, where reduced gene expression was related to metastasis and short overall survival in pancreatic cancer patients (B. Liu et al. 2014), and in lower grade glioma (Figarella-Branger et al. 2012). Also among the list of differentially hypermethylated genes were a number of the Homeobox gene family, *HOXD8*, *HOXD13* and *HOXC4*, which have been described in many cancers. Hypermethylation of HOXD genes has been reported previously in glioma (Wu et al. 2010), mantle cell lymphoma (Leshchenko et al. 2010) and HOXC aberrant expression in leukaemia and prostate cancer (Miller et al. 2003). The transcription factor (*TFAP2A*)

142

was found to be epigenetically silenced in prostate cancer and head and neck carcinoma (Bennett, Romigh, and Eng 2009). Also, the glutamate receptor gene *GRM6* was identified in clear cell renal cell carcinoma as a CIMP marker in ccRCC (Arai et al. 2012). CBLN proteins are expressed throughout the cerebellum and genes from the CBLN family were found to be involved in many regulatory signalling pathways and receptors (Wei et al. 2012). Hypermethylation of *CBLN1* and *CBLN4* had not been reported prior to this study in any type of cancer. All these genes were hypermethylated in the short-term survival patients, suggesting a prognostic significance. The Dickkopf-2 gene (*DKK2*) is a negative regulator of the Wnt/β-catenin signalling pathway. Studies have reported that *DKK2,* a homolog of *DKK1*, accelerates tumour cell growth and positively induces angiogenesis (DKK1 negatively regulates angiogenesis) (Park et al. 2014, 1). Also, *DKK2* is a target gene for multiple micro-RNAs which activate the Wnt/β-catenin signalling pathway and promote tumourigenesis such as MicroRNA-222 and MicroRNA-21 in oral cancer cell invasion (Li et al. 2013; Kawakita et al. 2013). Furthermore, *DKK2* was reported as a pro-metastatic gene in Ewing sarcoma and overexpression was associated with high invasiveness and differentiation in Ewing sarcoma cells (Hauer et al. 2013). The current study confirmed that promoter hypermethylation of *DKK2* was significantly associated with loss of expression which was also associated with poor prognosis and worse overall survival in STS tumours. Our study suggests that *DKK2* methylation should be analysed further to assess its potential as a prognostic marker and may be a diagnostic marker in the assessment of glioma patients.

3.4.4 Conclusion

The aims of this study were to investigate the methylation profile of STS and LTS glioblastoma tumours and to identify differentially methylated genes that may be associated with worse patient prognosis using the latest technology from Illumina, the Infinium HumanMethylation450 array. This 450K methylation array is a cost effective method for epigenome wide analysis that utilises a large number of patient samples. This platform has more advantages compared to the old methylation platforms (27K array and GoldenGate array). The significantly increased number of probes in the HumanMethylation450 array allows the analysis of >96% of CpG sites in more than 99% of RefSeq genes in addition to CpGs outside the island regions such as CpG shores and CpG shelves which have been shown to be associated with gene expression.

The generated list of differentially methylated genes consisted of novel and known genes that were hypermethylated in STS tumours. Further work is required to confirm these findings in a separate, larger cohort, and to investigate the biological and functional relevance of these genes in glioma development. Some of the genes, such as *DKK2, NR2F2* and the transcription factor *TFAP2A*, are good candidates for additional analysis in grade IV glioblastoma to investigate whether these genes would be candidates as markers for disease outcome. Also, it would be beneficial to study the hypermethylated genes identified in CIMP[+] LTS glioblastoma tumours to gain an insight into the role of *IDH* mutation and CIMP in gliomagenesis, which would later be helpful in cancer therapies.

In summary, since the molecular genetic and epigenetic factors behind long-term and short-term survival in glioblastoma are still unclear, this study has used the latest

available technology to provide needed insight into the methylation profiles of STS and LTS glioblastoma.

This work was presented at The *Epigenetics* conference (part of genomic research Europe) in Frankfurt, Germany as a poster presentation on 4th – 5th September 2012.

Chapter Four: Epigenetic Targets in Breast Cancer Brain Metastasis

4.1 Introduction

Brain metastasis is the most common type of central nervous system cancer with an incidence of 10:1 of brain metastasis to primary brain tumours, while lung, liver, skin or breast are usually the primary site for brain metastasis. Following on from the hypothesis of "seed-and-soil", 15% of breast cancer patients develop metastasis in the brain where breast tumour cells prefer to disseminate, invade and adapt in the brain microenvironment. Therefore, brain metastasis usually indicates a patient's poor prognosis with a median survival time of <8 months. Consequently, it is essential to understand the molecular mechanisms behind brain metastasis development from breast cancer (BCBM) and to identify the markers responsible for predisposing brain metastasis formation.

Unlike tumour suppressor genes, metastasis suppressor genes are not involved in cell growth and proliferation; however, they regulate the processes required for cells' dissemination, invasion and adaptation at the distant sites. Studies have shown that these genes are rarely mutated and other mechanisms, for instance epigenetic changes, may be responsible for the dysregulation of metastasis suppressor genes.

Loss of gene expression due to DNA hypermethylation is a frequent event in cancer. The HumanMethylation450K array is the most appropriate method for investigating the methylation events within regulatory regions, such as gene promoters, in cancer samples. Probes located within CpG island proximity, and within the gene's transcription start sites, 5'UTR, and 1st exon can be selected to ensure relevance between DNA

hypermethylation and loss of expression. In this preliminary study, the HumanMethylation450K array was utilised to identify potential signature genes for early prediction of breast cancer brain metastasis (BCBM) development.

Aims of the study

- To investigate DNA methylation profiles in BCBM samples and to identify gene(s) that may play a role in brain metastasis formation from breast cancer.

- To compare the methylation level of these genes in brain metastasis samples from other primary organs, i.e. lung tumours and melanoma.

- To identify genes where hypermethylation/loss of expression is associated with patients' outcome.

4.2 Results

4.2.1 Methylome of breast tumour vs. brain metastasis

4.2.1.1 Cluster analysis

In this study, methylation profile analysis of breast cancer brain metastasis (BCBM) was conducted by utilising HumanMethylation450K BeadChips on 5 brain metastasis (BM) derived from breast tumours. The HumanMethylation450K array data of breast tumours without metastasis (n=28) were obtained from the TCGA online data portal (batch 96) which were selected based on the criteria of having matched normal sample and having never formed metastasis to distant organs (by October 2012). Appendix 7.5 contains patients' clinical information.

To gain an insight into the overall methylation pattern in BCBM and non-metastatic breast tumours, unsupervised Euclidean hierarchical clustering was conducted on the most 1000 variable loci from the combined data of 5 BM from breast cancer and non-metastatic breast tumours (n=28) from the TCGA data portal (Figure 4.1). The clustering separated breast tumours (without metastasis) from BCBM samples and divided them into two major groups. The first group contains 6 samples, which were split into 2 subgroups; one group contains 4 of the BM samples (67%), while the other group contains 2 of the TCGA samples (33%). The second major group contains the remaining TCGA samples (n=26, 96%) and one of the BM samples. Samples within the first group, which mainly contains BM samples, demonstrate the intermediate methylation level (mean β-value range 0.32) of the most 1000 variable loci, while the second group showed a high overall methylation level (mean β-value range 0.64). This observation indicates that the methylation profiles of the breast tumours without metastasis differ from the methylation profiles of BCBM.

4.2.1.2 CpG loci distribution in the genome

Following data assembly from GenomeStudio Illumina software, the raw data were normalized. A filtration process was then applied as follows: (i) probes with an unreliable detection p-value (>0.01) were removed as this reflects poor hybridization according to Illumina quality control (Dedeurwaerder et al. 2013) and (ii) probes for imprinted genes or overlapping SNPs were excluded from further analysis (Dedeurwaerder et al. 2013). Following the filtration process, probes showing β-value >0.6 in >60% of the tumour

samples were considered hypermethylated, while hypomethylation was considered if >60% of tumour samples had β-value <0.2.

In order to obtain an insight into the methylome of breast tumours (without metastasis) and brain metastasis derived from breast tumours, analysis of hypermethylated and hypomethylated CpG loci in the array with respect to gene function and features (Figure 4.2A) and in relation to CpG island (Figure 4.2B) was carried out. The majority of the hypermethylated probes in TCGA breast tumours (without metastasis) and breast cancer brain metastasis (BCBM) samples were located in the body of the genes, 48% and 50% respectively, and 20% and 19%, respectively, were located in the promoter regions. Distribution of the hypermethylated CpG loci in relation to the genomic location was also similar between breast tumour and BCBM samples, where the majority reside in CpG islands (39% and 38 respectively). Furthermore, 21% and 20% of the loci resided in CpG shores (up to 2Kb away from CpG islands) and 14% in CpG shelves (2-4Kb away from CpG island) respectively. This is consistent with other findings that many alterations in DNA methylation occur within the adjacent regions of CpG islands (Reyngold et al. 2014). Analysis of hypomethylated loci was also carried out and the majority of the probes in TCGA non-metastatic breast tumours were located in the promoter regions (70%) as well as in BCBM samples (66%) (Figure 4.2B). With respect to genomic location, CpG loci were roughly equally distributed between CpG islands (38% and 35% respectively), shores (22% and 32% respectively) and open sea (39% and 32% respectively) in non-metastatic breast tumours and BCBM samples. The above data indicates that breast tumours without metastasis and BCBM show a similar distribution of methylated probes in relation to gene features and the genomic location of CpG probes.
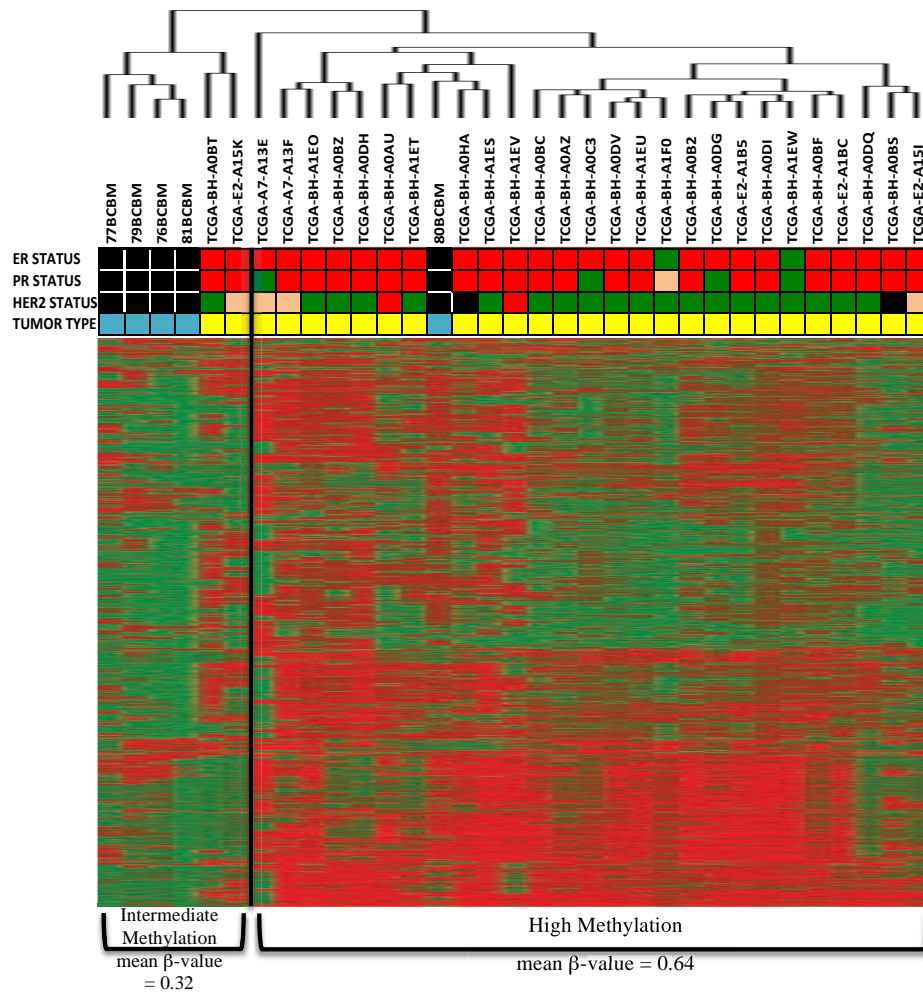
Figure **4.1 Unsupervised clustering of the most 1000 variable CpG loci in 5 BCBM samples and 28 non-metastatic breast cancer samples from the TCGA data portal**

CpG loci are colour coded according to the methylation status (hypermethylation, red; low methylation, green). The type of sample is shown above the clustering (blue, brain metastasis sample; yellow, TCGA breast cancer samples) along with the status of the Estrogen receptor (ER), Progesterone receptor (PR) and HER2 receptor (positive, red; equivocal, orange; negative, green; not available, black) in breast tumour samples. The two groups are separated by a black line and the mean β-value of each group is shown below each group.
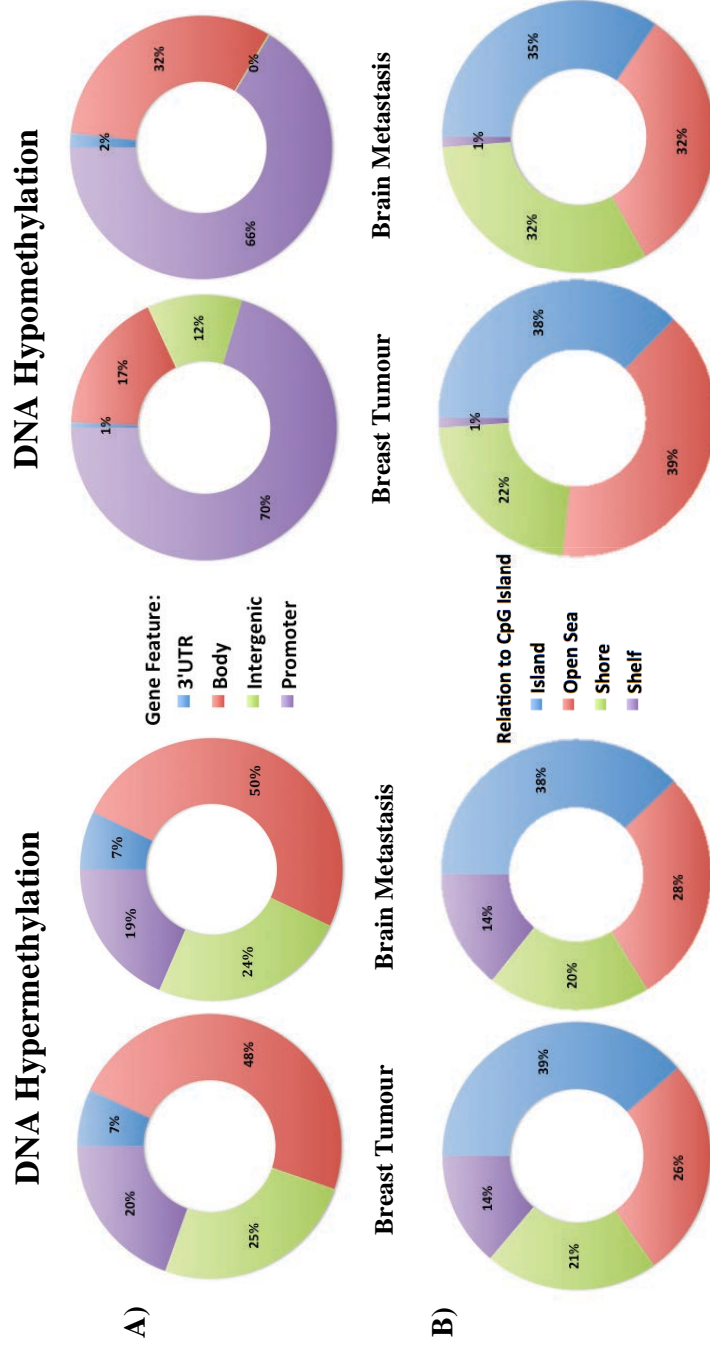
Figure 4.2 Hyper/Hypomethylated CpG loci distribution in non-metastatic breast tumours and BCBM samples with respect to A) Gene feature; 3' UTR, gene body, intergenic and promoter regions. B) Relation to CpG island; island, open sea, shore and shelf

4.2.2 Identification of hypermethylated signature genes in BCBM

In order to identify signature genes in BCBM, differential methylation analysis between 5 BM and 28 non-metastatic breast tumours was conducted on CpG loci located within gene promoter regions (200bp and 1500bp from TSS, 5' UTR, and within 1st exon) to enrich the biological relevance between DNA methylation and gene silencing. Using the criteria of β-value <0.2 methylation in >60% non-metastatic breast tumour samples and β-value >0.6 methylation in >60% of BM samples in addition to *P* value below 0.05, 37 CpG loci (representing 28 genes) were identified (Table 4.1).

To assess the function and types of gene within this panel, the DAVID functional analysis tool (http://david.abcc.ncifcrf.gov/) was used to assess the properties of the genes. The functional annotation tool identified tumour suppressor genes (2 genes), genes involved in cell differentiation (5 genes), cell migration and motility (3 genes), regulation of cell death (3 genes), cell adhesion (1 gene), regulation of cell proliferation and division (3 genes), and regulation of transcription (9 genes) (Figure 4.3). Ingenuity Pathway Analysis (IPA) of the list showed that 24 genes (86%) were associated with cancer. Molecular and cellular functional analysis of the list divided the gene list into 9 functional categories that were involved in cell death and survival (11 genes; 39%), cell-to-cell signalling (7 genes; 25%), cell cycle (2 genes; 7%), cellular growth and proliferation (9 genes; 31%), cellular movement (10 genes; 36%) and cell mediated immune response (3 genes; 11%) (Table 4.2). Taken together, the gene list produced was considered for further analysis.

	Gene	Probe ID	P-value	CHR	Gene Feature	Relation to CpG Island
Genes with multiple probes	EOMES	cg22383888	0.019	3	TSS1500	Shore
		cg06014401	0.002	3	TSS1500	Island
	MIR125B1	cg24603444	5.51E-07	11	TSS1500	Open sea
		cg06749053	0.002	11	TSS1500	Open sea
	WDR69	cg00063174	9.05E-06	2	TSS200	Island
		cg20647610	0.0001	2	TSS200	Island
	NTM	cg20881910	0.006	11	TSS1500	Island
		cg15617814	2.12E-06	11	TSS1500	Island
	TRIM68	cg03122735	0.002	11	5'UTR	Island
		cg01719157	0.006	11	TSS200	Shore
	WDR63	cg26012941	0.001	1	TSS1500	Open sea
		cg02213139	0.033	1	TSS1500	Open sea
		cg15825692	0.013	1	TSS200	Open sea
		cg12062099	0.036	1	TSS1500	Open sea
	DLX6AS	cg00337916	0.0003	7	TSS1500	Shelf
		cg25707676	0.0003	7	TSS200	Shelf
Genes with single probe	AMBRA1	cg24374161	0.001	11	5'UTR	Open sea
	CLIC6	cg17935217	0.023	21	1stExon	Island
	CREB5	cg23281552	0.010	7	5'UTR	Open sea
	CXXC5	cg03992638	0.0003	5	5'UTR	Shelf
	DZIP1	cg22991101	0.004	13	5'UTR	Island
	FAM198B	cg03450635	0.005	4	TSS200	Open sea
	HOXD13	cg10418524	0.046	2	TSS1500	Island
	MAP3K14	cg18700940	2.57E-06	17	5'UTR	Open sea
	MEF2D	cg10128479	0.0111	1	5'UTR	Shelf
	NCALD	cg07056138	0.0042	8	5'UTR	Open sea
	NR3C1	cg26720913	0.0010	5	1stExon	Open sea
	PHOX2B	cg15542784	0.0179	4	1stExon	Shore
	PIK3IP1	cg12273325	5.37E-05	22	TSS1500	Shelf
	PRDM13	cg00126261	0.001	6	TSS1500	Shore
	PNF170	cg01981354	9.88E-05	8	5'UTR	Shelf
	SERPINA5	cg00917847	0.0002	14	TSS200	Open sea
	SOX1	cg10245273	0.0006	13	1stExon	Shore
	TLR5	cg23035715	0.0431	1	5'UTR	Open sea
	WIPF1	cg27117828	1.13E-05	2	5'UTR	Open sea
	ZF323	cg07369507	1.65E-10	6	TSS1500	Open sea
	ZNF365	cg03961010	1.30E-07	10	5'UTR	Open sea

Table 4.1 List of 37 CpG loci identified that were frequently methylated in BCBM

Twenty-eight hypermethylated genes in BM were identified using the Infinium 450K array. Illumina probe ID, percentage of methylation in BCBM and in non-metastatic breast tumours, P-value, chromosome, gene feature with promoter region and probe in relation to CpG island were indicated for each probe.

Tumor Suppressor Genes (2 genes)
SOX1 DZIP1
Cell Differentiation (5 genes)
DZIP1 AMBRA1 EOMES MEF2D PHOX2B
Cell Migration and Motility (3 genes)
SOX1 PRDM13 PHOX2B
Regulation of Cell Death (3 genes)
MEF2D MAPK13 NR3C1
Cell Adhesion (1 gene)
NTM
Regulation of Cell Proliferation and Division (3 genes)
HOXD13 NCALD PRDM13
Regulation of Transcription (9 genes)
PHOX2B ZNF323 PRDM13 SOX1 CREB5 EOMES HOXD13 MEF2D NR3C1

Figure 4.3 **David functional analysis of hypermethylated signature genes in BCBM**

The identified hypermethylated genes in BCBM were analysed by the DAVID functional analysis tool and categorised based on relevant function to tumorigenesis. The number of genes in each functional group is shown within parentheses.

Ingenuity Analysis	Category	No. of Genes	p-value Range	Gene Names (Cancer genes in bold)
Diseases and Disorder	Cancer	24/28 (86%)	1.34E-03 - 3.83E-02	*AMBRA1, CLIC6,CREB5, CXXC5, DZIP1, WDR69, EOMES, FAM198B, HOXD13, MAP3K14, MIR125B1, NCALD, NR3C1, NTM, PHOX2B, PIK3IP1, SERPINA5, SOX1, TLR5, WIPF1, ZNF323, ZNF365*
Molecular and Cellular Functions	Cell-To-Cell Signalling	7/28 (25%)	3.25E-04 - 4.09E-02	*MAP3K14, MIR125B1, NR3C1, NTM, PHOX2B, SERPINNA5, TLR5*
Molecular and Cellular Functions	Cellular Development	11/28 (39%)	9.53E-04 -4.09E-02	*AMBRA1, EOMES, MAP3K14, MEF2D, MIR125B1, NR3C1, PHOX2B, PIK3IP1, SERPINA5, TRIM68, WIPF1*
Molecular and Cellular Functions	Cellular Growth and Proliferation	2/28 (7%)	1.34E-03 - 3.92E-02	*MIR125B1, NR3C1*
Molecular and Cellular Functions	Cell Death and Survival	9/28 (32%)	1.34E-03 - 3.89E-02	*EOMES, MAP3K14, MIR125B1, NR3C1, PHOX2B, SERPINA5, SOX1, TLR5, WIPF1*
Molecular and Cellular Functions	Cellular Assembly and Organization	10/28 (36%)	2.50E-03 - 3.44E-02	*CXXC5, EOMES, MAP3K14, MIR125B1, PHOX2B, PIK3IP1, SERPINA5, SOX1, TLR5, WIPF1*
Molecular and Cellular Functions	Cell Death and Survival	7/28 (25%)	1.34E-03 - 4.21E-02	*EOMES, MAP3K14, NR3C1, PIK3IP1, SOX1, TLR5, WIPF1*
Molecular and Cellular Functions	Cellular Assembly and Organization	3/28 (11%)	1.34E-03 - 4.21E-02	*NR3C1, SERPINA5, WIPF1*
Molecular and Cellular Functions	Cell Death and Survival	3/28 (11%)	1.34E-03 - 2.65E-02	*EOMES, MAP3K14, TLR5*
Molecular and Cellular Functions	Cellular Assembly and Organization	11/28 (39%)	1.34E-03 - 3.89E-02	*CREB5, DZIP1, EOMES, MAP3K14, MEF2D, MIR125B1, NR3C1, PHOX2B, SOX1, TLR5, WIPF1*

Table 4.2 **IPA Ingenuity Analysis of the 28 hypermethylated genes in BCBM**

IPA analysis demonstrating the associated cancer gene and top molecular and cellular functions.

4.2.2.1 Analysis of genes with multiple probes

Within the list of candidate genes, seven genes (*EOMES*, *WDR69*, *MIR125B1*, *NTM*, *TRIM68*, *WDR63* and *DLX6AS*) were represented by more than one probe (Table 4.1). *DLX6AS* was excluded from further analysis because it is a non-translated gene. Analysis was carried out by confirming the methylation data of the remaining 6 genes with a further set of 15 BCBM tumours using CoBRA to amplify the region containing both of the methylated CpG sites. BCBM tumours demonstrated high methylation levels for *EOMES* (71%; 10/14), *WDR69* (93%; 14/15), *MIR125B1* (50%; 7/14), *NTM* (67%; 10/15) and *WDR63* (53%; 8/15) (Figure 4.4). (Appendix 7.6 shows raw CoBRA data for this analysis). Hence, we have confirmed the HumanMethylation450K array data using our in-house CoBRA assay.

The methylation status of *TRIM68* in brain metastasis could not be obtained by CoBRA analysis, as there were no restriction enzyme sites that include a CpG dinucleotide around or within the two CG probes contained on the array; hence this gene was excluded from further analysis.

4.2.2.2 Analysis of genes with a single probe

Four genes (*DZIP1*, *SOX1*, *PRDM13* and *PHOX2B*) were selected from the remaining 21 genes with a single probe in the list to be further tested on 15 BCBM. These 4 genes were selected because of: (i) their location within the CpG island, shore or shelf (regions with a strong relation to gene expression) (Irizarry et al. 2009), (ii) restriction enzyme cutting

site was within the CG array probe, and (iii) within the list identified by the DAVID functional analysis tool (section 4.3.2).

As expected, these genes were also hypermethylated in the new set of BCBM samples analysed by CoBRA at frequencies of 9/15 (60%) for *DZIP1*, 9/12 (75%) for *SOX1*, 9/14 (64%) for *PRDM13* and 11/15 (73%) for *PHOX2B* (Figure 4.4).

Furthermore, to validate that the selected 9 genes from the above analysis (5 with multiple probes (*EOMES*, *WDR69*, *MIR125B1*, *NTM*, *WDR63)* and 4 (*DZIP1*, *SOX1*, *PRDM13* and *PHOX2B)* with single probes) are unmethylated in non-metastatic breast tumours, a further set of non-metastatic breast tumour samples from the TCGA data portal (n=18) was selected and their β-values were examined. These 9 genes remained unmethylated (β-value <0.2) in >60% of the new set of breast tumours without metastasis. Since we were able to confirm the data in a further subset of samples and with all previous studies validating the HumanMethylation450 array (Roessler et al. 2012; Hill et al. 2014), analysis was continued with confidence for the methylation analysis of matched pairs of breast tumours and corresponding brain metastasis.

Gene	Probe ID	Methylation %	Relation to CpG island
EOMES	cg22383888		Shore
	cg06014401	71%	Island
WDR69	cg00063174		Island
	cg20647610	93%	Island
MIR125B1	cg24603444		Open sea
	cg06749053	50%	Open sea
NTM	cg20881910		Island
	cg15617814	67%	Island
	cg26012941		
WDR63	cg02213139		Open sea
	cg15825692	53%	Open sea
	cg12062099		
DZIP1	cg22991101	60%	Island
SOX1	cg10245273	75%	Shore
PRDM13	cg00126261	64%	Shore
PHOX2B	cg15542784	73%	Shore

Table 4.4 Validation of highly methylated probes on a further set of BCBM samples

Highly methylated probes in BM were further tested on a second set of BM (n=15). The result from CoBRA product digestion is indicated with a red circle (digested, methylated) or a green circle (undigested, unmethylated. Black circles represent that no PCR product for the particular sample was amplified. Frequency (%) of methylation in brain metastasis is shown above for each gene.

4.2.3 Methylation analysis of selected genes in paired breast tumour/brain metastasis samples

To assess the specificity of the DNA methylation of these 9 genes to metastatic breast tumours, DNA methylation analysis was carried out in a cohort of matched breast tumours and brain metastasis paired samples (up to 10 pairs). Using the CoBRA PCR assays described and validated in the previous section, 7 of the 9 genes (*EOMES, WDR69, MIR125B1, DZIP1*, *SOX1*, *PRDM13* and *PHOX2B*) demonstrated a high methylation level in breast tumours with BM, and also in the matching BM, compared to non-metastatic breast tumours from TCGA (Figure 4.5).

EOMES, WDR69 and MIR125B1 were hypermethylated in 70% (7 of 10), 100% (8 of 8), and 71% (5 of 7) of the breast tumours with brain metastasis, respectively (Figure 4.5A, B, C). Also, the single probed genes, *DZIP1*, *SOX1*, *PRDM13* and *PHOX2B,* showed a high level of methylation in 50% (4 of 8), 75% (6 of 8), 100% (6 of 6) and 83% (5 of 6), respectively, in the breast tumours (Figure 4.5D, E, F, G). These results suggest that these genes may play an early role in BCBM formation. The CoBRA results for the other 2 genes, *NTM* and *WDR63*, showed a random methylation pattern between primary breast tumours and matching brain metastasis samples with no specific direction or selection for an early or late metastasis event. The data for *NTM* and *WDR63* suggests that these genes may be passengers for methylation changes in BCBM.
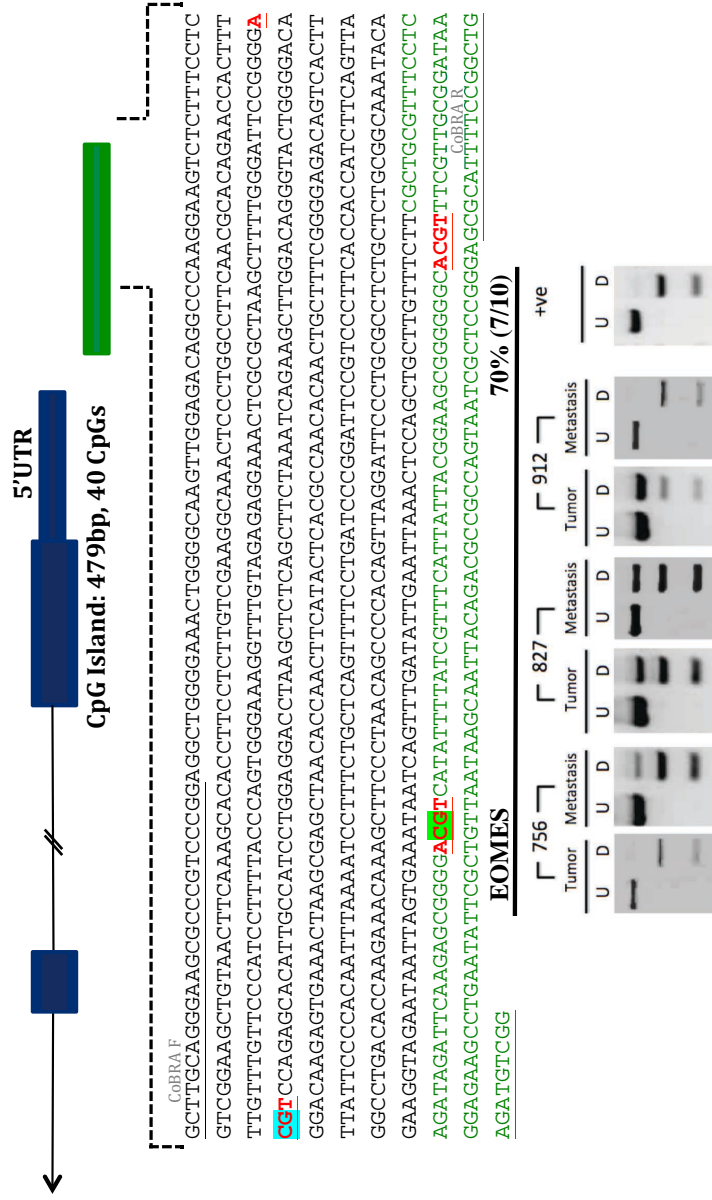
Figure 4.5A A schematic diagram of part of *EOMES* with CpG island location within the promoter region (green)
The amplified region by the CoBRA assay is indicated by dotted lines which contains the two probes, cg22383888 (blue) in CpG Shore and cg06014401 (green) in CpG island (green text), along with CoBRA forward and reverse primers (underlined sequences), and *HpyCH4IV* restriction sites (underlined red sequence). CoBRA digestion results are shown below the diagram for selected breast tumours and matching brain metastasis. In each case, undigested products (U) and digested (D) products run next to each other along with a positive control. The frequency of methylation in metastatic breast tumours is shown next to the gene name. CoBRA for *EOMES* in paired samples was kindly done by Rajendera Pangeni (School of Biology, Chemistry and Forensic Science, University of Wolverhampton).
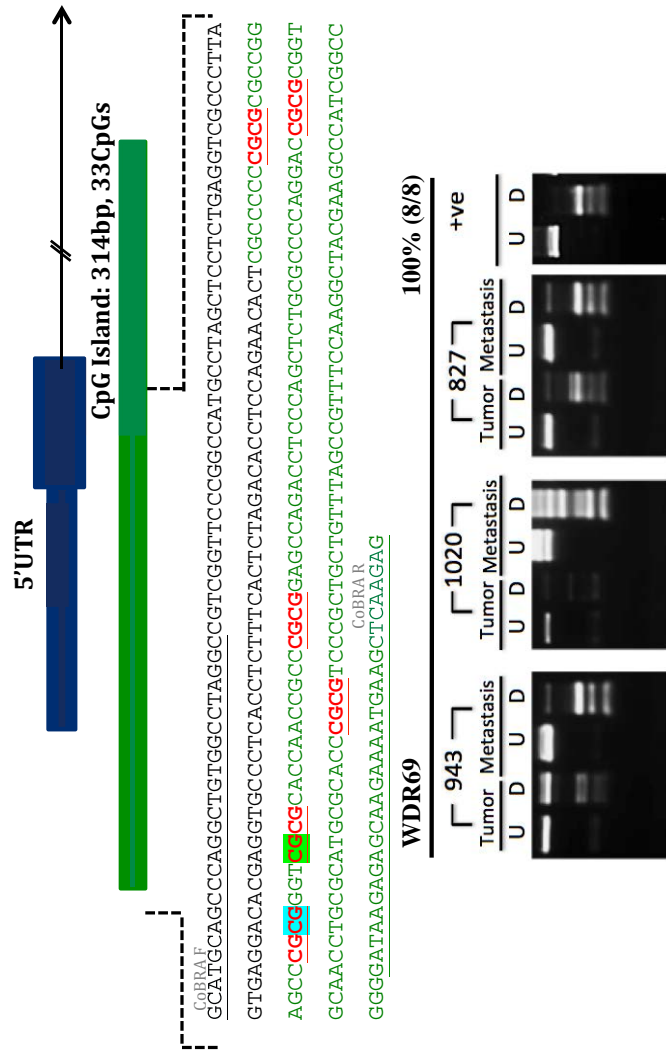
Figure 4.5B A schematic diagram of part of *WDR69* with CpG island location within the promoter region (green)

The amplified region by the CoBRA assay is indicated by dotted lines which contains the two probes, cg0063174 (blue) and cg20647610 (green) in CpG island (green text), along with CoBRA forward and reverse primers (underlined sequences), and *BstUI* restriction sites (underlined red sequence). The CoBRA digestion results are shown below the diagram for selected breast tumours and matching brain metastasis. In each case, undigested products (U) and digested (D) products run next to each other along with a positive control. The frequency of methylation in metastatic breast tumours is shown next to the gene name.

MIR125B1

CoBRA F

GGCTATAGAAGTGTAAAGACGAATAGCGTTCTGAACTCACCCAT**TCGA**AGCGGAGCGGGCCCGCACAGCCTCCTGTGGCACCA
TC**CCGA**GCAAGGGGGAAAATGAGAGCTCTCAGTGCGCTAGGCTGGGCGAGGTGTTCAATAGAGGACGGAGAACGGGGGCGGG
GAGGGGAGG<mark>CG</mark>CGGGGCAGGAAGCGATACACCAAATC**TCGA**GACTGTAACTCTGTAGCTTATGCTTTGACGAGAGAAAAAGGGC
AGTTGGGTGGGAGTGGAGGGAGGGGCTGGAACTGATCAAATTGGAAAATATTCACTTAAAAATTAAAAAAAGGCGGGGGGGGG
GGGGGAGTATGAAACCTCCCCAGAGCTGCTTTTTGAAAAATTTCTTCCCTTCTTTCTTTAGCGTGCACTTTGCCA_GTCCTGCAAA
_CGGAGGCTTCGTCTGCATCAG

CoBRA R

MIR125B1

71.4% (5/7)
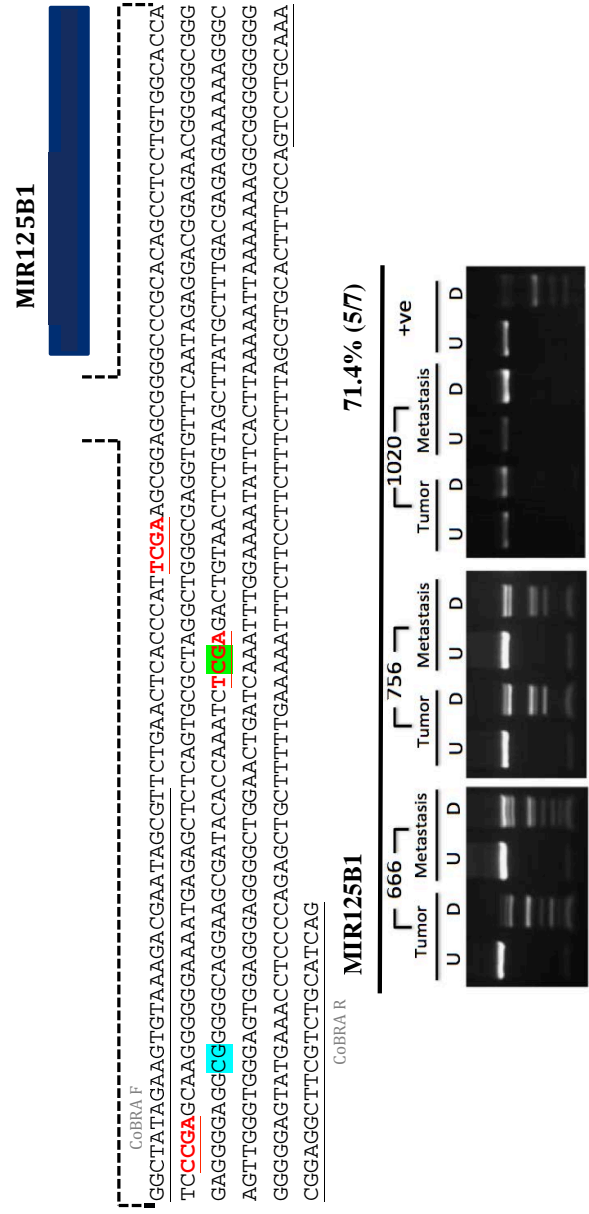
Figure 4.5C A schematic diagram of part of *MIR125B1* with CpG island location within the promoter region (green)

The amplified region by the CoBRA assay is indicated by dotted lines which contains the two probes, cg06749053 (blue) and cg24603444 (green) as isolated CpGs within 1500bp from TSS, along with CoBRA forward and reverse primers (underlined sequences), and *Taq1* restriction sites (underlined red sequence). The CoBRA digestion results are shown below the diagram for selected breast tumours and matching brain metastasis. In each case, undigested products (U) and digested (D) products run next to each other along with a positive control. The frequency of methylation in metastatic breast tumours is shown next to the gene name.

5'UT

CpG Island: 916bp,

CoBRA F
TTGCATAGGAGGAGGACCGGGGCGGGTCTTTACCCAGCCTGGGCGCGCCTCCCGGCCGCCCGCGCCCACAGCCCTCAGG
AGCGGGAGAAGGCCGGGTTCCTCGCTTCCGCGCGCGGCGGGGCCTAAGGTCTGGGCGTCCAGGACGTTGCTAT
AGCAGCGCCCAGGTCCGGACCTGGAGCAAAGGGCGCGTCTGCCGCAGGGTCAGGGTCAGCCCAGAACCCT
GCGGGACCAGAGAAGACCTCTTGGACGAGCCCGAGTGGACACCGCGTTCACCCGAAGAGAGGGTACCTCCTGGG
TACCGGCTCCCCAGGGGGCCACCCTCCCTGGCGTTTGGGGCTGGACACTCCATGTCCCTAAGGCAGGGCGTGGGG
AGCC
CoBRA R

DZIP1

50% (4/8)

912			827				756				+ve	
Tumor		Metastasis	Tumor		Metastasis		Tumor		Metastasis			
U	D	U D	U	D	U	D	U	D	U	D	U	D
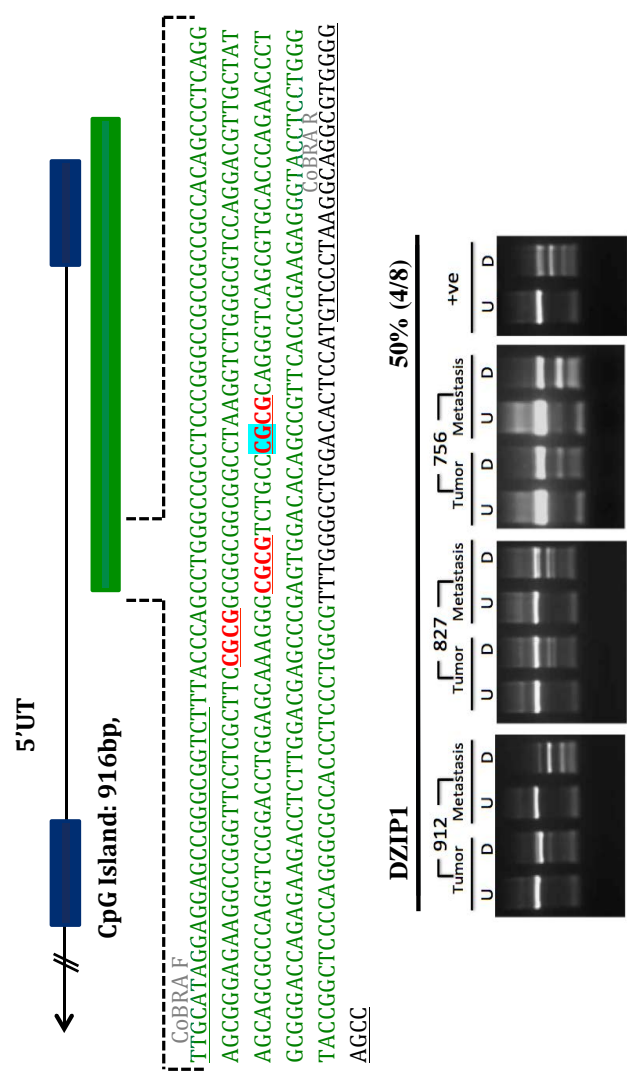
Figure 4.5D A schematic diagram of part of *DZIP1* with CpG island location within the promoter region (green)
The amplified region by the CoBRA assay is indicated by dotted lines which contains the CpG probe cg22991101 (blue) in CpG island (green text), along with CoBRA forward and reverse primers (underlined sequences), and *BstUI* restriction sites (underlined red sequence). The CoBRA digestion results are shown below the diagram for selected breast tumours and matching brain metastasis. In each case, undigested products (U) and digested (D) products run next to each other along with a positive control. The frequency of methylation in metastatic breast tumours is shown next to the gene name.

CpG island: 3018bp, 274CpGs

^{CoBRA F}
GGTCGAATGGGACCTTGAGGGTTTTCTCTGAAATATACAAACTTAAAGGACTCTCTCGAGGTTCTTTGACTGA**ACG**
TCCACTCTCAGTCTGCGCCCTGCTCCCCTGTGTCCCCTGTGCTGAGTTTCTGTGTCCAATTGTTGGCATCTAGGTCT
TGGCTCAAGATTAGGATGTGGGCCCCACTTTAGAGGCACAGACTATGAAAAGCTGAGTTAGTGCGCCCGGG**ACG**CC
AGGCAAGCAGCTTTTACAGTTTGGCATCTTATTGCCAGTGCTTCGTGCACAGTCAGCTGAAATAGCCAATGCCAGG
TGCTCCAACCACCTTATTTCCTTGTTTGTTGATTAGAACAACACAGAAAAAAGCAAATATAAAATTTTTAATGACT
CCATTTAAAAATATCACAGGGTGGGGCAAGGAAATTAGCTGAGATTCATCTCAGGATTGAGATTCTATCCCCCCT
TCCCGGCCCCAGCAGTGTCGCTCCAATTCAAATTAGTGGAGAAAGATTACAGTAGGCC^{CoBRA R}

SOX1

75% (6/8)

560	119	1020	+ve
Tumor	Metastasis	Tumor	Metastasis	Tumor	Metastasis
U D	U D	U D	U D	U D	U D	U D

Figure 4.5E A schematic diagram of part of *SOX1* with CpG island location within the promoter region (green)
The amplified region by the CoBRA assay is indicated by dotted lines which contains the CpG probe cg10245273 (blue) in CpG island (green text), along with CoBRA forward and reverse primers (underlined sequences), and *HpyCH4IV* restriction sites (underlined red sequence). The CoBRA digestion results are shown below the diagram for selected breast tumours and matching brain metastasis. In each case, undigested products (U) and digested (D) products run next to each other along with a positive control. The frequency of methylation in metastatic breast tumours is shown next to the gene name.

CpG Island: 302bp, 38CpGs

CoBRA F

TGGTTTTAGGGAAGTTAGTCCAAGGACATCTTCTAGGGGACGTGATCAATAACCCACTGGTAGAACTTCTT

CCTCTTCAACCCTGGGCTGCGGACCGGGGGGCGCAGGATCCCCAT**TCGA**AACTGTTGGTTCTGGGGCTGTTTC

CCATTCTCCAGAGAAAGTGCAGGCCCGGGGGTGTGC**CCGA**GAAAAGTGATGAGCAGTTGACCTCCCCTTCTC

AACTTGGTCTGGAGGCTTCATCATATTTCGGACGAAATGGGACAGAAGTGAGAATTAAAGCGGTTCTGAA
CoBRA R

PRDM13   **100% (6/6)**

119		707		712		+ve
Tumor	Metastasis	Tumor	Metastasis	Tumor	Metastasis	
U D	U D	U D	U D	U D	U D	U D

Figure 4.5F A schematic diagram of part of *PRDM1* with CpG island location within the promoter region (green)

The amplified region by the CoBRA assay is indicated by dotted lines which contains the CpG probe cg0126261 (blue) in CpG shore, along with CoBRA forward and reverse primers (underlined sequences), and *Taq1* restriction sites (underlined red sequence). The CoBRA digestion results are shown below the diagram for selected breast tumours and matching brain metastasis. In each case, undigested products (U) and digested (D) products run next to each other along with a positive control. The frequency of methylation in metastatic breast tumours is shown next to the gene name.

Figure 4.5G A schematic diagram of part of *PHOX2B* with CpG island location within the promoter region (green)
The amplified region by the CoBRA assay is indicated by dotted lines which contains the CpG probe cg15542784 (blue) in CpG shore, along with CoBRA forward and reverse primers (underlined sequences), and *HpyCH4IV* restriction sites (underlined red sequence). The CoBRA digestion results are shown below the diagram for selected breast tumours and matching brain metastasis. In each case, undigested products (U) and digested (D) products run next to each other along with a positive control. The frequency of methylation in metastatic breast tumours is shown next to the gene name.

4.2.4 Expression analysis in BCBM

In order to assess whether DNA methylation of these genes affects gene expression, RT-PCR was performed to assess the mRNA expression of these genes in methylated and unmethylated BM samples. Correlation between DNA methylation and loss/low expression was observed in *EOMES, WDR69, PRDM13* and *PHOX2B* in BM samples (Figure 4.6A). MicroRNA are small sized molecules and analysis of their expression requires a more sensitive method. Therefore, analysis of *MIR125B1* (cDNA size = 88bp) expression was performed using quantitative real-time RT-PCR. The results also showed decreased expression in methylated BM compared to unmethylated BM tumours (Figure 4.6B). Furthermore, expression of these genes in non-metastatic breast tumours from TCGA was also investigated by analysing the publically available gene expression data from the Agilent Gene Expression Microarray in the TCGA data portal. *EOMES*, *PHOX2B* and *SOX1* demonstrated expression in the majority of the samples, 95.8% (23/24), 100% (24/24) and 91.6% (22/24) respectively in non-metastatic breast tumours from TCGA. Taken together, these results highlight the importance of epigenetic silencing of these genes in metastasis development. However, the remaining genes (*WDR69*, *DZIP1* and *PRDM13*) showed low expression in unmethylated TCGA breast tumours, which may be caused by other genetic alterations such as mutations or deletions/allelic loss of the gene. Expression data for *MIR125B1* was not available in the TCGA data portal.
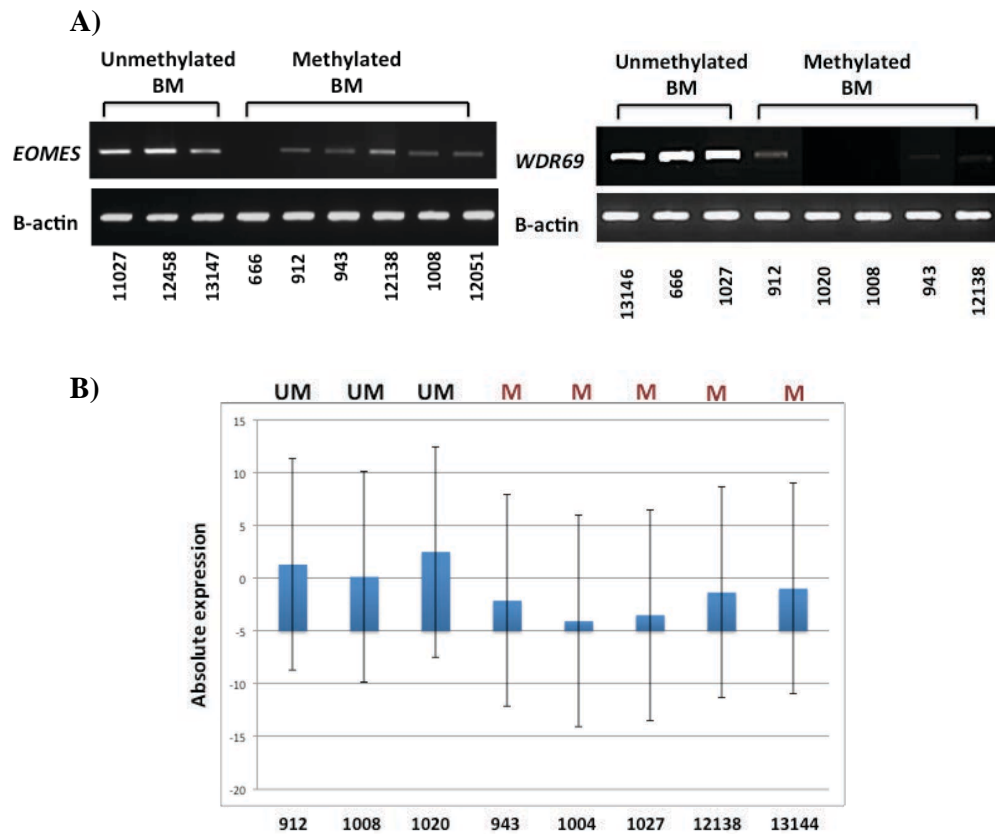
Figure 4.6 Expression analysis results for selected genes

A) Expression results are shown for *EOMES* and *WDR69* in methylated (loss of expression) and unmethylated BM samples (expression). For each case, β-actin is used as a control. B) Quantitative real-time expression results for *MIR125B* are shown for umethylated (UM, n=3) and methylated (M, n=5) samples, where unmethylated samples show high expression compared to the hypermethylated samples. The endogenous control, RNU24, was used for normalization.

4.2.5 Association of gene expression and patient prognosis

To determine the potential clinical relevance of these genes, the relationship between gene expression and breast cancer prognosis was examined using the PrognoScan online database (Mizuno et al. 2009) (*http://www.abren.net/PrognoScan/*). This tool uses publically available cancer microarray datasets to determine the relationship between gene expression and patient prognosis (section 2.10.4). In PrognoScan, patient populations are divided into high and low expression groups by selecting an appropriate cut-off point that maximises the statistical significance of difference in outcome, and results are corrected for multiple testing using the method of Miller and Siegmund (Miller and Siegmund 1982).

Overall survival (OS), relapse free survival (RFS) and distant metastasis free survival were assessed for the 7 genes identified in this study. Three datasets, GSE1456-GPL96, GSE1456-GPL96 and GSE299, from GEO (Gene Expression Omnibus) were used for PrognoScan expression analysis. Low expression of *DZIP1*, *PRDM13* and *PHOX2B* were correlated inversely with relapse free survival (RFS) in *DZIP1*, p=0.02 and p=0.005 (in GSE1456-GPL96 and GSE2990 dataset respectively), *PRDM13* in GSE2990 (p=0.01) and *PHOX2B* in GSE1456-GPL96 (p=0.007). Low expression of *DZIP1* and *PRDM13* were also significantly associated with worse distant metastasis free survival in GSE2990 (p=0.001 and p=0.009, respectively). In addition, worse overall survival (OS) in GSE7390 (p=0.048) was correlated with reduced expression of *PHOX2B* (Figure 4.7). This analysis suggests a significant association between the loss of expression of these genes and the reduced time of patient survival. No significant correlation between clinical

outcome and expression of *WDR69*, *EOMES* and *SOX1* was identified and data was not available for *MIR125B1*.

A) Relapse Free Survival

DZIP1 p=0.005 (GSE2990)

DZIP1 p=0.02 (GSE1456-GPL96)

PRDM13 p=0.01 (GSE2990)

PHOX2B p=0.007 (GSE1456-GPL96)

B) Distant Metastasis Free Survival

DZIP1 p=0.001 (GSE2990)

PRDM13 p=0.009 (GSE2990)

C) Overall Survival

PHOX2B p=0.048 (GSE7390)
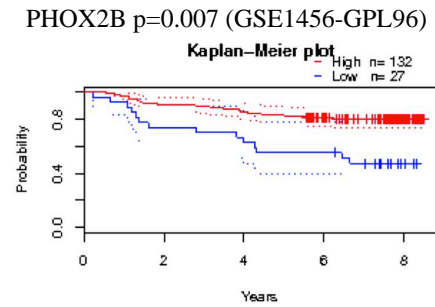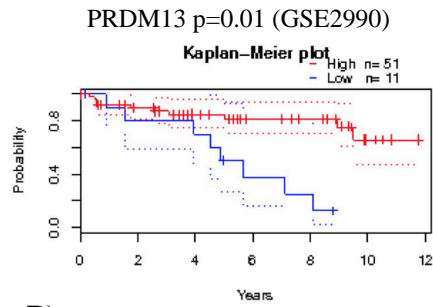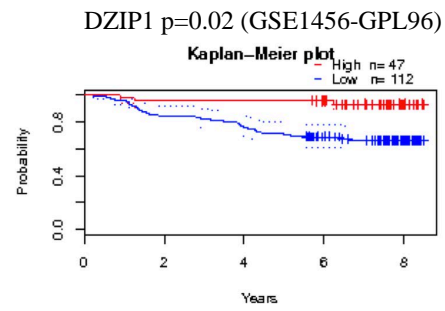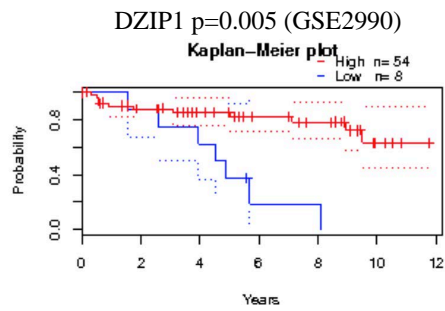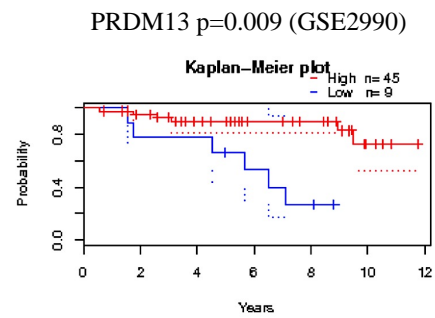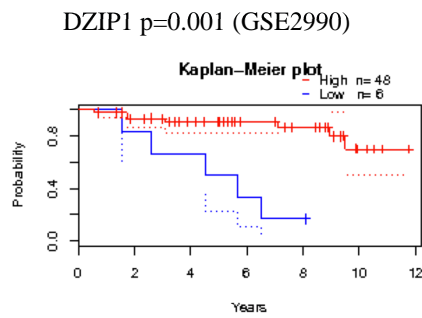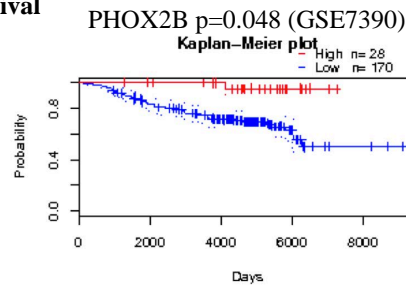
Figure 4.7 Kaplan-Meier survival analysis for *DZIP1, PRDM13* and *PHOX2B*
Graphs are shown for the association between gene expression (high, red; low, blue) and
A) relapse free survival (FRS) for *DZIP1*, *PRDM13* and *PHOX2B*, B) Distant metastasis
free survival for *DZIP1* and *PRDM13*, and C) overall survival for *PHOX2B* (OS). The
GEO dataset and p-value are shown for each case. Kaplan-Meier plots were generated by
the PrognoScan online tool.

4.2.6 ETS1 proto-oncogene is regulated by MIR125B1

The microRNAs are small noncoding RNAs that regulate many biological processes (Jansson and Lund 2012). Many MicroRNAs are associated with cancer and their dysregulation contributes to tumorigenesis (Wang et al. 2012). *MIR125B1* is known to affect various types of cancer and can act as either a tumour suppressor gene or an oncogene depending on their mechanism or target genes (Sun, Lin, and Chen 2013).

In a previous study that was carried out by Yan Zhan *et al.* (Zhang et al. 2011), *MIR125B1* was reported to be hypermethylated and downregulated in invasive breast tumours which is associated with poor patient survival. Also, it has been found that *ETS1* is a proto-oncogene and a direct target of *MIR125B1* and that it showed overexpression in invasive breast cancer, suggesting its role in breast tumour progression (Zhang et al. 2011). Here in this study, *MIR125B1* was also frequently methylated in the primary breast tumours with brain metastasis (71.4%), which matches the findings of Zhang's group and it was also frequently methylated in BM samples (79%). Furthermore, the gene expression analysis of *ETS1* was carried out on 13 brain metastasis samples. As expected, ETS1 was expressed in *MIR125B1* methylated brain metastasis samples (12/12, 100%) and ETS1 expression was reduced in a tumour where the *MIR125B1* gene promoter was not methylated (Figure 4.8). Taken together, down regulation of *MIR125B1* and high expression of its target gene *ETS1* in invasive breast tumours and brain metastasis suggests they may play an important role in breast cancer progression and metastasis development.
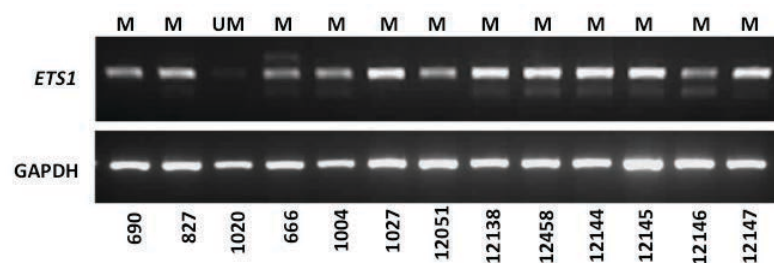
Figure 4.8 Expression analysis for *ETS1* in brain metastasis samples

Expression of *ESTS1* and *GAPDH* is shown for selected BM samples. Methylation status of *MIR125B1* in BM is shown above each case, where M=methylated and UM=unmethylated.

4.2.7 DNA methylation in BM from other organs

To determine whether these changes in DNA methylation of *EOMES, WDR69, MIR125B1, DZIP1*, *SOX1*, *PRDM13* and *PHOX2B* are specific to BM from breast tumours, methylation status analysis was performed on 40 BM which arise from other organs such as lung tumours (n=20) and melanoma (n=20) via CoBRA PCR, as described in previous sections (Figure 4.7). The methylation levels of *EOMES, WDR69* were low in BM from lung tumours (27% and 11%, respectively) and melanoma (10%, and 16%, respectively), suggesting that DNA hypermethylation of these genes was mostly restricted to BM from breast tumours. On the other hand, the digestion of CoBRA PCR products shows that *MIR125B1, DZIP1*, *SOX1*, *PRDM13* and *PHOX2B* were highly methylated in BM from lung tumours and melanoma. The results showed that 80% (*MIR125B1*), 65% (*DZIP1*), 70% (*SOX1*), 90% (*PRDM13*) and 60% (*PHOX2B*) of the BM from lung tumours were hypermethylated and 75% (*MIR125B1*), 63% (*DZIP1*), 66% (*SOX1*), 100% (*PRDM13*) and 58% (*PHOX2B*) were highly methylated in BM from melanoma. The above findings suggest that the DNA methylation of *MIR125B1, DZIP1*, *SOX1*, *PRDM13* and *PHOX2B* is not restricted to BCBM but may play a broader role in the development of BM from other tumours.

A)

Gene	BM from Lung Cancer	BM from Melanoma	Methylation % in Lung Cancer	Methylation % in Melanoma
EOMES			27%	10%
WDR69			11%	16%
MIR125B1			80%	75%
DZIP1			65%	63%
SOX1			70%	66%
PRDM13			90%	100%
PHOX2B			60%	58%

B)

BM from Lung   **BM from Melanoma**   **+ve**

EOMES, WDR69, MIR125B, DZIP1, SOX1, PRDM13, PHOX2B (U / D lanes)
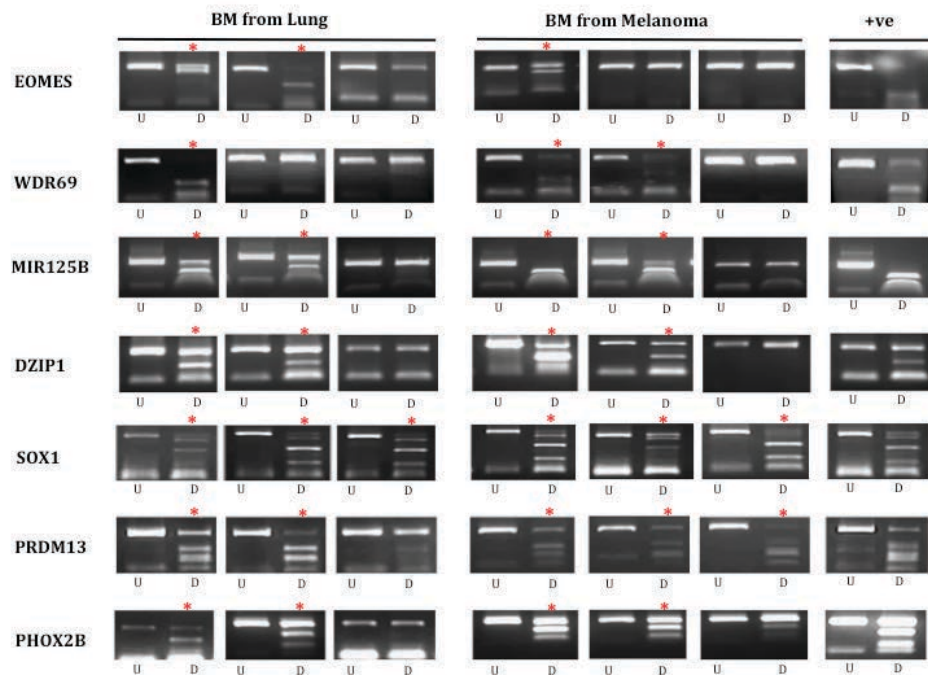
Figure 4.7 Methylation results of BM from lung cancer and melanoma

A) CoBRA results of the 7 genes when analysed in 20 BM from lung cancer and 20 BM from melanoma. Digestion results are indicated with a red circle (digestion, methylation) or a green circle (no digestion, absence of methylation). Black circles indicate no PCR product for that particular sample. The frequency (%) of methylation in brain metastasis from lung tumours and melanoma are also shown for each gene. B) Examples of CoBRA digestion in selected BM from lung and melanoma samples. Digested positive control is shown for each gene and methylated samples (digested) are indicated with a red star.

176

4.3 Discussion

Distant metastasis is the last stage of solid tumour malignancies. Multiple primary tumour sites metastasize to the brain; the most frequent tumour sites include lung, breast and melanoma (Gavrilovic and Posner 2005). The incidence of brain metastasis from breast cancer is increasing. Currently, 15%-20% of breast cancer patients develop brain metastasis with a median survival time of 7.8 months (Wikman et al. 2014) and most breast cancer associated deaths occur due to the failure of metastasis treatments (Salhia et al. 2014; Li et al. 2011). Studies have found that even with early diagnosis and the treatment of breast cancer patients, some cases form metastasis, suggesting that initiating events in breast cancer are responsible for predisposing metastasis formation (Webb 2003). This highlights the importance of understanding the underlying molecular biology and identifying poor prognosis signatures in primary breast tumours from early stages of diagnosis.

In breast cancer, tumour cells usually spread and invade certain distant sites, for example, breast cancer metastasis to the bones, lungs and brain is common but metastasis to the liver is less common (Minn, Kang, et al. 2005). This observation of "seed-and-soil", where disseminated tumour cells (seed) progress and adapt in certain sites (soils), depends on the microenvironment of the metastatic organ and genetic constitution of the tumour cells (Valastyan and Weinberg 2011; Wan, Pantel, and Kang 2013). For example, overexpression of *CXCR4*, *IL11* and *OPN* genes in breast cancer showed a high affinity to metastasize to the bone (Kang et al. 2003), while breast cancer cells that overexpress *COX, EREG* and *ANGPTL4* favour metastasis to the lung (Minn, Gupta, et al. 2005). In

addition, *COX2, HBEGF* and *ANGPTL4* genes were specifically overexpressed in BCBM (Bos et al. 2009).

The brain is a unique organ and differs from any other organ due to its specialized microenvironment. Successful metastasis to the brain requires specialised interaction between the circulating tumour cells and brain defences and microenvironment, which are characterised by the presence of a blood-brain-barrier (BBB) and the absence of a lymphatic drainage (Martín et al. 2008). Therefore, investigating signature genes in human brain metastasis and their matching primary breast tumours is important in order to identify new prognostic and diagnostic markers, which may be helpful in developing new therapeutic approaches.

Loss of expression of tumour suppressor genes (TSGs) is a fundamental step in tumour formation. However, metastasis suppressor genes are different from TSGs since they are not involved in tumour growth but they control cancer cell dissemination, survival during circulation and adaptation at the distant organ site (Wikman et al. 2014). Since metastasis suppressor genes are rarely mutated (HORAK et al. 2008), epigenetic alterations, such as DNA methylation, could be one of the mechanisms involved in the dysregulation of gene expression leading to metastasis.

To expand current knowledge of the epigenetic changes arising in BCBM, this study used the Infinium HumanMethylation450 array to investigate DNA methylation events in BM from breast cancer. This data was combined with the methylation array data of non-metastatic breast tumours from the TCGA data portal to identify potential genes that may enhance brain metastasis in breast cancer cells and that are involved in BCBM development.

178

In order to identify BCBM signature genes, differential analysis was applied, which led to the identification of 28 genes (37 probes) that were differentially methylated between non-metastatic breast tumours and BCBM. Seven genes from the list (*EOMES*, *WDR69*, *MIR125B1, DZIP1, SOX1, PHOX2B* and *PRDM13*) were chosen for further analysis based on their relative function and the position of the CpG probe within the promoter region and CpG island/shore/shelf regions. CoBRA analysis was performed on matching paired breast tumours and brain metastasis samples and the results have shown that hypermethylation was not only found in BCBM, but also in the corresponding primary aggressive breast tumours, suggesting their role in the early stages of metastasis formation. The functional relevance between DNA methylation and gene expression was investigated and the lower expression level of *EOMES, WDR69, MIR125B1, PHOX2B* and *PRDM13* was observed in methylated BCBM. Any clinical relevance for these genes was determined by assessing the publically available gene expression datasets (GEO database) and was analysed using the PrognoScan tool. Analysis showed that down regulation of *DZIP1*, *PRDM13* and *PHOX2B* genes was significantly associated with worse patient survival.

4.3.1 Signature BCBM genes

4.3.1.1 EOMES

Eomesodermin *(Eomes, aka TBR2)* is a member of the T-box transcription factor. The overexpression of *EOMES* in colorectal cancer was associated with significantly reduced lymph node metastasis (Atreya et al. 2007). In this study, *EOMES* was hypermethylated in paired breast tumours and the corresponding BCBM samples, but was unmethylated in

the non-metastatic breast tumours and the hypermethylation of BM was associated with the loss of gene expression. Also, methylation of the gene was less frequent in BM from lung tumours (27%) or in melanoma (10%). When considering all the identified methylation and expression results of *EOMES*, these results suggest that methylation of the *EOMES* promoter may be an early indicator of BCBM formation in breast cancer patients.

4.3.1.2 WDR69

Also known as *DAW1, WDR69* is a dynein assembly factor with WD repeat domains. This gene is located on chromosome 2 and contains one CpG island at the 5' UTR within the promoter region. The size of the island is relatively small (314bp) as it contains 33 CpG sites. WD repeat domain proteins are involved in various cellular functions such as signal transduction, cell cycle control, transcription regulation and apoptosis (Li and Roberts 2001). Genetic alterations of these genes can cause a wide range of diseases; examples include mutations in the WD repeat domain of AHI1 in some autosomal recessive disorders, chromosomal translocation of WDR11 in glioblastoma, upregulation of WD repeat protein, endonulein, in adenocarcinoma and others (Smith 2008). Currently, not much is known about WDR69 biology, hence further studies are required in order to understand the role of this gene in the development of brain metastasis.

4.3.1.3 SOX1

SOX1 is a one-exon gene located on chromosome 13. It has a 3018bp CpG island containing 274 CpG sites covering the TSS, 3'UTR and a large part of the 1st exon. *SOX1*

is a member of the SOX family and is known as a tumour suppressor gene because it deactivates Wnt/b-catenin signalling pathways in liver cancer (Tsao et al. 2012). Promoter hypermethylation of *SOX1* has been observed in many types of cancer and showed an association with gene down regulation (Lai et al. 2014; Shih et al. 2013). Also, studies showed that the over expression of *SOX1* is associated with the suppression of cell proliferation, cell migration and cell invasion in hepatocellular carcinoma (Tsao et al. 2012), ovarian cancer (Su et al. 2009) and cervical cancer (Lin et al. 2013). The results in this chapter showed that *SOX1* was hypermethylated in the majority of metastatic breast tumours (75%; 6/8). Although the samples assessed here did not show evidence of methylation correlating with reduced expression, the overwhelming evidence of *SOX1* methylation playing a role in other cancers suggests that further analysis of additional samples would be worthwhile. Hypermethylation was frequently found in BM from lung tumours (70%) and melanoma (66%) as in BM from breast tumours (75%). Our findings, presented here, indicate that DNA hypermethylation of *SOX1* may contribute to metastasis to the brain from different primary sites.

4.3.1.4 PHOX2B

The homeobox gene, *PHOX2B*, consists of 3 exons and contains 2 CpG islands where one is located in the 3'UTR region (31 CpG sites) and the second island is situated in exon 2 of the gene. Previous studies have found a relationship between *PHOX2B* alterations and cancer; for instance, somatic and germ-line mutations and LOH at *PHOX2B* were previously reported in neuroblastoma (Perri et al. 2005). Other small studies have found hypermethylation of *PHOX2B* in colorectal cancer and neuroblastoma

(Yang 2012; de Pontual et al. 2007). In this study, *PHOX2B* was only methylated in metastatic breast tumours (83%) but not in the non-metastatic breast tumours from TCGA. The *PHOX2B* promoter was also highly methylated in brain metastasis from breast cancer, lung cancer and melanoma (73%, 60% and 58%, respectively). Hypermethylation was strongly associated with the loss of expression. The unmethylated breast tumours without metastasis showed high expression of the gene, while expression was lost in the methylated brain metastasis. Loss of expression of *PHOX2B* was associated with poor patient prognosis which suggests *PHOX2B* as a potential biomarker for BCBM.

4.3.1.5 PRDM13

PRDM13 (PR domain containing 13) is a four exon gene situated at 6q16.2. Very few studies have been done on PRDM13, but PR proteins containing the PR domain have been reported to be tumour suppressive by regulating chromatin mediated gene expression (Jiang and Huang 2000). Results from the current study have shown a high level of methylation in breast cancer and brain metastasis pairs as well as in brain metastasis from lung and melanoma (92% and 100%, respectively). This suggests *PRDM13* methylation may be involved in the development of brain metastasis from primary breast, lung and melanoma tumours. Loss of expression of PRDM13 was strongly correlated with promoter hypermethylation and was associated with poor relapse free survival and distant metastasis free survival in breast cancer patients.

4.3.1.6 DZIP1

DAZ interacting zinc finger protein 1 is located on chromosome 13. It is known as a tumour suppressor gene that regulates the hedgehog signalling pathway in many types of cancer. It has been reported that mutations of *DZIP1* might cause the aberrant expression of downstream genes in the hedgehog pathway (Kikuyama et al. 2012). So it is possible that the aberrant methylation of *DZIP1* can also affect this signalling pathway during cancer formation. The results of this thesis showed frequent hypermethylation of *DZIP1* in BCBM samples (60%) as well as methylation in BM from lung cancer (65%) and melanoma (63%). Prognoscan data showed that loss of expression was strongly associated with poor survival in breast cancer patients. Although the expression of *DZIP1* could not be determined in the small number of BM samples assessed during this study, the previous links of *DZIP1* with tumour suppressor functions, the loss of expression associating with worse survival and the results from our methylation analysis suggests that further investigation of *DZIP1* and its biological functional relevance in BCBM is warranted.

4.3.1.7 *MIR125B1* and its target gene *ETS1*

MicroRNAs are single stranded non-coding RNAs that are known to control many biological processes, including cell proliferation, death, differentiation and mRNA translation (Iorio et al. 2005). *MIR125B1* is located on chromosome 11 and is involved in the posttranscriptional regulation of the mRNA expression of target genes. Recent studies have reported the important role and effect of *MIR125B1* in carcinogenesis. *MIR125B1* was observed to be aberrantly expressed in different cancer types and may act as an

183

oncogene such as in prostate cancer and glioma (Shi et al. 2007; Schaefer et al. 2010; Wu et al. 2013) or as a tumour suppressor gene in breast, liver and bladder cancers (Iorio et al. 2005; Tang et al. 2012; Jia et al. 2012; Huang et al. 2011). To date, several *MIR125B1* target genes have been identified and are known to be involved in cancer, for example, *Bak1* in prostate cancer (Shi et al. 2007), *Bmf* in glioma (Xia et al. 2009), *ERBB2/ERBB3* and the most recent *ETS1* in breast cancer (Scott et al. 2007; Zhang et al. 2011).

A significantly high expression of *MIR125B1* was observed in breast tumours but not in invasive breast cancer, suggesting that *MIR125B1* may play a role as a pro-metastatic factor in metastatic breast tumours (Tang et al. 2012). *MIR125B1* was found to be methylated and down regulated in aggressive breast tumours, which was associated with poor patient survival. According to a study by Zhang and his group (Zhang et al. 2011), the overexpression of *MIR125B1* in breast cancer cells promotes the G_1 cell cycle to arrest through the target gene *ETS1*. Generally, the transcription factor *ETS1* regulates cellular processes such as cell proliferation and migration and it has been associated with poor prognosis of cancer patients when it is upregulated (Davidson et al. 2001). The loss of expression of *MIR125B1* in invasive breast tumours induces the expression of the target gene *ETS1* and consequently, the activation of cell proliferation and cell growth. In this study, *MIR125B1* showed hypermethylation in metastatic breast cancer as well as in brain metastasis samples but showed low methylation in breast tumours without metastasis. Hypermethylation and low expression of *MIR125B1* also correlated with the overexpression of *ETS1* in BM. *MIR125B1* hypermethylation was not restricted to brain metastasis from breast tumours. Hypermethylation was also observed in BM from lung tumours (80%) and melanoma (75%).

4.3.2 Conclusion and further work

Metastasis is the most fatal stage of cancer. Many studies using expression or methylation profiling provided a closer look at the molecular biology events in cells contributing to metastasis. Recent studies suggest that the genes responsible for metastasis formation are expressed by primary tumour cells themselves (Ramaswamy et al. 2003; Bernards and Weinberg 2002; Bos et al. 2009; Minn, Gupta, et al. 2005; Kang et al. 2003). Therefore, the aim of this study was to identify methylation events that are specific to the development of BCBM. Using the latest Illumina HumanMethylation450K BeadChip in this study, a list of novel candidate genes was identified, which is considered worthy of further investigation. One of the advantages of using the HumanMethylation450K array is that probes can be selected based on their gene feature. The chosen probes were located in the promoter region of the gene to ensure relevance between DNA methylation and gene silencing. However, not all the identified genes were analysed further and a number of selected genes were chosen for further confirmation. The CoBRA digestion of *TRIM68* was not done due to the lack of a restriction enzyme cutting site within the CpG probe. Therefore, analysis was performed for the remaining selected genes based on the availability of a cutting site within the CpG probe. This could be corrected by analysing the whole region using bisulfite clone sequencing. In the current study, the CoBRA assay was favoured because it is a fast, cost effective and widely accepted method for assessing DNA methylation (Xiong and Laird 1997). Therefore, the CoBRA assay provided an opportunity to rapidly and effectively screen additional cohorts to confirm our findings.

For this type of study, where paired breast tumour and matching brain metastasis samples are required for analysis and validation, samples should be used with care. It is difficult to obtain matching metastases paired samples, where patients may develop metastasis several years after primary tumour diagnosis. We were fortunate to have 10 paired primary breast tumours and matching BCBM samples to validate the methylation data, and to our knowledge, this is one of the few studies where such a cohort of matched samples was available for analysis.

A particular weakness encountered in this study was the use of only 5 BM samples in the array. Although the methylation of the identified genes was confirmed in a different set of BM samples, it would have been beneficial to include more BM samples in the array to produce more significant genes in the list. Yet, we have identified a set of genes that may be important for the development of BCBM and could be useful prognostic markers for BCBM risk assessment.

Overall, this study was able to identify a list of genes that were significantly methylated in breast tumours that form brain metastasis compared to non-metastatic breast tumours. Some of the genes presented potential candidates to be epigenetically downregulated in metastatic breast cancer with an association with poor survival. Among the list, *DZIP1* was a good candidate to be analysed further and therefore, expression analysis was carried out for *DZIP1* in brain metastasis samples to assess its functional relevance. However, several attempts, using three different sets of primers each time, were carried out for *DZIP1* expression analysis and none of the attempts were successful. Since this study was the last project undertaken in my final year, no adequate time was available to carry out further analysis. Therefore, it would be worth investigating and carrying out

further biological and functional studies to confirm if these genes from the list contribute to metastasis formation from breast cancer. For example, cell migration and invasion assays can be performed on breast cancer cell lines that show hypermethylation and expression loss to investigate if reduced gene expression is associated with increased cell motility and invasion. Alternatively, expression could be knocked-out in non-methylated, expressing breast cancer cell lines.

Additionally, it would be beneficial to confirm this data in a larger, independent cohort of samples and also to analyse our list more comprehensively before considering these genes as true prognostic and diagnostic markers for BCBM metastasis formation.

Finally, another consideration could be taken into account by analysing the genes involved in the late stage of tumour progression rather than in the early stage of metastasis formation. This data can be analysed by finding genes that are hypermethylated in BM samples but show low/no methylation in metastatic breast cancer tumours. Identifying such genes will increase our understanding of tumour progression and metastases since these genes may be involved in cell survival during circulation in addition to adaptation and colonization in the brain.

In conclusion, we have identified a preliminary set of genes/probes that may be developed into useful molecular diagnostic markers (gene signatures) for the early prediction of BCBM formation.

This work was presented at The National Cancer Research Institute conference (NCRI) in Liverpool as a poster presentation on 4th November 2014.

Chapter Five: Genetic Alterations in Vestibular Schwannoma

5.1 Introduction

The benign tumour, vestibular schwannoma, is characterised by the continuous proliferation of schwann cells around the nerves which may cause impaired hearing, imbalance and may lead to death at later stages. Vestibular schwannoma has been studied extensively in the past few years and has been associated with *NF2* gene mutation and/or Ch22 LOH. However, a subset of schwannoma patients showed no alterations in *NF2* gene, suggesting other causative genes are responsible for schwannoma tumour development. Recent studies have identified other genes that are involved in vestibular schwannoma and located on chromosome 22 near to *NF2* gene, such as *SMARCB1* and *LZTR1*.

With the expansion of cancer research studies, scientists are looking for cheap, fast and high-throughput technologies to better understand the molecular biology and genetics of cancer formation. The development of NGS has offered many advantages to cancer research through WGS and WES technologies. Nowadays, WES technology is widely used because it is faster, lower in price and targeted to the sequence of the protein-coding regions of the genome to identify disease-causing variants.

The current study utilises WES in combination with bioinformatics tools and applications to identify disease causing gene variants in non-NF2 schwannoma tumours. Despite the success of WES technology, this technology is full of challenges that were resolved and limited in this study by applying different strategies to reduce false positive variants and attempted to identify the true causative gene(s) in schwannomas.

5.2 Aim of the study and approaches

- Identification of somatic mutations and driver genes that may lead to non-NF2 vestibular schwannoma formation using whole exome sequencing technology.

- Utilising various bioinformatic tools, in combination with algorithmic software to investigate variants that may be involved in schwannoma development.

5.3 Results

Whole exome sequencing technology utilises parallel sequencing of the protein coding regions of the genome (exomes). It is a powerful and cost effective technique for the identification of causative mutations in many human diseases including cancer.

This study used whole exome sequencing techniques on a panel of 8 sporadic unilateral vestibular schwannoma tumours (S14T, S21T, S36T, S37T, S44T, Lp152T, Lp272T and Lp303T) from Spanish patients, in addition to matching normal blood DNA for 3 samples (S14S, Lp272S and Lp303S). These samples did not harbour *NF2* mutations by combined dHPLC/PCR and MLPA (MRC-Holland) analysis and demonstrated no LOH at 22q microsatellite markers (Torres-Martin et al., 2013). Exon 1 analysis was not conducted in the above samples, since exon 1 mutations are rare in SNP databases. Information about the tumour samples can be found in Appendix 7.7.

Before starting the analysis, the gender of the samples was estimated and compared with patients' clinical information to rule out a sample mix up during the sample preparation

189

for exome sequencing. This was conducted by counting the number of heterozygous X chromosome variants in each sample. Typically in exome sequencing data, female samples show an excessive number of heterozygous SNPs on the X chromosome (>200) compared to male samples (<100) (Guo Y et al. 2014). After counting the number of variants, the results showed that exome sequencing was able to estimate the sample's gender (Table 5.1).

Sample ID	Gender	Number of heterozygous X chromosome variants
S14	Female	239 variants
S21	Male	31 variants
S36	Male	28 variants
S37	Male	50 variants
S44	Female	219 variants
Lp152	Male	39 variants
Lp272	Female	264 variants
Lp303	Female	295 variants

Table 5.1: Estimation of Schwannoma samples' gender

Gender estimation was performed by counting the number of heterozygous X chromosome variants in each sample. A sample showing >200 hits is considered female, while a sample with <100 hits is considered male.

5.3.1 Newly identified mutated genes in meningioma

A previous study by Clark *et al.* identified somatic mutations in 4 genes: *TRAF7* (TNF receptor associated factor 7), *AKT1* (v-akt murine thymoma viral oncogene homolog 1), *SMO* (Smoothened frizzeled family receptor), and *KLF4* (Krupple like factor 4), in meningioma samples that were exclusive of *NF2* mutation (Clark et al., 2013). According to the paper, mutations in *TRAF7* occurred in 25% of non-*NF2* meningioma samples. These mutations co-occurred with either a recurrent mutation in *KLF4* or with *AKT1*. On the other hand, *SMO* mutations were identified in less than 5% of meningioma samples and were exclusive of mutations to the other 3 genes (Table 5.2).

Mutations in these genes were screened for as both meningioma and schwannoma are intracranial nervous system tumours and associated with genetic alterations in NF2 genes. This part of the study was carried out before receiving the exome sequencing data and the investigation of mutations in these genes was carried out in 20 non-*NF2* schwannoma samples.

Firstly, analysis was started by sequencing the recurrent mutations in *KLF4, AKT1* and *SMO*. The hotspot mutations of *KLF4* and *AKT1* were missense mutations, p.K409Q and E17K respectively, which always co-occurred with *TRAF7* mutations. On the other hand, the recurrent mutation L412F of *SMO* was exclusive of mutations in the other 3 genes. Conventional Sanger sequencing was performed on schwannoma tumour samples using pre-designed primers from the study of the 3 hotspot mutations. From the results of the electropherograms, sequencing was unable to identify these mutations in any of the non-*NF2* samples (Figure 5.1).

Clark et al. identified 12 mutations in 8 out of 21 exons of *TRAF7* in meningioma (Table 5.2). Analysis of these mutations in schwannoma samples was carried out using Sanger sequencing. These mutations were not detected in any of the samples (Figure 5.1). These mutations of the 4 genes may be exclusive to meningioma tumours and at the time of writing this part of the thesis, no other studies have identified such mutations in schwannoma tumours.

After receiving and investigating exome sequencing data, none of these mutations were found in the samples included in exome sequencing, hence confirming the Sanger sequencing data.

Gene Symbol	Gene Name	Position of Hotspot Mutation
TRAF7	TNF receptor associated factor 7	p.T145M
		p.F337S
		p.C338Y
		p.E353insFRRDAS
		p.G390R
		p.G536S
		p.S561N
		p.L580del
		p.K615E
		p.Q637H
		p.R641C
		p.R653Q
KLF4	Krupple like factor 4	p.K409Q
AKT1	v-akt murine thymoma viral oncogene homolog 1	p.E17K
SMO	Smoothened frizzeled family receptor	p.L412F

Table 5.2: Hotspot mutations identified in meningioma

Table showing the recurrent mutations previously identified in meningioma by Clark et al. and that have been analysed in schwannoma tumours.

KLF4:

AKT1:

TRAF7:

Exon11
p.C338Y
p.E353insFRRDAS

Exon13
p.G390

Exon18
p.S561N
p.L580de

Exon19
p.K615E

Exon10
p.F337S

Exon17
p.G536S

Exon20
p.Q637H
p.R641C
p.653Q

SMO:

Exon6
p.L412F
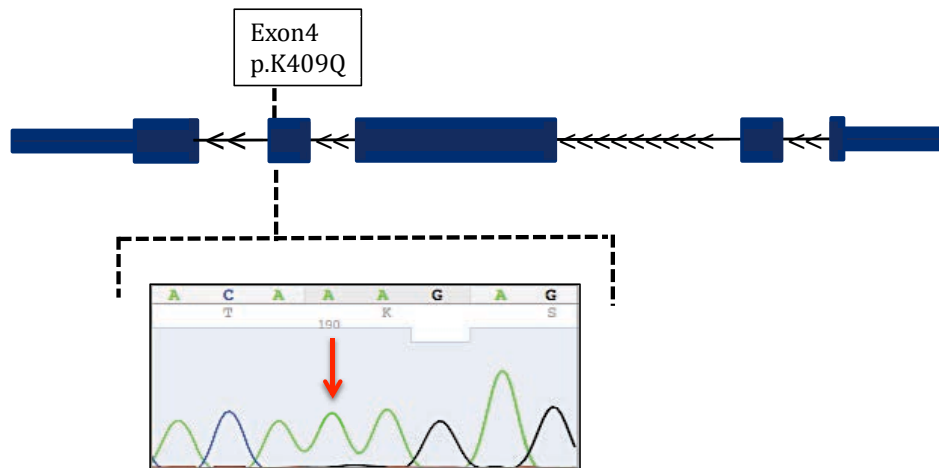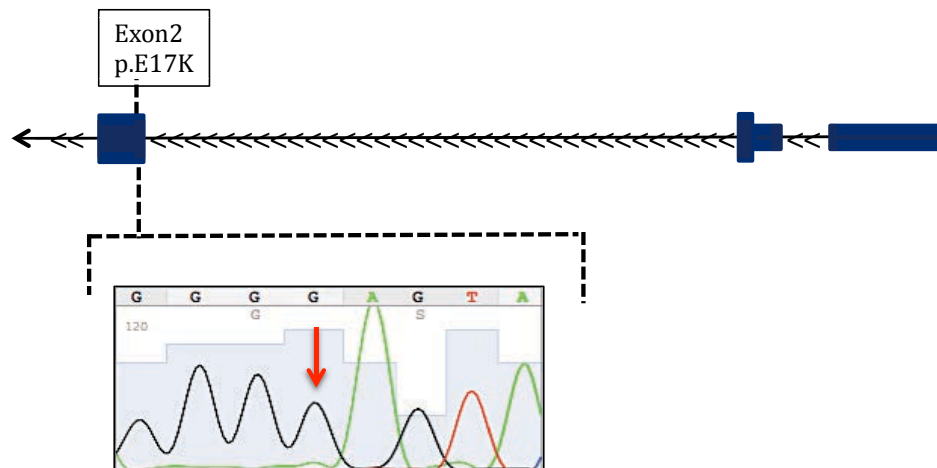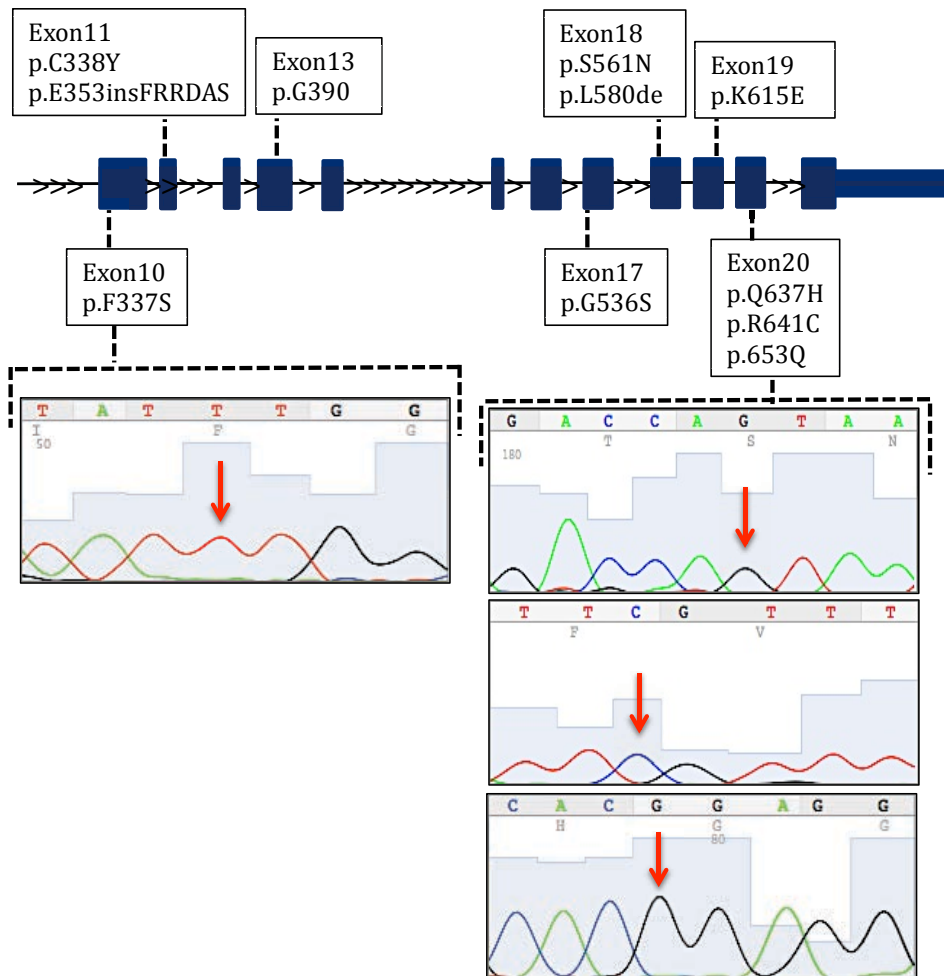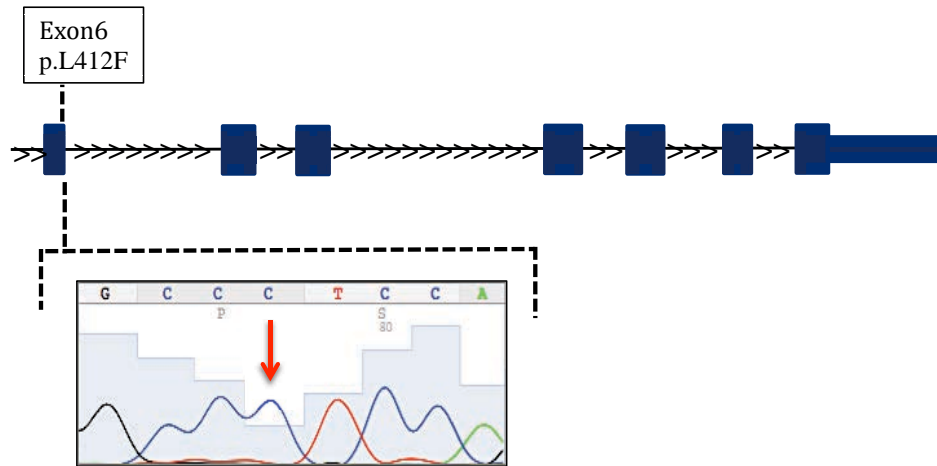
Figure 5.1: Analysis of meningioma genes (*AKT1, KLF4, SMO* and *TRAF7*) in non-*NF2* schwannoma

A schematic diagram of *KLF4, AKT1, SMO* and *TRAF7* genes (blue) that shows the location of the hotspot mutations. In each case, Sanger sequencing of each hotspot is shown in non-*NF2* schwannoma tumours and indicated by a red arrow. For *TRAF7*, 4 out of the 12 mutations are shown. No mutations were detected in any of the schwannomas analysed and the Sanger sequencing scan shows a wild type sequence for each of the mutations analysed.

5.3.2 Analysis of exome sequencing data using in-house pipeline

5.3.2.1 Exome sequencing data filtration pipeline

After confirming the samples' gender, the received exome sequencing data of all the schwannoma tumour samples were combined in one datasheet for analysis. The total number of variants was 194,318 with an average of 24,290 variants per sample. To reduce the vast number of variants in exome sequencing data, a series of filters and approaches were applied to identify the genes and variants that could be involved in non-NF2 schwannoma development (Figure 5.2). Analysis initially started by removing synonymous SNVs and variants present in intronic regions. After that, germ-line mutations found in any of the normal blood samples were also excluded. To eliminate common SNPs from the remaining 3,493 variants, variants that showed a minor allele frequency (MAF) greater than 0.001 (1%) in the 1000 genomes database and/or more than 10 times within the in-house data (data from genome sequencing centre's database) were also removed. Furthermore, any splice-site candidate mutation that was >3bp away from the exon-intron boundary was also excluded.
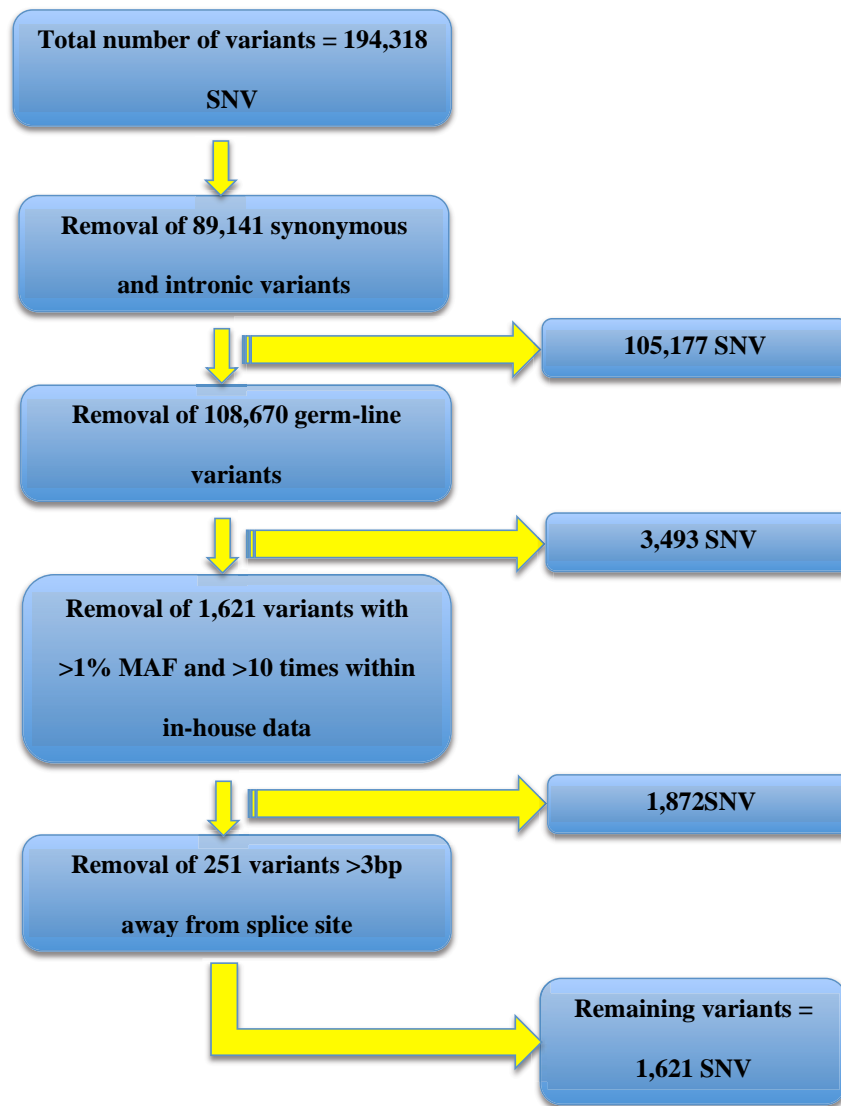
Figure 5.2: Filtration process applied to exome sequencing data to reduce the massive number of detected variants

After applying these filters, the number of remaining candidate variants was 1,621. Among these variants, 631 variants were novel, 662 variants had rs ID in dbSNP (with MAF <1%) and 328 variants were not identified in any SNP databases but had been reported multiple times within the in-house data (Figure 5.3A). Moreover, based on the type of coding variants, 47 variants were frameshift mutations (26 frameshift deletion and 21 frameshift insertion), 39 variants were non-sense mutations (corresponding to 36 genes), 152 variants were splice-site mutations (145 genes) and the majority were missense mutations with 1,383 variants (1,227 genes) (Figure 5.3B).

It was noticeable that the number of variants present in the samples that had exome sequencing data for the matching normal blood samples (S14, Lp303 and Lp272) was less than the other samples with no matching normal blood samples (Lp152, S36, S37 and S44). Samples S14T, Lp303T and Lp272T contained 23, 34 and 25 variants respectively, while Lp152, S36T, S37T and S44 contained 311, 323, 319 and 251 variants respectively. This major difference in the number of variants is a result of the filtration steps when removing germ-line variants (Figure 5.3C).

Afterwards, different approaches were applied to analyse the data. Initially, in-silico analysis was carried out to conduct an in depth investigation of the candidate variants and later bioinformatics tools were used, such as IGVtools and VarScan to support the analysis.

A)

Novel: 631
MAF <1% (rs ID): 662
In-house <10 times: 328

B)

Frameshift: 47
Stop-gain: 39
Splice-site: 152
nonsynonymous: 1383

C)

Samples with matching normal: S14T (23), Lp272 (25), Lp303 (34)
Samples without matching normal: S21 (335), S36 (323), S37 (319), S44 (251), Lp152 (311)
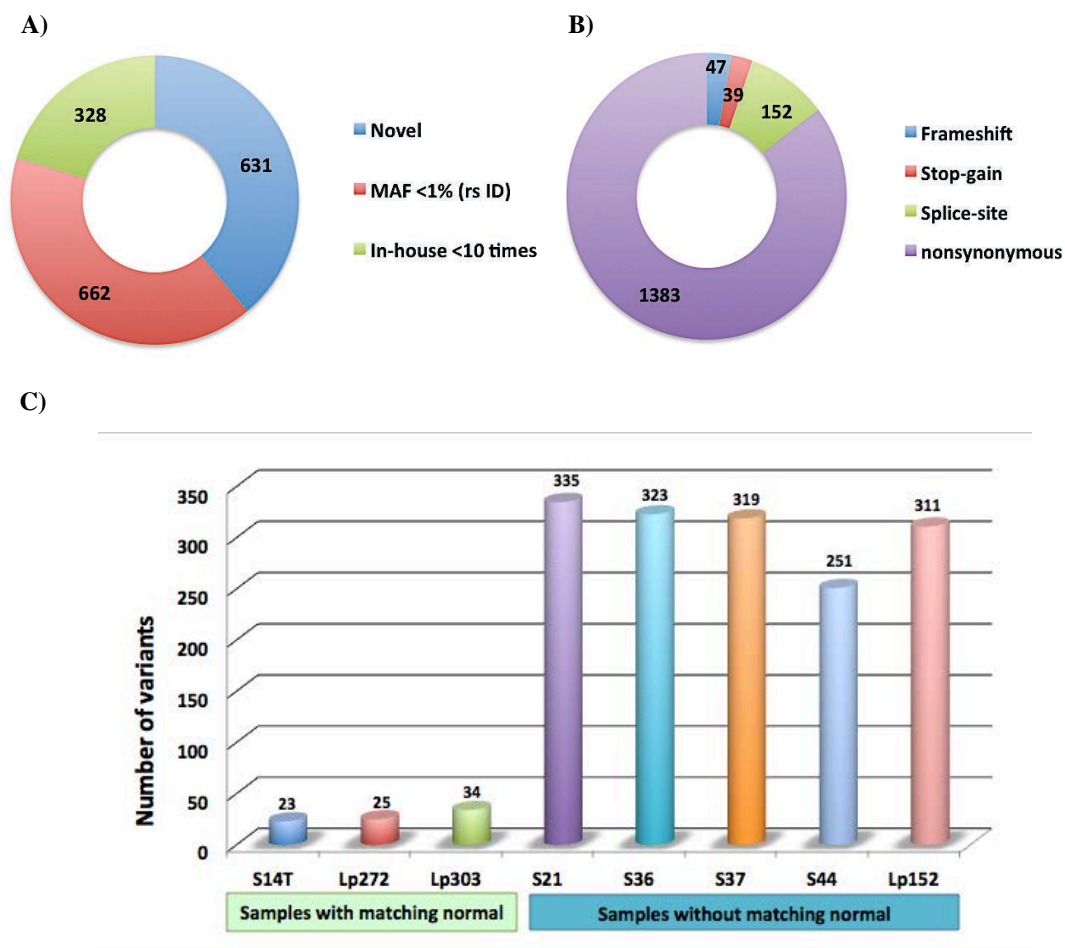
Figure 5.3 Distribution of variants after the filtration process A) Distribution of SNVs according to variant novelty. B) Distribution of SNVs based on the type of coding variants. A description of the colour legends is shown adjacent to the charts. C) Number of variants in each sample. Samples were divided into two groups, with or without matching normal as shown.

5.3.2.2 Exome sequencing in-silico analysis

The analysis began by prioritising the list of variants using different online tools. Genes or variants found in more than one sample were labelled as a high priority candidate. At the beginning, all genes were investigated in an exome variant server website for high impact mutations (non-sense, frameshift and splice-site) and those with fewer SNPs are more likely to be significant. Also, genes were checked to see if they were mutated in other types of cancer using the TCGA and COSMIC databases. In the COSMIC database, genes with high non-sense and frameshift mutation distribution among different types of cancer were labelled as significant. Similarly, genes were considered important if they showed mutations in more than 5 types of cancer with a mutation frequency of >2.5% in each type in the TCGA database. Variants were also investigated using the 1000 genomes Project database by looking at the frequency of the mutation in the Spanish population (IBS=Iberian population in Spain) compared to the total population (n=26 populations around the globe). Furthermore, different web-based tools were applied to the candidate variants. For splice-site variants, a splice-site prediction tool was used to predict the effect of intron splicing on the protein. Additionally, Polyphen-2 was used to assess the function effect of missense variants on proteins (benign or damaging). Figure 5.4 illustrates the various online tools applied in the in-silico analysis using *NF2* gene mutations as an example.
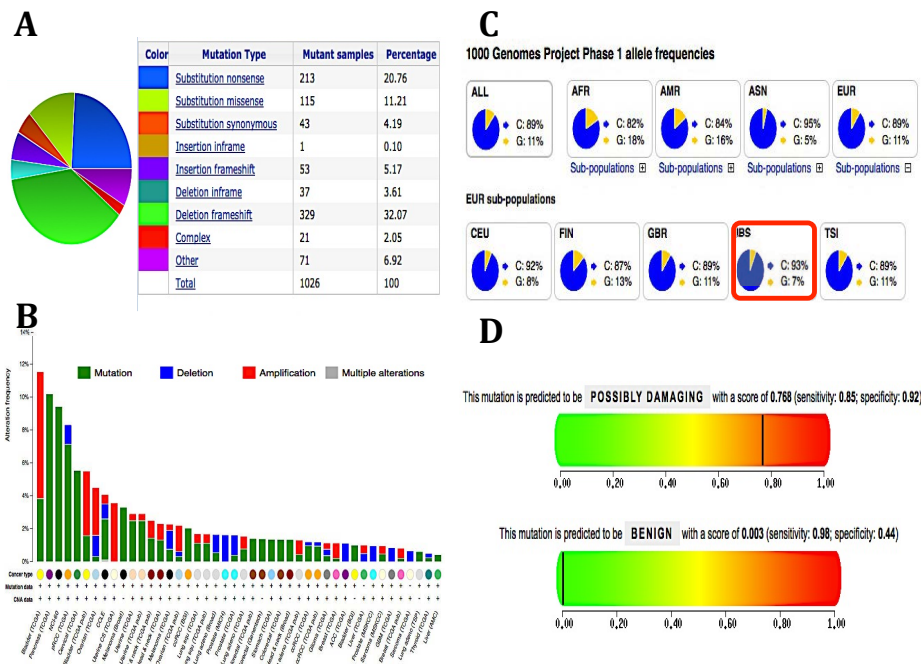
Figure 5.4 Examples of different online tools used in the study

All examples represent mutations in *NF2* genes. A) Distribution of mutation types in COSMIC database for *NF2* gene displayed as a pie chart and each colour represents a mutation type. The adjacent table shows the number of samples and the mutation percentage out of the total number of mutations in *NF2*. B) A summary of alterations in *NF2* across different types of cancer in the TCGA database. In this study, a gene is considered significant if it shows mutations (green bars) in more than 5 types of tumour with >2.5% mutations in each type. C) Allele frequency of a variant in different populations in the 1000 Genomes Project database. The Spanish population is the Iberian population in Spain (IBS) which is a subpopulation of Europeans (EUR) (marked by a red triangle). D) Polyphen-2 protein function prediction tool. The figure shows the score of 2 different variants; the variant of the top bar was predicted to be possibly damaging (score of 0.77) and the other variant was predicted to be benign with a score of 0.0 (lower bar).

5.3.2.3 IGV tools

IGV was used in conjunction with the in-silico analysis to support the list of candidate genes. IGV is a visualization tool that allows real time exploration and the analysis of large-scale genomic data. It supports a wide range of data types including NGS to visualize SNPs by comparing them to a reference genomic sequence, plus the deletion and insertion of regions of the genome. When uploading a special format file of the sample (BAM file), sequencing data that is aligned to the most recent human genome sequence (hg19) is displayed. When displaying sample information, the level of detail varies according to the zoom level. As the view is zoomed out, IGV displays the general read coverage of the sample (Figure 5.5A). This is useful for evaluating the overall coverage quality and for detecting technical issues in the sequencing of the interesting area. When the view is zoomed in IGV displays the sequence of the sample aligned to the reference genome. Nucleotides are colour coded and SNPs are highlighted as a mismatch. Bases that match the reference genome are displayed in the same colour, while mismatched reads are displayed in different colours. For example, in Figure 5.5B, IGV window displays the sequence of *NF2* after zooming in. The histogram on the top displays the coverage quality of this area and the highlighted area represents where the stop-gain mutation p.W60X was found in the schwannoma sample Lp152T. The mismatched bases are displayed in a different colour than the reference base and when the cursor is moved over the base pair, a pop-up window shows the number of reads and the percentage of each base. IGV is a powerful tool for eliminating false positives, which will aid in identifying true findings.

The analysis was conducted to find (i) SNV in the genes involved in cancer pathways, (ii) variants located in chromosome 22 near *NF2*, (iii) high impact variants and finally, (iv) the remaining non-synonymous variants.

A)
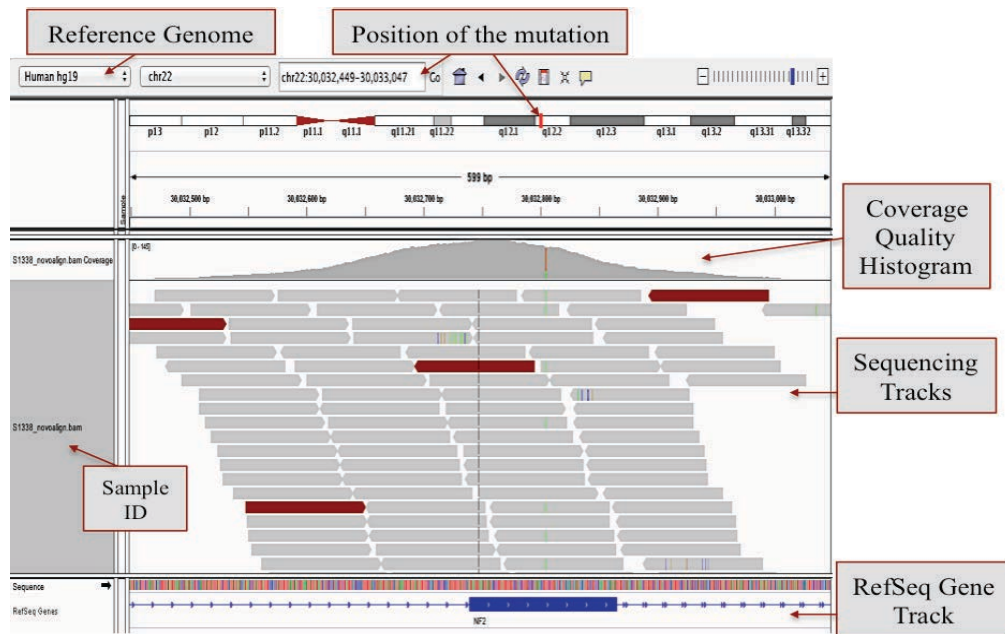
B)
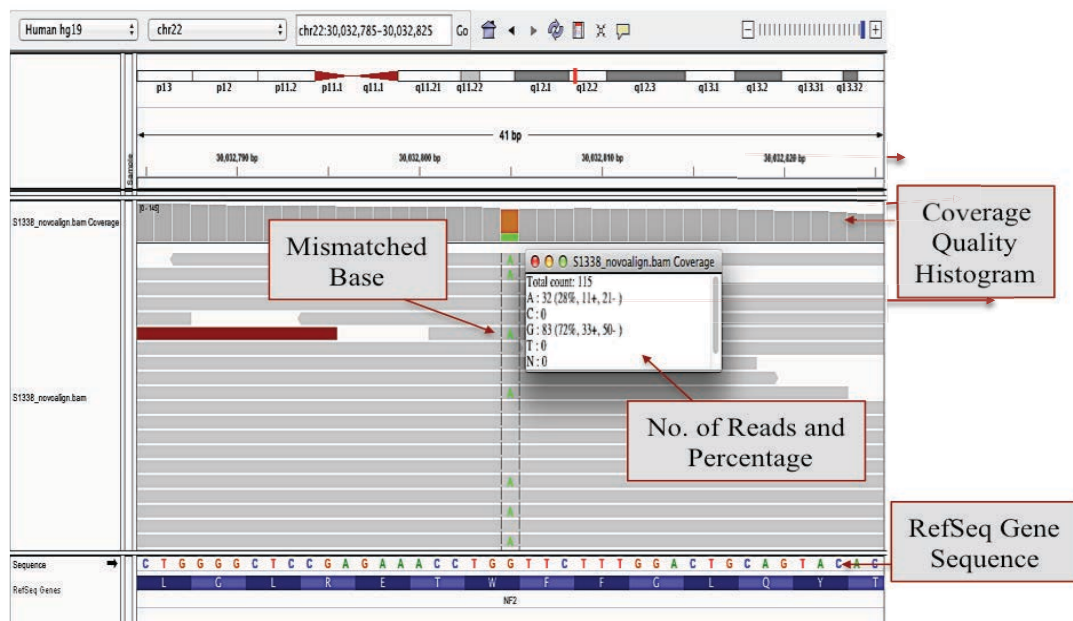
C)
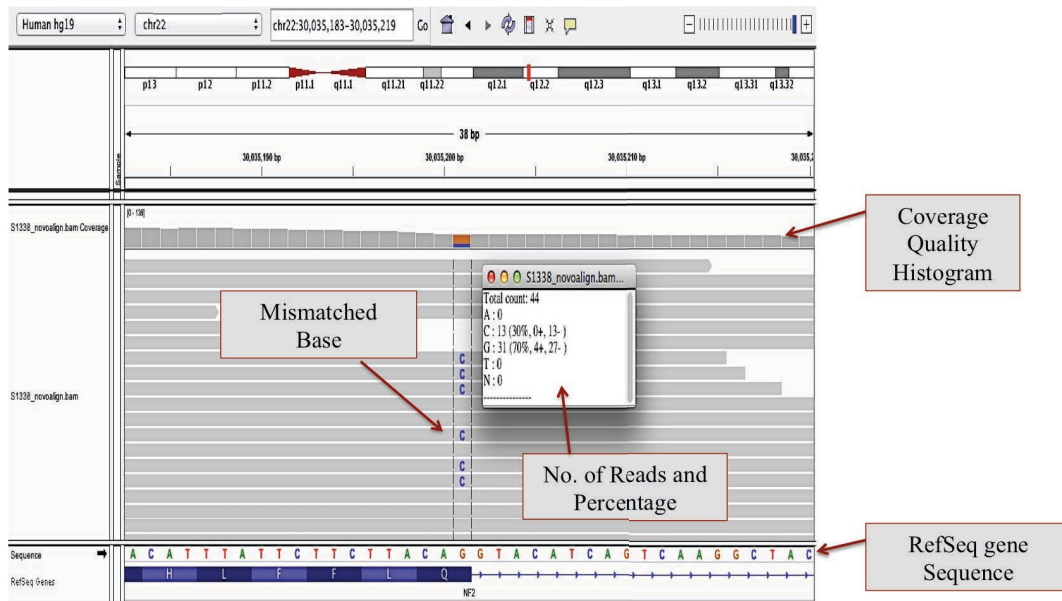
Figure 5.5 IGV display window using *NF2* as an example

The IGV application shows information about the sample according to the zoom level. A) Zoomed out view of *NF2* gene showing the good coverage quality of the gene. B) Zoomed in view where the stop-gain mutation p.W60X is highlighted. The mismatched base (A) is given a different colour (green) and the information window shows that the SNP base (A) appeared 32 times out of 115 reads (28%), while the reference base (G) appeared 83 times (72%). C) Another zoomed in view where the nonsynonymous mutation p.Q121H is highlighted. The mismatched base (C) is given a different colour (blue) and the information window shows that the SNP base (C) appeared 13 times out of 44 reads (30%), while the reference base (G) appeared 31 times (70%).

208

5.3.3 Identification of novel NF2 mutations

After receiving and analysing the data, *NF2* mutations were found in 4 tumour samples. Sample S14T contained a novel stop-gain mutation in exon 11 (p.Q387X), while sample S21T had two novel mutations, stop-gain (p.K15X) and frameshift deletion (p.24_24del), both in exon 1. Sample Lp152T showed a novel stop-gain mutation (p.W60X) and a novel missense (p.Q121H) mutation in exon 2 and exon 3, respectively. The fourth sample, S44T, contained one stop-gain mutation in exon 5 (p.E141X). This later mutation was previously identified as a SNP in various SNP databases (RS74315495) (Table 5.3) (Figure 5.6).

The Human Gene Mutation Database (HGMD) is a free online resource that contains all published gene mutations in addition to some unpublished mutations from around the world. According to the HGMD database, *NF2* contains 386 mutations, the majority of which were frameshift mutations (35%), nonsense mutations (24%), and splice-site mutations (20.5%). On the other hand, *NF2* gene in the COSMIC database contains 1,026 coding mutations that were identified in 31 different tumour tissues (17,692 samples). Based on tissue distribution, the majority of the mutations were identified in schwann/nerve cells (41.55%), Mesothelium tissue (37.65%) and meninges tissue (33.33%). Regarding coding mutation types, the majority were frameshift mutations (37%), 20.76% were nonsense mutations and 11.21% were missense mutations. The majority were confirmed to be somatic mutations.

The *NF2* mutations identified in this study were checked through the HGMD database. The SNP mutation (p.E141X) and the novel stop-gain mutation (p.W60X) were previously reported in NF2 and schwannoma patients respectively in two different studies

(MacCollin et al., 1994) (Jacoby et al., 1994). While none of the remaining novel variants were found in the HGMD database or the COSMIC database.

These mutations were subsequently checked using IGV tools. Interestingly, the stop-gain mutation (p.Q387X) in S14T appeared as a somatic mutation in 25% (2/8) of total reads that showed the mutated base, while the corresponding normal blood sample showed a 100% reference base. The remaining samples had no exome data for the corresponding normal samples, therefore; only the tumour samples were checked in IGV tools. The stop-gain mutation (p.K15X) and frameshift deletion (p.24_24del) of sample S21T were adjacent to each other and the two mutations appeared in 22% (4/18) and 37% (18/49) of the reads, respectively. The stop-gain (p.W60X) and the missense (p.Q121H) mutations of Lp152T were 28% (32/115) and 30% (13/44) mutated in total reads, respectively. Finally, the SNP in sample S44T appeared in 60% (75/125) of the total reads.

In order to confirm the presence of these variants in the samples, primers were designed around the mutation sites and the tumour samples and their corresponding normal samples were sequenced using Sanger sequencing. The results from the sequencing confirmed the presence of these variants in the tumour samples only but not in the corresponding blood samples (Figure 5.7).

It was noticeable that the stop gain mutations p.Q387X, p.K15X and p.W60X for S14T, S21T and Lp152T, respectively, were very weak. Low detection frequency is common in somatic hits in cancer and has previously been reported in other papers (Thomas et al., 2007 & Lim et al., 2014). This is possibly due to the heterogeneous mixture of normal and tumour tissues in one sample and also the low sensitivity of conventional sequencing, which makes it difficult to identify somatic mutation in tumours.

The sequencing results were able to confirm the mutations and therefore, these samples were excluded from further analysis. However, these findings confirm the validity of the whole-exome sequencing for identifying variants in tumour samples.

Sample ID	Mutation Type	CDS Mutation	AA Mutation	Novelty	Percentage of Reads in IGV
S14T	Stop-gain	c.C1159T	p.Q387X	Novel	25%
S21T	Stop-gain	c.A43T	p.K15X	Novel	22%
	Frameshift	c.71-72del	p.24_24del	Novel	37%
Lp152T	Stop-gain	c.G180A	p.W60X	Novel	28%
	Nonsynonymous	c.G363C	p.Q121H	Novel	30%
S44T	Stop-gain	c.G421T	p.E141X	RS74315495	60%

Table 5.3: *NF2* **mutations identified in exome sequencing data**

Mutation type, the position of mutations in cDNA and amino acid sequence, and novelty of the variants are shown in the table along with the percentage of mutation appearance in total reads.
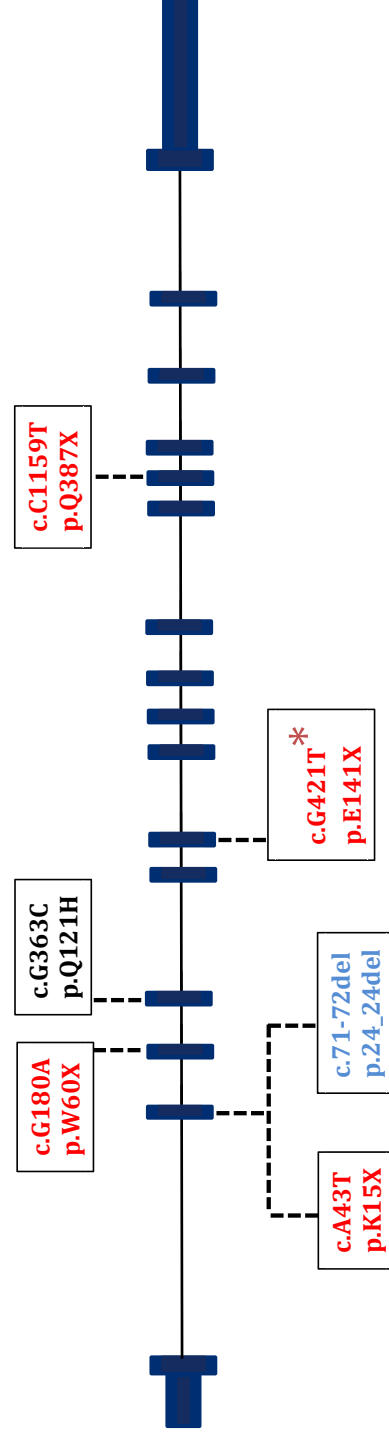
Figure 5.6 Schematic diagram of *NF2* gene showing the identified SNV

A schematic diagram of *NF2* (blue) showing the location of the identified SNV from whole exome sequencing. Stop-gain mutations (red) were identified in exon1, exon2, exon5 and exon11. A frameshift mutation (blue) was identified in exon1 and a nonsynonymous mutation (black) was identified in exon3. All these variants were novel except for the stop-gain mutation, p.E141X (with a red star), which was a common SNP (RS74315495).

Normal Sample S14S — Tumour Sample S14T

c.C1159T: p.Q387X

Normal Sample S21S — Tumour Sample S21T

c.A43T: p.K15X

Normal Sample S21S — Tumour Sample S21T

c.71-72del: p.24_24del
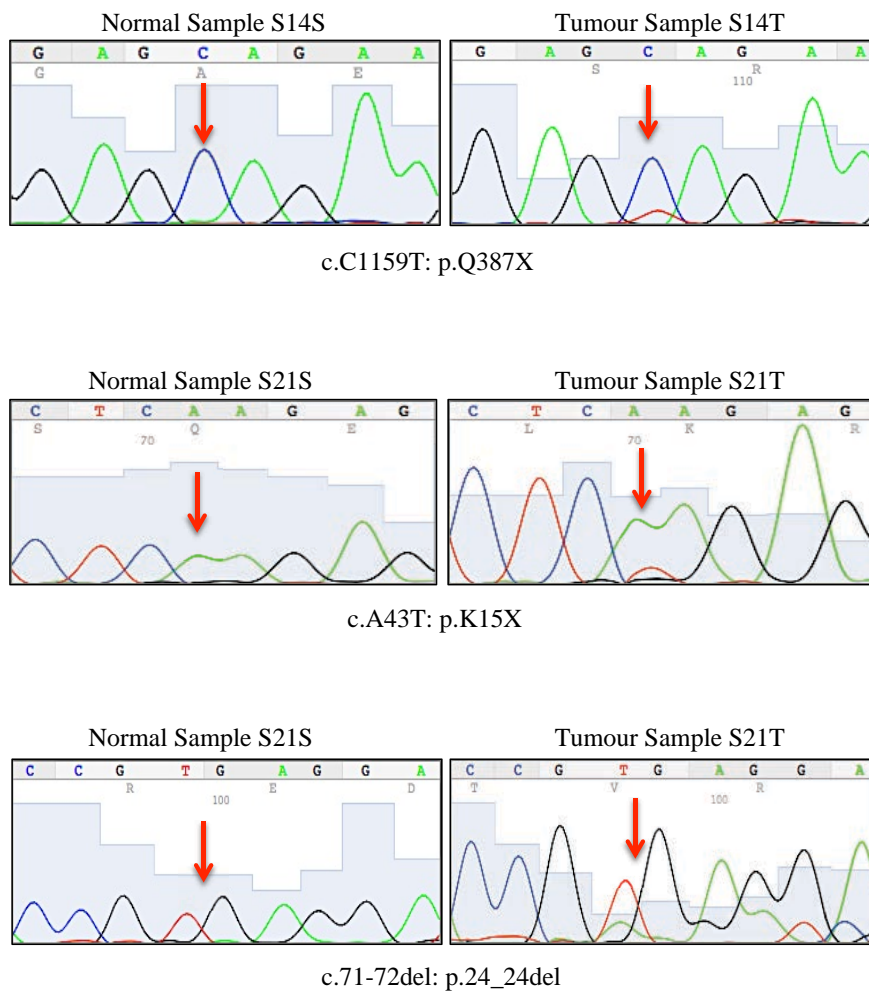
Figure 5.7 Identification of *NF2* mutations in schwannoma

Sanger sequencing confirms the presence of *NF2* mutations in schwannoma tumours. The corresponding normal samples show a wild type sequence. The mutation point is indicated by a red arrow.

5.3.4 SNV in cancer pathways

Genes involved in cancer pathways have been discovered and are often mutated in more than one type of cancer. Therefore, this part of the study focused on genes in major cancer pathways.

In order to identify the mutations present in genes involved in cancer pathways, a list of genes from 10 major cancer pathways were collected from the KEGG cancer database (http://www.genome.jp/kegg/). Analysis of the exome datasheet identified 3 genes (2 novel variants and 1 SNP) that were involved in the SWI/SNF pathway, 6 genes (4 novel variants and 2 SNPs) in the Hippo pathway and 2 genes (1 novel variant and 1 SNP) in the VEGF signalling pathway (Table 5.4).

The 3 nonsynonymous SNVs in the SWI/SNF pathway were found in the same sample (S36T). The mutations of *ARID1B* (p.D172A) and *SMARCE1* (p.R297H) were novel mutations, while the mutation p.V766G of *SMARCA4* was a known SNP (RS201128299) with no MAF data in the 1000 Genomes Project website. Interestingly, the in silico PolyPhen-2 prediction tool predicted these mutations to be "possibly damaging" to the corresponding proteins. Furthermore, the p.R297H mutation in *SMARCE1* was previously found in the COSMIC database in a colon tumour sample (n=1) which was confirmed to be a somatic mutation. To confirm the existence of these mutations in the samples, Sanger sequencing was performed in the tumour sample and in the corresponding normal sample using primers covering the mutation sites. The results of the electropherograms showed that the two mutations p.D172A and p.V766G in *ARID1B* and *SMARCA4*, respectively, were false positive. On the other hand, the p.R297H

215

mutation in *SMARCE1* was identified as a germ-line mutation and existed in the tumour and the corresponding blood samples (Table 5.4A).

The 6 identified mutations of the Hippo pathway genes were nonsynonymous SNVs and were also checked in the COSMIC database and PolyPhen-2. Two genes from the RASSF family members were identified in the list, *RASSF4* and *RASSF1A*. The p.V224I mutation of *RASSF4* was a novel mutation, while the p.R106Q mutation of *RASSF1A* was a known SNP with MAF=0.002. Neither mutations had been reported previously in any other type of cancer, according to the COSMIC database, and also based on the PolyPhen-2 prediction tool, these missense mutations were predicted to be "probably damaging" to the proteins. The remaining 4 identified genes of the Hippo signalling pathway were cadherin related genes, *FAT1*, *FAT3*, *FAT4* and *DCHS2*. The 3 mutations, p.R4359W, p.N1035S and p.R3342Q, of *FAT1*, *FAT3* and *FAT4*, respectively, were novel nonsynonymous SNVs, while the mutation in *DCHS2* (p.R263W) was a nonsynonymous common SNP (RS199621086). Interestingly, according to the COSMIC database, the same mutations of *DCHS2* and *FAT4* had been identified previously in other cancers. The p.R263W mutation in *DCHS2* was identified as a somatic mutation in the breast cancer sample and the p.R3342Q mutation in *FAT4* was previously reported in endometrium carcinoma and colon cancer samples. In order to confirm these mutations in the tumour samples, primers were designed to cover the mutation sites and sequencing was performed in the tumour/normal pair samples. Sanger sequencing was able to confirm all these mutations in the tumour samples as well as in the corresponding blood samples (Table 5.4B), hence these were germ-line changes in our samples.

Analysis of VEGF signalling pathway genes was interesting at the beginning. The identified SNV of *MAPK12* was a novel frameshift insertion mutation (p.V33fs), while the nonsynonymous p.S611P mutation of *ARNT* was found in two different tumour samples. Primers were specifically designed to cover the mutation sites and samples were sequenced to confirm the presence of these mutations. The frameshift mutation in *MAPK12* was a germ-line mutation that was identified in the tumour and the corresponding normal blood samples. The missense mutation was a false positive and was not detected by the conventional Sanger sequencing in both samples (Table 5.4C). The presence of this mutation in more than one sample could possibly be due to an error of exome capture sequencing. The summary of the identified cancer pathway's genes and the results from the Sanger sequencing are listed in Table 5.4.

A)

Gene	Mutation	Novelty	Sequencing Results	Tumour Sequencing	Normal Sequencing
SWI/SNF					
ARID1B	p.D172A	Novel	False Positive		
SMARCE1	p.R297H	Novel	Germ-Line		
SMARCA4	p.V766G	SNP	False Positive		

B)

Hippo Pathway

Gene	Mutation	Novelty	Sequencing Results	Tumour Sequencing	Normal Sequencing
FAT1	p.R4359W	Novel	Germ-Line		
RASSF4	p.V224I	Novel	Germ-Line		
FAT3	p.N1035S	Novel	Germ-Line		
FAT4	p.R3342Q	Novel	Germ-Line		
RASSF1A	p.R106Q	SNP	Germ-Line		
DCHS2	p.R263W	SNP	Germ-Line		

C)

Gene	Mutation	Novelty	Sequencing Results	Tumour Sequencing	Normal Sequencing
	VEGF				
MAPK12	p.V33fs	Novel	Germ-Line		
ARNT	p.S611P	SNP	False Positive		
ARNT	p.S611P	SNP	False Positive		

Table 5.4 SNV identified in cancer pathways

These tables show a list of SNVs in genes involved in cancer pathways. A) Three genes in the Hippo signalling pathway. B) Six genes in the SWI/SNF pathway. C) Two genes in the VEGF pathway. All tables also show the position and the novelty status of each mutation. In addition are the Sanger sequencing electropherograms of the tumour and the corresponding normal samples. For each case, a red arrow indicates the location of the mutation.

5.3.5 Variants located on chromosome 22

Genetic alterations of schwannoma are mainly associated with the loss of heterozygosity of chromosome 22 and mutations of *LZTR1, NF2* and *SMARCB1* which are associated with familial or sporadic schwannoma (Piotrowski et al., 2013). Therefore, variants located on chromosome 22 were investigated to identify schwannoma predisposing genes. Fifteen variants (1 frameshift, 1 splice-site and 13 nonsynonymous) located on chromosome 22 were identified from the list and none of the variants had ever been identified previously in any SNP database except for the p.I739V mutation of *CECR2* (RS199565531), which had been identified in 0.2% (0.0002) of the European American population. By looking at the TCGA data and EVS, genes that showed excessive SNPs and a low number of mutations in the TCGA cancer data were excluded. This reduced the list to 8 variants. Also, the splice-site mutation (c.1603-3->TCCC) was checked using the splice-site prediction tool and the deletion of 4bp from the intron showed no effect on the splicing. Therefore, this splice-site was not inspected further. Afterwards, the remaining variants were checked using IGV tools to reduce the chance of finding false positive results. The identified variants were found in tumour samples that had no exome data for their matching normal blood sample. Therefore, variants in each sample were investigated by IGV tools individually. Variants that showed less than 15% of total reads and low coverage of the mutation site were excluded due to the high chance of false positive mutations being detected by conventional sequencing. The remaining candidates were one novel frameshift insertion mutation (p.V33fs) of *MAPK12* (previously analysed in section 5.3.4) and 4 nonsynonymouse mutations, the p.V258I mutation of *MCM5*, the

p.L333P mutation of *PRODH*, the novel p.V265A mutation of *ACO2* and the novel p.E283K mutation of *XBP1*. These mutations showed a high percentage of mutation reads and good coverage according to the IGV tools (33%, 56%, 60%, 60% and 45%, respectively). The Polyphen-2 prediction tool predicted these nonsynonymous mutations to be probably damaging to the protein. To analyse these mutations in the samples, primers were specifically designed to cover the mutation site and Sanger sequencing was performed in the tumour and matching normal samples. Sanger sequencing was able to identify these mutations in the tumour samples as well as in the matching normal blood samples (Table 5.5). Finding germ-line mutations in all samples was unfortunate; however, these results indicated the importance of using paired samples when performing whole exome sequencing for somatic mutation identification.

Gene	CDS Mutation	AA Mutation	Sequencing Results	Tumour Sequencing	Normal Sequencing
MAPK12	c.99_100insG	p.V33fs	Germ-Line		
MCM5	c.G772A	p.V258I	Germ-Line		
XBP1	c.G847A	p.E283K	Germ-Line		
ACO2	c.T794C	p.V265A	Germ-Line		
PRODH	c.T998C	p.L333P	Germ-Line		

Table 5.5 Identification of variants located on chromosome 22

Sanger sequencing of the schwannoma tumour and its matching normal sample is shown above for each gene. Amino acid mutation along with the corresponding cDNA sequence mutation is mentioned in the table and is indicated by a red arrow in the electropherogram.

5.3.6 High impact variants

Analysis of the remaining exome sequencing data was started by investigating the most deleterious variants. Analysis of high impact (non-sense, frameshift and splice site) variants was conducted by applying several criteria (Figure 5.8). To reduce the large number of variants in the list (n=238), genes were initially checked through the exome variant server. Generally, genes with an excessive number of high impact SNPs are more likely to have another germ-line mutation when new variants are identified within the gene. Therefore, variants corresponding to such genes were removed from the list. This step reduced the list to 97 variants. After that, the TCGA and COSMIC databases were used to check whether these genes contained several somatic mutations in different cancer types. Genes that showed none or a low number of mutations in the COSMIC database or with a mutation frequency of <2.5% in less than 5 types of cancer in the TCGA database were also removed. The remaining 62 variants consisted of 14 frameshift variants, 4 stop-gain variants and 44 splice-site variants. All splice-site variants within the list were checked using the splice-site prediction tool and those showing no effect on the splicing were excluded. Moreover, variants with MAF <0.001 and with rs ID were checked in the 1000 Genomes Project database to investigate the frequency of these mutations in the Spanish population. All these variants have never been reported previously in the Spanish cohort. Finally, to eliminate false positive results, the remaining 22 variants were analysed using IGV tools to investigate the quality and the reading coverage of the sequencing. Variants with poor coverage in the tumour samples or that

were present in >5% of the normal blood samples (for the samples that had matching normal blood samples) were removed from the list.

After applying these stringent criteria, the overall variants within the list were 10 novel frameshift, 4 novel stop-gain and 3 splice-site (2 novel) variants (Table 5.6).
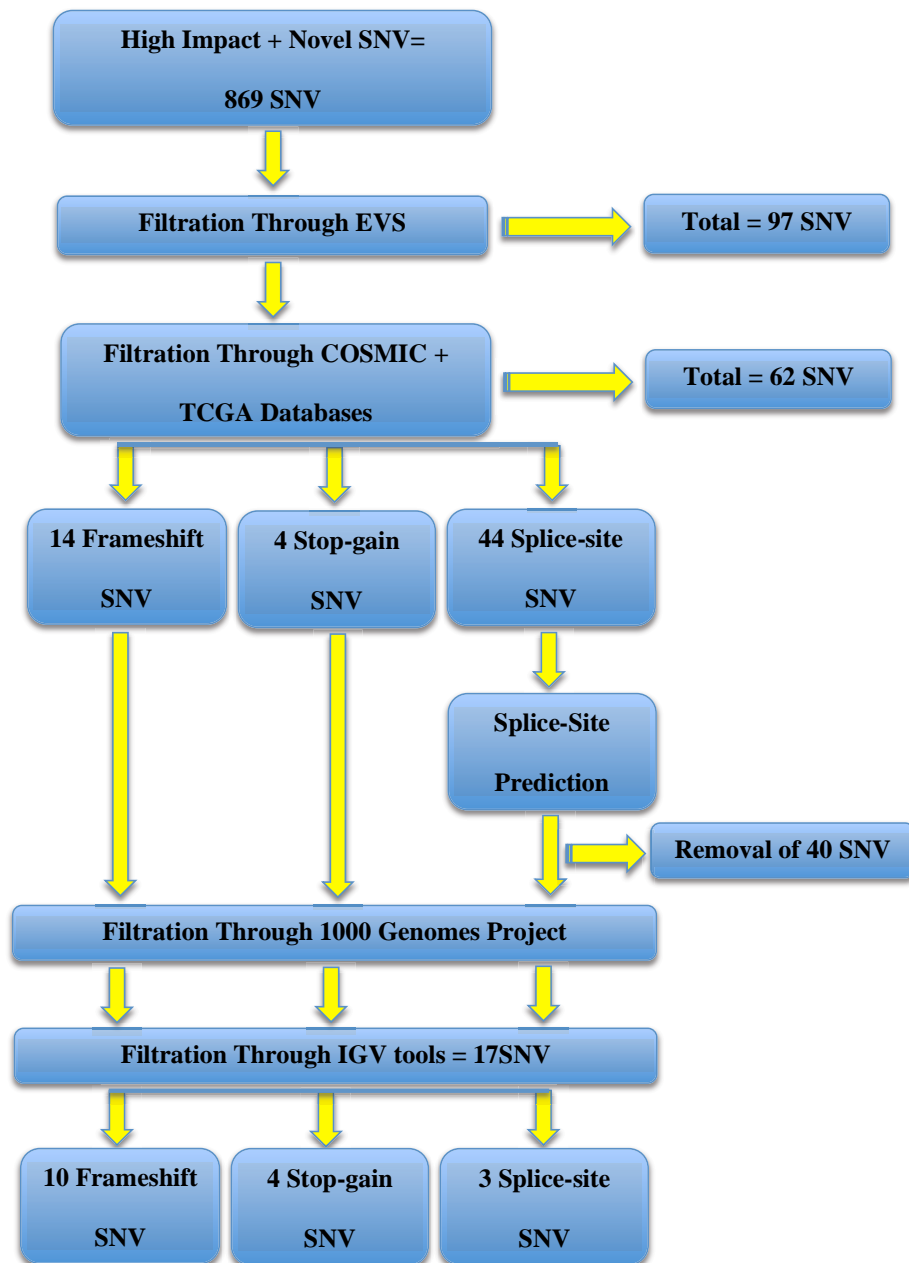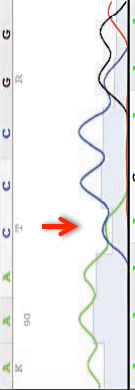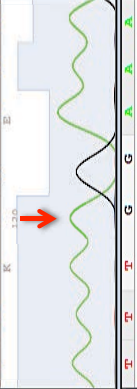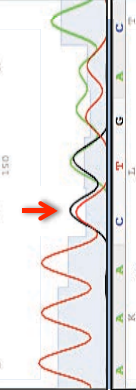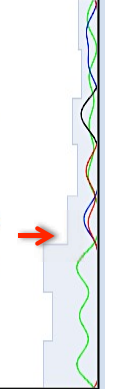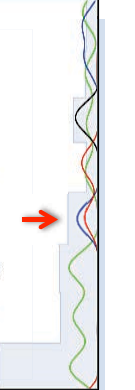
Figure 5.8 Filtration process applied to reduce the number of high impact variants
A combination of manual and web-based tools were applied in the filtration process.

Interestingly, the novel splice-site variant (c.121-3->TTTTTTT) of *CACNB2* was found in two schwannoma samples and was predicted to affect the splicing, according to the splice-site prediction tool, and the region was read 36 times when analysed by IGV tools (Figure 5.9A). *CACNB2* (calcium channel, voltage-dependent, beta 2 subunit) was considered as a good candidate since it is involved in the MAPK signalling pathway and its protein functions as calcium ion transport and cell-cell signalling (Figure 5.9B). To confirm the presence of the variant in the tumour samples, primers were designed around the mutation point and then tumours and their matching normal samples were sequenced by conventional Sanger sequencing. When analysing the electropherograms, this mutation was not found in either of the two samples (Figure 5.9C). This could be due to a technical error of exome sequencing in this region since the exon-intron boundary contains multiple A bases.

The remaining candidate variants within the list were analysed by Sanger sequencing in the tumour and matching normal blood samples. *MAPK12* is involved in the cancer pathway and was analysed previously in section 5.3.5. However, the results detected in the tumour and normal samples were either false positives or germ-line variants (Table 5.6).

Figure 5.9 Analysis of the splice-site variant in *CACNB2*

A) The IGV tool displays the total reads and the coverage quality of the mutation area. B) *CACNB2* is in the upstream of the MAPK signalling pathway (indicated by a red arrow). The source of the pathway picture is DAVID Bioinformatics; http://david.abcc.ncifcrf.gov/. C) Sanger sequencing covers the mutation area in 2 tumour samples. The wild-type sequence contains 25 (A) bases at the exon-intron boundary site, while the mutated sequence has an insertion of extra 7 (A) bases. The assumed splice-site insertion variant is indicated by an arrow.

Frameshift SNV

Gene	CDS Mutation	Sequencing Results	Tumour Sequencing	Normal Sequencing
LNX1	c.6_9del	Germ-line		
CYP3A4	c.1457_1458insA	Germ-line		
CFHR5	c.486_487insAA	False Positive		
SAMD9	c.3645_3646insT	Germ-line		
NRD1	c.3451_3454del	Germ-line		

USP40	c.8_9insT	Germ-line
PLEC	c.60_61insC	Germ-line
COPZ2	c.17_18insC	False Positive
TNFAIP6	c.832_833del	Germ-line

Stop-gain SNV

| *CYBA* | c.G433T | False Positive |
| *SYTL2* | c.2537_2538insA | Germ-line |
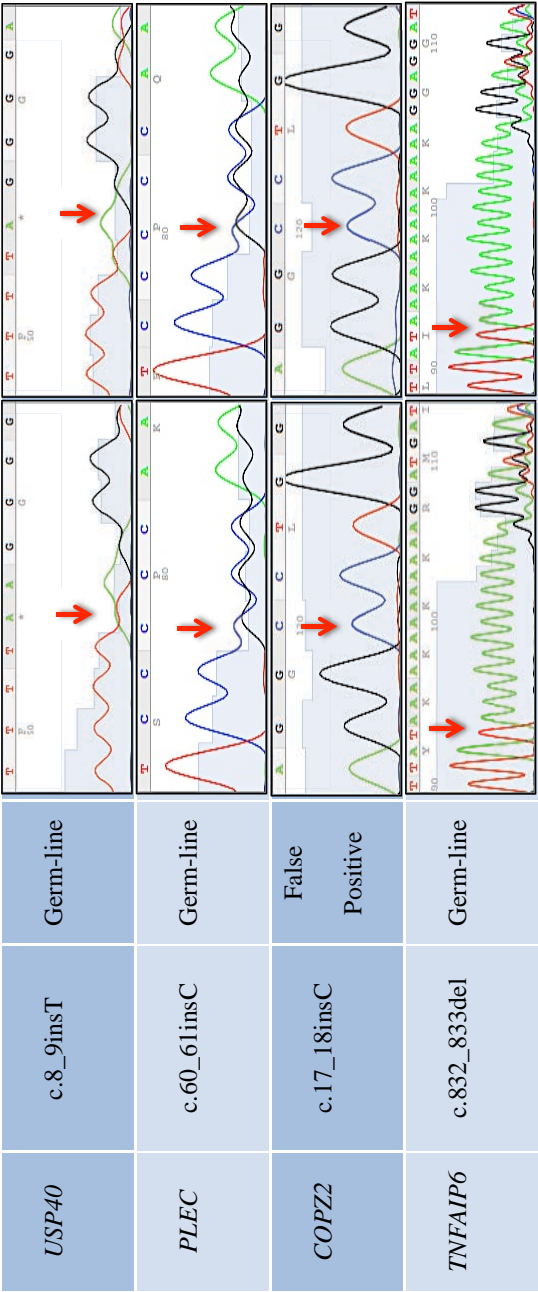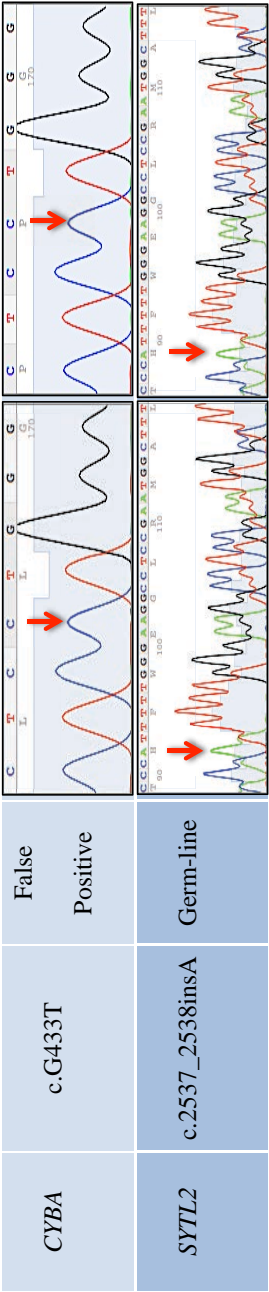
ATRAID	c.G3 1T	Germ-line
FASN	c.8_9insT	False Positive

Splice-site SNV

ABCE1	c.406-2A>-	Germ-line
SNX30	c.1791+1A>C	False Positive

Table 5.6 A summary of high impact mutations identified in exome data

Genes with frameshift, stop-gain and splice-site mutations are shown in the table along with the position of the mutation in cDNA and the results from the Sanger sequencer. Sequencing electropherograms from schwannoma tumours and their matching normal are shown with an arrow indicating the mutation site.

5.3.7 Nonsynonymous mutations

The remaining missense mutations (n=1383) were investigated at the end of the analysis. The same criteria as above (section 2.4.6) were applied to the list to reduce the vast number of nonsynonymous variants. In addition, the Polyphen-2 tool was used to assess the effect of the mutation on the function of the proteins. All selected variants ranged from possibly damaging to probably damaging (50% - 99%). Variants predicted to be benign were excluded. The analysis began by looking for recurrent mutations that were found in more than one sample. Two variants that met the criteria were p.S611P of *ARNT* in the S36T and S37T samples and p.Q62P of *RNF10B* in Lp272T and Lp303T. Primers were designed around the region of interest and the tumour and corresponding normal blood samples were sequenced. These mutations were false positives and the reason for identifying a recurrent mutation in more than one sample is due to the errors that occurred during the sequencing (Table 5.7). Next, genes that showed multiple mutations were investigated. A list of 7 genes was identified and they were analysed by Sanger sequencing. The results were either false positive or germ-line variants (Table 5.7).

Afterwards, the remaining nonsynonymous mutations that fit the criteria were investigated. The list contained variants that have previously been analysed in sections 5.3.4 to 5.3.5. Interestingly, among the list, the possibly damaging p.S169N mutation of *FRG1* in the tumour sample Lp303 was previously reported by COSMIC in 3 cases of ovarian cancer and 6 cases of melanoma (COSM225855) and was confirmed to be somatic in all cases (Figure 5.10A). To confirm the presence of the variant in schwannoma, tumour and matching normal blood samples were sequenced through Sanger sequencing. The result of the tumour sample was false positive as conventional

sequencing was unable to identify it (Figure 5.10B). The other nonsynonymous variants within the list were also investigated in the tumour samples and their corresponding normal blood samples were sequenced by Sanger sequencing. The results showed that none of the variants were confirmed to be somatic in any of the samples (Table 5.7). It was noticeable that the majority of the identified nonsynonymous mutations were in the two samples that have no exome data for the matching normal sample (S36 and S37). Having data on the tumour and matching normal samples will reduce germ-line variants, as seen in the other two samples.

Gene	CDS Mutation	AA Mutation	Sample ID	Novelty	Sequencing Results
Recurrent Variants					
ARNT	c.T1831C	p.S611P	S36T	No (no rs ID)	False Positive
			S37T		False Positive
RNF19B	c.A185C	p.Q62P	Lp272T	Novel	False Positive
			Lp303T		False Positive
Genes with Multiple Mutations					
LPRB1	c.A13676G	p.N4559S	S36T	No (RS148848707)	Germ-line
	c.C6568A	p.L2190M	S36T	Yes	Germ-line
	c.A8039G	p.H2680R	S36T	Yes	Germ-line
ENOSF	c.T38C	p.M13T	S36T	No (no rs ID)	False Positive
	c.T20C	p.M7T	S36T	No (no rs ID)	False Positive
	c.T83C	p.M28T	S36T	No (RS201382707)	False Positive
SP7	c.G839T	p.R280L	S36T	No (RS200582631)	Germ-line
	c.C799T	p.R267C	S36T	No (RS201666834)	False Positive
RCSD1	c.T620G	p.L207W	S37T	No (RS191098895)	Germ-line
	c.A1235G	p.Q412R	S36T	No (RS140531524)	False Positive
C2CD4A	c.C554G	p.P185R	Lp272T	Yes	Germ-line
	c.G310C	p.G104R	S37T	No (RS199710631)	Germ-line
MESP2	c.C556G	p.Q186E	S36T	No (no rs ID)	False Positive
	c.G554C	p.G185A	S36T	No (no rs ID)	False Positive
LINGO1	c.C1284A	p.D428E	S37T	Yes	False Positive
	c.C1184G	p.T395R	S37T	Yes	False Positive
Nonsynonymous Variants					
NRXN2	:c.C31G	p.P11A	Lp272T	Yes	False Positive
GALNT12	c.G182A	p.R61H	LP303T	Yes	Germ-line

MLL2	c.G11129C	p.G3710A	S36T	Yes	Germ-line
RHOU	c.G391A	p.V131M	S37T	Yes	Germ-line
NQO2	c.T116C	p.V39A	S36T	Yes	Germ-line
EPPK1	c.C3577T	p.R1193C	S37T	No (no rs ID)	Germ-line
RETN	c.A245G	p.D82G	S36T	No (no rs ID)	False Positive

Table 5.7 Nonsynonymous variants analysed in non-*NF2* schwannoma

Different approaches were applied to analyse nonsynonymous variants; recurrent variants, genes with multiple variants and single variants. The position of mutation in the cDNA and amino acid sequence, sample ID, novelty status and Sanger sequencing result are shown in the table.

A)
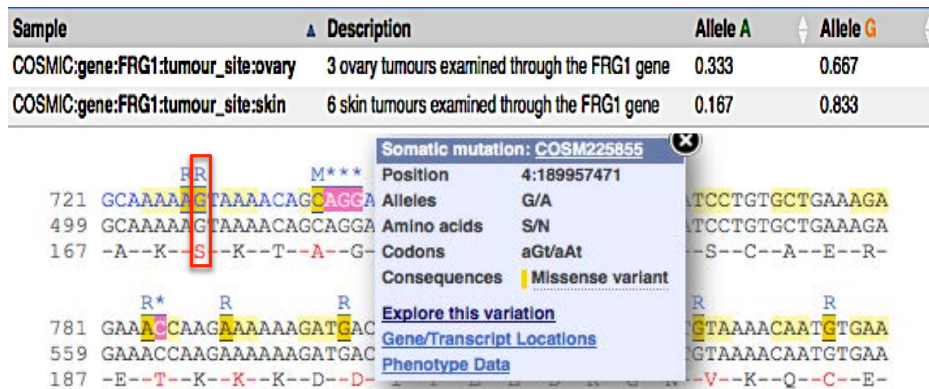
B)

Lp303 Schwannoma Tumour   **Lp303 Normal Blood**
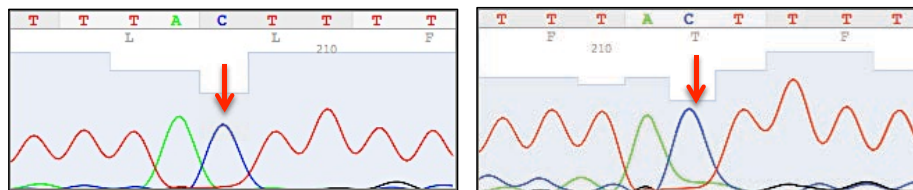
Figure 5.10 Analysis of the non-synonymous mutation p.S169N of *FRG1*

A) The c.G506A SNV was previously identified in the COSMIC database in skin and ovarian cancer (COSM225855). The position of the mutation in cDNA is marked in red. B) Electopherogram of the mutation site. The red arrow indicates the mutation site.

5.3.8 VarScan2 tool

In-silico analysis of the exome sequencing data did not reveal any somatic mutation in schwannoma tumours. Using simple steps to subtract the variants present in tumour and matching normal samples, or applying certain filtration criteria is probably inadequate for exome sequencing analysis. Many algorithmic tools have been developed to evaluate the next generation sequencing data of normal/tumour paired samples in order to detect somatic mutations (Koboldt et al. 2012).

VarScan2 is an exome data analysis tool for somatic mutations and the detection of copy number alterations (section 2.9.5.3). It reads normal/tumour pair samples simultaneously to identify somatic/germ-line mutations and loss of heterozygosity (LOH) events. The analysis of exome data by VarScan2 was conducted by bioinformatician Michael Simpson of Kings College London. In brief, when Illumina exome sequencing reads are generated and aligned against reference sequences, the output reads will be converted to SAM (Sequence Alignment/Map) and then to BAM (binary version of SAM) file formats by the SAM tools package. Afterwards, the VarScan2 algorithm analyses BAM files from tumour and matching normal samples simultaneously by comparing the base cells and the depth of the sequence at each base position. To determine the somatic status of a variant, certain criteria were applied for coverage and base quality. Variants present in >15% of total reads in the tumour sample and <5% in the matching normal sample were classified as somatic, common variants found between tumour and matching normal samples were classified as germ-line, and variants found to be heterozygous in the normal sample but homozygous in the tumour sample were considered as LOH (Figure 5.11).
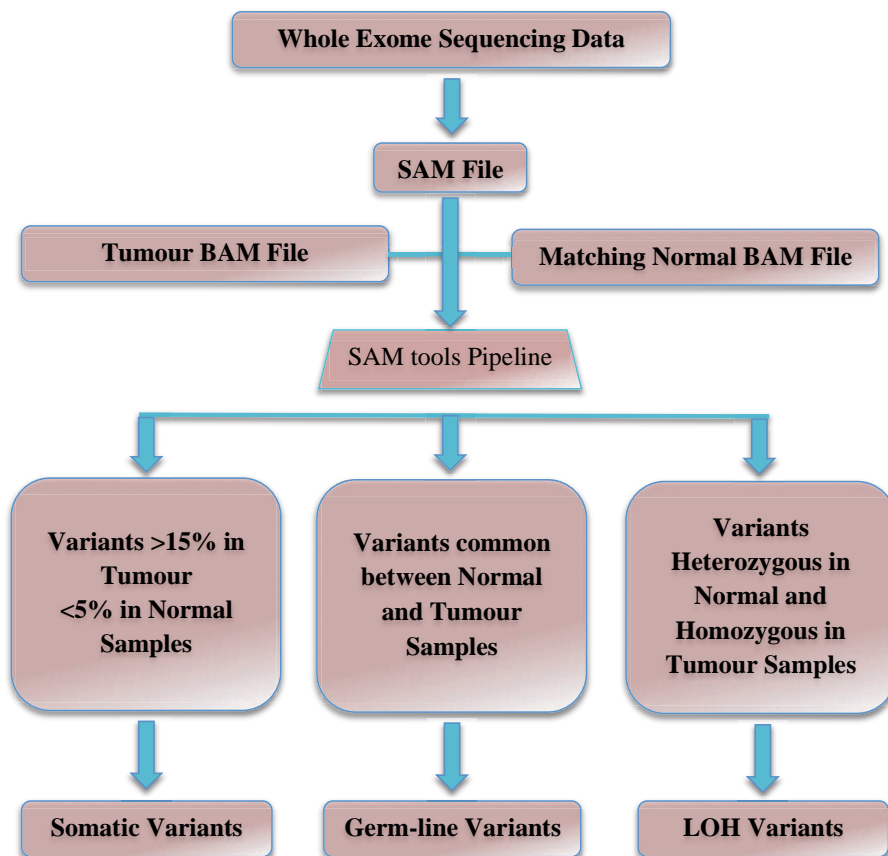
Figure 5.11: VarScan2 variants detection algorithm

A flow chart showing the steps for the identification of variants from whole exome sequencing using the VarScan2 tool.

In the paired Lp272 samples, VarScan2 identified 6 LOH and 19 somatic hits. The identified somatic calls were 1 in UTR3, 1 in UTR5, 2 intronic, 9 splicing and 7 exonic variants (4 synonymous, 2 nonsynonymous, and 1 nonframeshift deletion). On the other hand, the identified somatic variants in Lp303 paired samples were 1 intronic, 1 in UTR3, 2 splicing and 3 exonic (3 nonsynonymous) in addition to 10 LOH. In both cases, none of the LOH events were found on chromosome 22 (Table 5.8). None of these variants were identified by in-silico analysis because they did not fit the in-house criteria; nonsynonymous variants and some splice-site mutations were found more than 10 times in in-house data, and in addition, some of the splice site variants were >3bp away from the exon intron boundary.

HRNR was a common gene between the two-paired samples Lp272 and Lp303. One synonymous and one nonsynonymous mutation were found in Lp272, while one nonsynonymous variant was found in the Lp303 sample. All these variants were known SNPs; RS76438416, RS61814943 and RS61814939, respectively, and were identified >10 times in in-house data (Table 5.8). Despite the fact that the 2 nonsynonymous mutations did not match our criteria and both were predicted to be benign to the protein when analysed by PolyPhen-2, the mutation regions were sequenced through Sanger sequencing and were not found in the tumour samples and hence were false positives.

Schwannoma Lp272

Gene	Mutation Type	Mutation Position	Percentage in Normal	Percentage in Tumor	Chr	Detection within In-house Data	Reason(s) for not being analysed
STK33	LOH		16.39%	2.56%	11	2	--
DLGAP3	LOH		18.18%	5%	1	12	--
RALY	LOH		17.54%	4.17%	20	23	--
CCDC50	LOH		20.59%	3.26%	3	237	--
CMAS	LOH		16.22%	4.88%	12	-	--
L1CAM	LOH		84.31%	95.05%	X	-	--
HRNR	synonymous SNV	c.T5076C	4.17%	16.28%	1	52	synonymouse variant (not deleterious)
HRNR	nonsynonymous SNV	c.C1942A	0%	23.81%	1	33	found >10 times within in-house data (common SNP)
GEMIN2	Splicing	c.406-7C>T	5%	15.38%	14	3	variant is located 7bp away from splice donor/acceptor sequence
ASGR2	Intronic		0%	19.23%	17	3	variant is located in non-coding region
IRS1	synonymous SNV	c.A3609C	4.88%	15.91%	2	33	synonymouse variant (not deleterious)
ZDHHC11	UTR3		0%	16.67%	5	7	variant is located in non-coding region
HLA-DRB1	Splicing	c.371-10T>C	5%	16.67%	6	76	variant is located 10bp away from splice donor/acceptor sequence + found >10 times within in-house data (common SNP)
REXO1L1	synonymous SNV	c.C1869T	0%	29.03%	8	7	synonymouse variant (not deleterious)
AR	nonframeshift deletion	c.1408_1419de	0%	15.22%	X	15	found >10 times within in-house data (common SNP)
PTEN	Splicing	c.802-4T>-	0%	22.22%	10	-	variant is located 4bp away from splice donor/acceptor sequence
AKR1D1	Splicing		4.55%	21.82%	7	-	variant is located far from splice donor/acceptor sequence
KRTAP9	Intronic		0%	15%	17	-	variant is located in non-coding region
FHDC1	synonymous SNV	c.A219C	0%	19.23%	4	-	synonymouse variant (not deleterious)
SASH1	Splicing		0%	16.13%	6	-	variant is located far from splice donor/acceptor sequence
UFL1	Splicing	c.1986-4T>-	4.55%	17.46%	6	-	variant is located 4bp away from splice donor/acceptor sequence
B4GALT6	Splicing	c.233-9C>T	0%	20%	18	-	variant is located 9bp away from splice donor/acceptor sequence
ATM	Splicing	c.2922-9T>-	4.55%	19.19%	11	-	variant is located 9bp away from splice donor/acceptor sequence
SLC2A2	Splicing		0%	15.38%	3	-	variant is located far from splice donor/acceptor sequence
HES1	UTR5		0%	15.38%	3	-	variant is located in non-coding region

Schwannoma Lp303

Gene	Mutation Type	Mutation Position	Percentage in Normal	Percentage in Tumor	Chr	Detection within in-house Data	Reason(s) for not being analysed
SMG7	LOH		22.95%	3.26%	1	17	--
KRTAP4-11	LOH		22.73%	5%	17	5	--
RPAIN	LOH		18.52%	3.39%	17	22	--
POP4	LOH		16.25%	3.53%	19	3	--
DPY19L4	LOH		16.39%	3.12%	8	1	--
TNFAIP8	LOH		15.91%	4.55%	5	-	--
ATP8B3	LOH		21.05%	3.64%	19	-	--
TARS	LOH		18.75%	4.55%	5	-	--
ARHGEF26	LOH		15.38%	5%	3	-	--
NUCKS1	LOH		19.05%	4.35%	1	-	--
HRNR	nonsynonymous	c.G6160A	4.44%	20.31%	1	90	found >10 times within in-house data (common SNP)
TCIRG1	nonsynonymous SNV	c.G754C	0%	31.15%	11	16	found >10 times within in-house data (common SNP)
KLHDC8A	nonsynonymous	c.C399G	4.35%	15%	1	76	found >10 times within in-house data (common SNP)
BDP1	Splicing	c.6023-7->T	4.26%	15.22%	5	23	variant is located 7bp away from splice donor/acceptor sequence + found >10 times within in-house data (common SNP)
BRAF	Intonic		4.65%	22.64%	7	1	variant is located in non-coding region
ARHGAP1	UTR3		3.23%	15.56%	11	-	variant is located in non-coding region
ADAMTSL1	Splicing	c.3935-9C>-	5%	15.62%	9	-	variant is located 9bp away from splice donor/acceptor sequence

Table 5.8 Somatic variants identification in schwannoma paired samples using VarScan2 tool

This table shows the analysis of 2 schwannoma paired samples, Lp272 and Lp303, using VarScan2. The type of mutation (LOH, exonic, splicing and others), position of the variants, chromosome and number of times detected within in-house data are shown above. All detected variants were found in <5% in the normal sample and >15% in the tumour sample and none of the LOHs were found on chromosome 22. The reasons for not analysing the variants further are also explained and shown in the table. *HRNR* gene was found in the two paired samples and variants were indicated by a red arrow.

241

5.4 Discussion

5.4.1 General discussion of main findings

Sporadic vestibular schwannoma is a benign tumour that is mainly affected by mutations in the tumour suppressor gene *NF2* or/and chromosome 22 LOH. However, a small percentage of sporadic vestibular schwannoma showed no alterations in NF2 gene. Mutations of other genes on chromosome 22 have been reported in schwannoma cases such as *SMARCB1*, which has been found in 50% of familial schwannoma and 10% of sporadic (Rousseau et al., 2011; Smith et al., 2012), and *LZTR1*, in familial schwnomatosis (Piotrowski et al., 2014). These different gene alterations indicate that genes on chromosome 22 play a critical role in vestibular schwannoma development.

Researchers are always looking for faster and cost-effective technologies to understand human disorders including human cancer. The advances in NGS have allowed this technique to characterise the molecular background of different cancer types mainly through whole genome sequencing (WGS) and whole exome sequencing (WES) techniques. While WGS is expensive for most researchers, WES is able to produce vast protein coding alterations data at a relatively low price. The application of exome sequencing in cancer studies has identified novel cancer causative mutations and genetic alterations that contribute to cancer development such as significant mutations in breast cancer (Banerji et al., 2012), colon cancer (Melissa S Derycke, 2013), leukaemia (Chang et al., 2013; Yan et al., 2011), lung cancer (Zheng et al., 2013) and many other cancer studies that are under study through TCGA.

In this study, whole exome sequencing was used in combination with bioinformatics tools to identify causative mutation for schwannoma tumours that do not contain *NF2* mutations. Different approaches and strategies were applied in order to increase the probability of identifying the causing variant. *NF2* mutations were identified in half of the samples and the remaining non-*NF2* samples showed no somatic mutation in any of the genes. Most of the identified variants were either germ-line mutations or false-positive variants. The identification of this amount of germ-line mutations was surprising, however, finding false positive variants is frequent in such a high-throughput sequencing technique (MacArthur and Tyler-Smith, 2010) and finding hundreds of deleterious variants in a healthy human genome is generally common (MacArthur et al., 2012). Despite the widespread success of whole exome sequencing technology, this technology is full of challenges and limitations which will be discussed in the next section.

5.4.2 Possible reasons for failing to identify somatic mutation in non-NF2 schwannoma

There are three possible explanations for failing to identify somatic driver mutation in non-*NF2* schwannomas. First, the causing variant is located in areas not covered by exome sequencing, such as the non-coding regions of DNA. Secondly, the coding mutation in non-*NF2* schwannoma was missed by exome sequencing or during the in-silico analysis and filtration process. Finally, the causing variant was not detected due to tumour heterogeneity.

5.4.2.1 The causative is non-coding mutation

Cancer sequencing studies mostly focus on driver mutations located in protein coding regions. However, a number of recent studies have identified significant cancer mutations in non-coding regulatory regions. A study by Horn and colleagues identified recurrent somatic mutations within the *TERT* promoter region in 33% of melanoma cases in addition to germ-line promoter mutations in familial melanomas (Horn et al., 2013). Moreover, another research group, led by Huang and colleagues, also found 2 recurrent somatic mutations in the promoter region of *TERT* in 71% of melanoma tumours using the whole exome sequencing technique (Huang et al., 2013). Other individual cases with non-coding mutation were also identified such as somatic and germ-line mutations within cis-regulatory genes found in colorectal cancer (Ongen et al., 2014) and the splice-site mutation in the non-coding region of *BRCA2* in familial breast cancer (Bakker et al., 2014).

These findings increase the evidence that mutations in non-coding regulatory regions in the human genome are significant, which may play a crucial role in tumour genesis and cancer progression. It is possible that non-*NF2* schwannomas are derived by mutations in non-coding regulatory genes that affect gene expression, especially regulatory elements on chromosome 22. This can be achieved by using whole genome sequencing, however, the whole genome sequencing is expensive and analysing the vast genome data is more challenging than the already overwhelming amount of data generated by whole exome sequencing.

An alternative explanation for the failure to identify somatic mutation in non-*NF2* schwannoma is that the causative variant is within a protein-coding gene, however, the

variant was accidently missed by exome sequencing due to technical errors or because of the filtration process during in-silico analysis.

5.4.2.2 Limitations of exome sequencing

Whole exome sequencing covers 95% of the exons throughout the human genome (Rabbani et al., 2014). The application of this sequencing technology in cancer studies has discovered somatic mutations in cancer genomes such as the identification of the recurrent somatic mutation in *IDH1* in glioblastoma (Parsons et al., 2008), *NOTCH1* in head and neck cancer (Agrawal et al., 2011) and others. However, it is still difficult to identify true somatic mutations using this technology due to many reasons, such as low coverage of some regions, tumour/normal contamination, mis-matching, sequencing errors and others.

Sequencing coverage is not uniform in all exons across the genome and even with newly discovered technologies and tools, to date no capture method can provide 100% high sensitivity with full coverage of all exons (Zhi and Chen, 2012). To overcome such problems, targeted sequencing is more effective in such cases because it provides a combination of different hybridization methods to increase sequencing efficiency. Examples of targeted sequencing specialised for cancer studies are Targeted Cancer Sequencing from Illumina and Ion AmpliSeq Cancer Hotspot Panel from Life Technology (Ion Torrent). Both technologies are based on sequencing a customised panel of cancer relevant genes and guarantee uniform coverage throughout the targeted region. Although this technology is cost effective, it is however, not a suitable technique for the identification of mutations in genes outside the panel.

When the exome sequencing data of each sample is generated, the data are aligned to the reference genome to produce SAM files and then BAM files. Typically, exome data match >99% of the bases of the reference genome and the remaining unmatched bases are considered as variants (Patel et al., 2014). Any fault or error during sequencing can cause mis-matches with the reference sequence and produce false positive calls. In contrast, false negative calls are when a variant is not detected due to low sequencing coverage of the mutation region. For instance, repetitive regions are enriched with false positive and false negative variants. Repetitive regions are hard to align and usually bases in such regions can easily be missed due to poor area coverage, but also, these regions can be falsely aligned and the mis-matched bases will be falsely flagged as variants.

False calls are major challenges when analysing exome sequencing data. False negative calls lead to the failure to identify true causative variants. Failure to identify a somatic mutation in schwannoma is possibly due to insufficient coverage. On the other hand, applying the stringent criteria of filtration is required to reduce the false positive variants.

5.4.2.3 In-silico analysis

Whole exome sequencing is able to identify mutations in most protein-coding genes in the genome. In order to reduce the massive number of variants and identify the causative somatic mutation, series of filtering strategy were applied to the exome data. The filtration process was based on stringent criteria to exclude germ-line variants and common SNPs (Figure 5.1). Firstly, many web-tools were used to exclude common variants such as dbSNP and the 1000 Genomes Project database. Other tools were applied to prioritise the variants; COSMIC and TCGA databases for cancer related genes,

Polyphen-2 and splice-site prediction tools for the estimation of variants' pathogenicity. Secondly, the application of the IGV tool was very important to reduce false call variants. Therefore, only variants/genes that were rare (MAF<0.001) or novel, commonly mutated in cancer, predicted to damage the coding protein, and showing good area coverage were analysed.

It is possible that the above criteria could falsely exclude the causing variant. For example, excluding splice-site variants that were >3bp from the exon/intron boundary might exclude some mutations that would have been relevant and possibly somatic. The splicing process is controlled by specific sequences that flank the exon/intron boundary, which are called splice donor and splice acceptor sequences. Mutations within these sequences are deleterious and may lead to the formation of non-functional proteins; therefore, this criterion was selected. However, other cancer studies have identified splice-site mutations that are pathogenic and located many base pairs away from the exons. For instance, a study on 52 cases of neurofibromatosis type 1 (NF1) has found that splice-site variants affecting the splicing process of *NF1* are the most common mutations in NF1 patients at a time when all studies were focusing on protein truncating mutations. In this study, the splice-site mutations ranged from ±1bp to -16bp from the exon (Ars et al., 2000). It is possible that splice-site mutations with different splicing distances are also common in schwannoma patients (neurofibromatosis type 2) in any other gene than *NF2* and analysing this type of mutation further is required.

In addition, the assumptions that only variants affecting the amino acid sequence are important and that synonymous variants have no pathogenic effect on diseases may not be accurate in light of recent data. Generally in cancer studies, synonymous variants are

considered as passenger mutations or have no cancer genesis effect and are rarely investigated. However, a study by Gartner and colleagues on melanoma found a recurrent somatic synonymous mutation in the *BCL2L12* gene and was associated with the overexpression of mRNA. By comparing the protein from WT and mutant *BCL2L12,* the two proteins were functionally similar but cells with a mutated version of the gene showed high cell survival and were more resistant to apoptotic signalling (Gartner et al., 2013). Moreover, a recent study by Supek and colleagues using the exome data of 11 different cancer tissues from the TCGA data portal and COSMIC, suggested that synonymous mutations can be oncogenic by affecting mRNA splicing and thereby altering protein function (Supek et al., 2014). This observation was mainly associated with the overexpression of oncogenes with mutations distributed non-randomly throughout the gene rather than in tumour suppressor genes. These studies indicate the importance of synonymous variants in human cancer and should be investigated in the cancer genomics field. Such analysis was beyond the scope of the present study.

Finally, sample size in this study must be re-considered since the analysis started by using 8 non-*NF2* schwannoma and 3 corresponding normal samples; hence exome sequencing was carried out on a total of 11 samples (3 normal/tumour pairs and 5 schwannoma tumour samples). After the identification of *NF2* mutations in some samples, the remaining samples of 2 normal/tumour paired samples and 2 tumour samples were used for the whole analysis. It is possible to identify somatic mutations with this number of samples, for example the study by Wang and colleagues on cholangiocarcinoma tumour, a type of liver tumour. They identified somatic mutations in *IDH1* and *IDH2* by using whole exome sequencing on one paired sample (Wang et al.,

248

2013). However, even in malignant tumours, such as lung and liver cancers, driver mutations are scarce and difficult to identify. These challenges are increased when screening benign tumours and searching for driver mutations (Zheng et al., 2014). It is more challenging to identify the causing variant in benign tumours with a small sample size.

Finding a mutation causing benign tumours among billions of variants is also challenging. The filtration and prioritization strategies applied in this study were strict. Even with these strict criteria, most of the analysed variants were germ-line mutations.

Despite the limitations of whole exome sequencing and the difficulty in identifying true somatic variants in the present report, this technology has shown fascinating results in the identification of variants in many cancer genomes in a short time frame and for a reasonable cost. The combination of sufficient sequencing coverage, appropriate sample size, stringent filtration criteria and external bioinformatics tools are essential factors to increase the success of the whole exome sequencing study.

5.4.2.4 Tumour heterogeneity

Another challenge that exome sequencing faces in detecting a single somatic mutation in the cancer genome is tumour heterogeneity and/or the contamination of tumour cells by adjacent normal cells. The cancer genome is complex and each clone varies in its genome and mutational stage. Studies carried out on different tumour samples from the same patient found diversity in the intratumour heterogeneity between different tumour sites, as well as between the primary tumour and the distant metastatic organs. This extensive heterogeneity within the single tumour explains the difficulty in understanding cancer

genetics and the discovery of tumour biomarkers (Gerlinger et al., 2012). Also, the genome of tumour cells is different from normal cells. Most studies that use high throughput sequencing technology include tumour and matching normal samples and increase the sample size to overcome such problems and exclude common variants. It was noticeable in the current study that the number of variants in samples without matching normal was high compared to samples with matching normal which lead to the identification of high germ-line mutations in these samples. Generally, mutations in cancer occur at a very low frequency and having a small number of samples reduces the chance of finding significant mutations.

5.4.3 Identification of novel NF2 mutations

The *NF2* gene, which functions as a tumour suppressor, has been widely investigated and many genetic alterations were reported in different types of cancer other than schwannoma, such as pleural mesothelioma (Sekido, 2001), meningioma (Vogan, 2013) and renal cell carcinoma (Dalgliesh et al., 2010). Loss of heterozygosity and other genetic mutations in *NF2* identified in different studies were uploaded to several online databases. To date, the number of identified mutations in *NF2* is 386 in HGMD, 1,026 in COSMIC and 2,124 in dbSNP. In this study, four deleterious mutations were identified in schwannoma samples and, to our knowledge; these mutations have never been reported in any SNP/genetic variation databases. These include the novel stop-gain mutations (p.Q387X and p.K15X), novel frameshift mutation (p.24_24del) and the novel nonsynonymous mutation (p.Q121H). Despite the fact that the objective of this study was to identify somatic mutations in non-*NF2* schwannoma tumours, finding new *NF2*

250

mutations using whole exome sequencing was validation of the ability of exome sequencing to identify significant mutations.

Chapter Six: Discussion

Cancer is one of the leading causes of death worldwide. The latest global statistics, in 2012, have revealed that around 32 million people are living with cancer, 8 million deaths were associated with cancer and 14 million individuals were newly diagnosed with cancer (Ferlay et al. 2015). According to the origin site of the cells, cancer is classified into more than 100 types which are categorised into 5 groups: carcinoma, sarcoma, leukaemia, lymphoma and CNS cancer.

Tumours of the CNS are relatively uncommon; however, malignant high grade tumours are fatal which increases the challenge to determine the genetic factors behind the tumour formation and progression. The number of cases associated with CNS tumours is increasing worldwide, with an estimated 256,000 CNS tumour cases diagnosed in 2012 (Ferlay et al. 2015). The development of techniques and research methods that help to understand the molecular mechanisms of tumour development and the identification of molecular markers are important to reduce mortality and improve patient therapy.

6.1 Genome-wide analysis of sporadic CNS tumours

Several genome-wide technologies have been utilised in cancer research to identify the genetic and epigenetic factors that are associated with cancer development and progression. Next generation sequencing (NGS) has been widely performed to identify single nucleotide variants associated with cancer initiation and progression. Recently, NGS and microarray technologies have been introduced to the field which complement each other. Epigenetic microarray based technology has provided a better understanding

of DNA methylation profiling on a genome-wide scale. Although NGS provides information about genome sequence variations, methylation microarrays identify epigenetic variants and cancer-associated markers. The work in this thesis aimed to characterise: (i) DNA methylation changes in brain cancer, presenting two studies that were performed to identify hypermethylated genes in (A) glioblastoma, and (B) brain metastasis from breast tumours, using Infinium Illumina HumanMethylation450K BeadChip; and (ii) genetic variants associated with sporadic vestibular non-*NF2* schwannoma by utilising whole exome sequencing (WES).

6.1.1 Epigenome-wide analysis of brain cancer

Numerous platforms are available to measure DNA methylation using either locus-specific or genome-wide methylation analysis. Examples of locus-specific DNA methylation analysis include methylation-specific PCR (MSP), Combined Bisulfite Restriction Analysis (CoBRA) and methylight real-time PCR. On the other hand, genome-wide methylation analysis includes either microarray based methods, such as Infinium Illumina or Affymetrix methylation arrays, or non-microarray based methylation methods including Restriction Landmark Genomic Scanning (RLGS), methylated DNA immunoprecipitation (MeDIP), Methylated CpG Island Recovery Assay (MIRA) and bisulfite converted DNA sequencing.

The most recent array, Infinium HumanMethylation450K BeadChip by Illumina, is a revolutionary assay that provides a comprehensive DNA methylation analysis of more than 485,000 CpG sites per sample across the entire genome. Chapters Three and Four

have utilised this most up to date technique to evaluate the methylation profile of brain tumours.

Grade IV glioblastoma multiforme (GBM) is the most fatal form of glioma in adults with a median survival rate of <1 year. However, a small fraction of GBM patients showed a longer survival time of up to 3 years, who are so-called long-term survivors (LTS). In Chapter Three, an evaluation of the methylation profile of 19 short-term survivors (STS) and 16 LTS glioblastoma tumours was conducted to identify methylation changes that contribute to STS patients' poor survival. During the analysis, a subset of LTS glioblastoma (n=5) showed massive hypermethylation across many CpG loci, indicating the existence of the CIMP+ phenotype. Interestingly, these five LTS tumours carried mutations in the *IDH1* gene and fulfilled the criteria of G-CIMP, which agrees with the findings in Noushmehr et al.'s paper. In addition, further analysis between G-CIMP[+] LTS glioblastoma and non-CIMP glioblastoma has identified 535 genes (2,377 CpG loci) that were statistically and significantly associated with CIMP[+]. Analysis of the gene list using a bioinformatics tool demonstrated that these genes are involved in various cellular regulatory processes such as cell migration, cell signalling, cell growth and apoptosis. Also, several genes within the list were associated with various types of cancer and tumorigenesis including colon cancer, epithelial cancer, metastasis, tumour angiogenesis and tumour growth.

This finding leads to the cohorts of the study being divided into three groups: STS tumours, LTS with *IDH*-wt (non-CIMP) and LTS with *IDH*-mut (CIMP[+]). However, the methylome of the three groups of glioblastoma was similar, where the majority of the hypermethylated CpG loci were located within CpG islands and in the promoter regions.

On the other hand, hypomethylated loci resided in the gene body and intergenic regions of the genome.

Furthermore, the study was also able to identify differentially hypermethylated genes (32 CpG loci) between STS and LTS glioblastoma that may be associated with STS patients' poor prognosis. Amongst the list, *NR2F2* was represented by multiple probes located within the CpG island and indicated a significant correlation between gene hypermethylation and poor survival. The association between the loss of *NR2F2* expression and patients' poor prognosis was also reported earlier in ovarian cancer (Hawkins et al. 2013). Hypermethylation of the *INA* gene was also associated with glioblastoma patients' poor survival. Alterations in INA expression have been related to short survival in other grades of glioma and pancreatic tumours (Liu et al. 2014; Figarella-Branger et al. 2012). Also, hypermethylation of the remaining genes has been found in other cancers, such as the Homeobox gene family in glioma, lymphoma and prostate cancer (Wu et al. 2010; Leshchenko et al. 2010; Miller et al. 2003), the transcription factor *TFAP2A* in prostate and head and neck cancer (Bennett, Romigh and Eng 2009), the glutamate receptor gene *GRM6* in ccRCC (Arai et al. 2012) and *DKK2* in Ewing sarcoma (Hauer et al. 2013). Hypermethylation of the generated list of genes in STS glioblastoma cases suggests their potential as prognostic markers and further investigation is required to assess their role in gliomagenesis.

Metastasis is an indication of a patient's poor prognosis. Around 15% to 20% of breast cancer patients acquire brain metastasis with a survival rate of less than 8 months. In Chapter Four, the HumanMethylation450K array was utilised to identify the

hypermethylation changes that promote brain cell metastasis from breast cancer and that are involved in metastasis development to the brain.

Differential analysis between breast cancer brain metastasis tumours and non-metastatic breast tumours from the TCGA data portal was carried out and identified 28 differentially hypermethylated genes (37 probes). Further bioinformatics analysis lead to the selection of 7 candidate genes (*EOMES*, *WDR69*, *MIR125B1*, *DZIP1, SOX1, PHOX2B* and *PRDM13*) from the list to be investigated further in the study. Methylation analysis by CoBRA of 10 breast tumours and their matching brain metastasis samples showed that the candidate genes were hypermethylated in the paired breast tumours and BCBM samples, indicating the significance of the 7 genes in the early stages of brain metastasis formation from breast tumours. Furthermore, lower expression of *EOMES, WDR69, MIR125B1, PHOX2B* and *PRDM13* genes was associated with DNA hypermethylation and down regulation of *DZIP1*, *PRDM13* and *PHOX2B* genes was associated with patients' poor survival. This preliminary study has provided some insights into candidate molecular markers for the prediction of BCBM development in the early stages. However, further biological and functional investigation must be carried out to confirm the involvement of genes in metastasis formation and progression.

6.1.1.1 Processing the HumanMethylation450K BeadChip data

The studies carried out in Chapters Three and Four utilised the most recently developed Illumina Infinium HumanMethylation450K BeadChip. This technology provides a comprehensive quantitative methylation measurement (β value) of each CpG dinucleotide, however, the array contains two differently designed probes (Infinium I and

Infinium II) that each evaluate DNA methylation status differently. Infinium I assay uses two separate probes (methylated and unmethylated) to assess the methylation intensity, while Infinium II assay measures the methylation intensity using one single probe for each CpG site. The difference between the two assays was considered in our studies and several quality control pre-analysis processes were applied to the array data to ensure an accurate analysis. For instance, using LUMI package pipeline to normalize the β value difference between Infinium I/II probes and removing probes with high detection p-values along with individual β value validation, all were applied to overcome any potential faults that may influence downstream analysis. Despite this minor limitation of the 450K array, this technology provides a cost-effective, high throughput epigenome-wide analysis that can assess more than 96% of CpG sites in the genome.

6.1.2 Genome-wide analysis of CNS cancer

NGS has been widely used in cancer research to identify nucleotide variants in genes involved in cancer formation and cancer progression. NGS, through WGS and WES, is a powerful and highly efficient technology that enables researchers to study tumour samples in depth and to identify cancer-associated variants within a short timeframe, especially with the development of targeted cancer sequencing that is based on a customised panel of cancer-associated genes. WES, which applies parallel sequencing of the protein-coding regions, was utilised in Chapter Five on 8 sporadic vestibular schwannoma and 3 DNA samples of matching normal blood. Several bioinformatic tools and different strategies were used to facilitate the analysis and to reduce false positive/negative variants. The analysis was carried out by investigating the SNVs

involved in major cancer pathways, genes located on chromosome 22 (where the *NF2* gene resides), high impact deleterious variants and then the remaining nonsynonymous SNVs. The initial analysis of the tumour samples identified four deleterious mutations in the *NF2* gene within half of the schwannoma samples, while the remaining half showed many germ-line or false positive variants. Identification of several germ-line mutations in our samples was displeasing, however, finding deleterious mutations in the genome of healthy humans has been frequently reported (MacArthur et al., 2012). The identification of multiple false positive variants in our study may be due to technical errors such as poor DNA hybridization, insufficient coverage or false alignment with the reference genome that occurs during the sequencing. Such errors are well known and considered as serious limitations to WGS. However, to overcome such problems in our study, combinations of bioinformatics tools (Polyphen-2 and 1000 Genomes Project database), external software (IGVtools) and stringent filtration criteria were applied to the exome sequencing data to raise the possibility of identifying disease-causing SNVs. Despite the fact that the aim of the study in this chapter was to identify somatic mutations in the novel genes involved in non-*NF2* schwannoma formation, WES was able to identify 4 deleterious variants in the *NF2* gene, and to our knowledge, these variants are novel to all SNP/SNV databases.

6.2 Future work

The most recent Illumina HumanMethylation450K BeadChip was utilised to: (i) investigate differentially methylated genes that may be associated with poor prognosis in STS glioblastoma patients. The study has generated a list of differentially hypermethylated genes in STS tumours and it would be beneficial to confirm these

findings in a separate and larger cohort. Additionally, further work would also be required to determine the functional relevance of the genes in gliomagenesis. Examples of genes from the list are *DKK2, NR2F2* and the transcription factor *TFAP2A* which are potential candidates as biomarkers for disease outcome.

Furthermore, the study has identified a list of CIMP$^+$ genes in a subgroup of LTS glioblastoma tumours. A thorough investigation would be beneficial to determine the association between *IDH* mutation and G-CIMP$^+$ as it would be helpful to understand the mechanism of G-CIMP formation. (ii) Identify a list of candidate genes that were hypermethylated and associated with poor survival in breast cancer brain metastasis patients. These genes may contribute to breast cancer brain metastasis formation and further analysis is required to assess the biological and functional relevance of the genes. Within the list, *EOMES*, *PHOX2B* and *DZIP1* were potential markers for BCBM and further investigation, such as migration and invasion assays, needs to be performed on breast cancer cell lines that show a correlation between DNA hypermethylation and loss of expression to find out if the reduced gene expression is correlated with increased cell motility and cell invasion. Other investigations may involve knocking out gene expression in unmethylated, expressing breast cancer cell lines.

6.3 Final conclusion

High-throughput technologies, such as NGS and methylation microarrays, are powerful technologies for the identification of disease-associated genetic/epigenetic changes without the knowledge of the exact location and function of the causative gene. Through

the use of such technologies in different CNS tumours, the results presented in this thesis have shown the validity of using high-throughput technologies to identify genetic/epigenetic markers. The results from Illumina Infinium HumanMethylation450K BeadChip have identified hypermethylated genes that are; (i) candidates for glioblastoma patients' poor survival, *DKK2*, *NR2F2* and *TFAP2A*, that require further investigation, (ii) a preliminary list of potential epigenetic markers for the early prediction of brain metastasis formation from breast tumours. WES was not able to identify the causative variants for non-*NF2* schwannoma, however, this technology was able to identify new SNVs in the *NF2* gene that, as far as we are aware, have not been previously reported in any SNP or SNV databases.

Chapter Seven: Appendices

7.1 Primer Sequences

7.1.1 CoBRA primers

Gene Name	Primers	Annealing Temperature
HOXA3	F 5' AGTATTTYGGGTTAGGGTGGTTTTATGTA 3' R 5' ATATACTTTCAAAACRCCCCAACCCTTAC 3' RN 5' TAACCCAAAAAATCCCTTCCCCTAACAAAT 3'	60 ^{O}C
RASSF1A	F 5' GAGYGYGTTTAGTTTYGTTTTYGGGTTTTAT 3' R 5' ATCRCCACAAAAATCRCACCACRTATACRTA 3' RN 5' TAACCACRACCAAAAACCAACTACCRTATA 3'	60 ^{O}C
FZD9	F 5' GAGGTTGTAGTGAGTYGATATYGYGTTAT 3' R 5' CTAAACCAAAACCRACRCAACCTCC 3' RN 5' ACTACCAACCCCCRAAACRAAACTC 3'	60 ^{O}C
SLIT2	F 5' GGGAGGTGGGATTGTTTAGATATTT 3' R 5' CAAAAACTCCTTAAACAACTTTAAATCCTAAAA 3' RN 5' ACTAAAACTTCCAACAACTACTAAAATACAAAAA 3'	54 ^{O}C
DKK2	F 5' TGAGTTTATAGGGTTTTGGGTTTTYGAATT 3' R 5' ACCACRCTCCRAATATTACAAAATACAATA 3' FN 5' ATTAGGAAGGGTTGAGGGAATATAATTTGT 3'	57 ^{O}C
EOMES	F 5' GGTATTGGGGATAGGATAAGAGTGA 3' R 5' CRCACTCAACAATCCTTTAACCATC 3' RN 5' CCRACATCTCAACCRAAAAATACRC 3'	58 ^{O}C
WDR69	F 5' GTATGTAGTTTAGGTTGTGGTTTAGGT 3' R 5' CCACRTAAAAATAACRAAACCRAACTAATAC 3' RN 5' CTCTTAAACTTCATTTTCTTACTCTCTTATC 3'	57 ^{O}C
MIR125B	F 5' GGTTATAGAAGTGTAAAGAYGAATAGYGT 3' R 5' ACCTTAACTTCAACCAAACAACTAATACAAAC 3' RN 5' CTAATACAAACRAAACCTCCRTTTACAAAAC 3'	58 ^{O}C
SOX1	F 5' GGATTTTGAGGGTTTAATGGGATTT 3' R 5' AACCTACTATAATCTTTTCTCCACTA 3' FN 5' GGTTTAATGGGATTTTGAGGGTTTT 3'	54 ^{O}C
PRDM13	F 5' AGTTTAGGTTGTTGGTTTTAGGGAA 3' R 5' TTCAAAACCRCTTTAATTCTCACTTCTA 3' FN 5' TGGTTTTAGGGAAGTTAGTTTAAGGAT 3'	55 ^{O}C
PHOX2B	F 5' AGGATTYGTAGGTAGAGGAATTGAG 3' R 5' AAACRCCAACAATAAAACCAACCRCTT 3' FN 5' AGGTTTTGGATGGTTTAGTTAAGTGGAA 3'	57 ^{O}C

Gene Name	Primers	Annealing Temperature
DZIP1	F 5' GTATTTTYGTYGGGGGYGTTTYGGTT 3' R 5' AACTCCCCACRCCTACCTTAAAAAC 3' RN 5' TTGTATAGGAGGAGTYGGGYGGTTT 3'	59 ^{O}C
ETS1	F 5' GGTTATAGAAGTGTAAAGAYGAATAGYGT 3' R 5' AACCTACTATAATCTTTTCTCCACTA 3' RN 5' CTAATACAAACRAAACCTCCRTTTACAAAAC 3'	58 ^{O}C

7.1.2 MSP/USP primers

Gene Name	Primers	Annealing Temperature
MGMT MSP	F 5' TTTCGACGTTCGTAGGTTTTCGC 3' R 5' GCACTCTTCCGAAAACGAAACG 3'	59 ^{O}C
MGMT USP	F 5' TTTGTGTTTTGATGTTTGTAGGTTTTTGT 3' R 5' AACTCCACACTCTTCCAAAAACAAAACA 3'	59 ^{O}C

7.1.3 Expression Primers

Gene Name	Primers	Annealing Temperature
DKK2	F 5' CTGCTGCCTGCTCCTACTG 3' R 5' CCATTATTGCAGCGGGTACT 3'	56 ^{O}C
GAPDH	F 5' GACCCCTTCAGTACCTCAACTACA 3' R 5' CTAAGCACTTGGTGGTGCAGGA 3'	56 ^{O}C
EOMES	F 5' CGGGCACCTATCAGTACAGC 3' R 5' CAGCACCACCTCTACGAACA 3'	59 ^{O}C
WDR69	F 5' TGTTTAGCCGTTTCCAAGGC 3' R 5' TGCTCTGTTCGTGAAGCTGT 3'	59 ^{O}C
SOX1	F 5' TGATGATGGAGACCGACCTG 3' R 5' CTCGGACATGACCTTCCACT 3'	58 ^{O}C
PRDM13	F 5' CTAGCGGTGCCAAAAGGC 3' R 5' CTTGCCCAGCTTGAAGGTAC 3'	58 ^{O}C
PHOX2B	F 5' TGATAGGGAGGTTGGACAGC 3' R 5' TATGGAGAAGGTGGCTGGAG 3'	58 ^{O}C
ETS1	F 5' GCTCACTCAGGATCTGCTCT 3' R 5' CCAAAAGGGGTAGCAAGGTC 3'	58 ^{O}C

7.1.4 Sequencing Primers

Gene Name	Primers	Annealing Temperature
TRAF7 Exon10/11	F 5' GACGGCAGGTGTGGGTG 3' R 5' CCACACACAAGACCCCTGAT 3'	59 ^{O}C
TRAF7 Exon13	F 5' ATCAGGGGTCTTGTGTGTGG 3' R 5' CAGGAATGAGTGAGGGAGCC 3'	59 ^{O}C
TRAF7 Exon17	F 5' CTGATGGCTGGCATGGAC 3' R 5' GAAGGTAGGAACAGGGCAGA 3'	59 ^{O}C
TRAF7 Exon18/19	F 5' CTGCCCTGTTCCTACCTTCG 3' R 5' GTGACACTGCCCTGGTGAC 3'	59 ^{O}C
TRAF7 Exon20	F 5' CGCCAGACCAGACCAAAGTC 3' R 5' CAGAGCCTGTCCACCTATGC 3'	59 ^{O}C
AKT1	F 5' CTGGCCCTAAGAAACAGCTC 3' R 5' CGCCACAGAGAAGTTGTTGA 3'	59 ^{O}C
SMO	F 5' GATGGGGACTCTGTGAGTGG 3' R 5' TCGTCTGCCTAGAGGGTCAC 3'	59 ^{O}C
KLF4	F 5' GTCCCGGGGATTTGTAGCTC 3' R 5' TCCCGGAGATCAAGGCGATA 3'	59 ^{O}C
NF2 Exon1/2/3	F 5' GGGCTAAAGGGCTCAGAGTG 3' R 5' AACCTCTCGAGCTTCCACCT 3'	59 ^{O}C
NF2 Exon5	F 5' TTACACGCCCTCTCTCTGTG 3' R 5' CCAACAATGAATGGGCCTCA 3'	59 ^{O}C
ARID1B	F 5' CAGCAACAGCAACATCCCAT 3' R 5' CAGCGCTGCCATAGTAATTA 3'	59 ^{O}C
SMARCE1	F 5' AGCTGACCTGTTGGCTGAAA 3' R 5' GAACGATGCTGCTCTGACTG 3'	59 ^{O}C
SMARCA4	F 5' TGTCACACGTGTCCATCGTT 3' R 5' GCGATGGTCTGGATGGTCTT 3'	59 ^{O}C
FAT1	F 5' CCGTCTTGGCCCATATACTT 3' R 5' GGCTTTAAGGGTAGATAGTGCG 3'	59 ^{O}C
RASSF4	F 5' AGGTTGGGAGAGGACAGGTT 3' R 5' CCCATCTGGAAGGGAGTACA 3'	59 ^{O}C
FAT3	F 5' ATAACCTTACTGTGCGGGCC 3' R 5' TTCCAAGACCACTGCCATCC 3'	58 ^{O}C

RASSF1A	F 5' GTTCTTGGTGGTGGATGACC 3' R 5' CCTTCAGGACAAAGCTCAGG 3'	59 °C
DCHS2	F 5' CGTTCTTCCAGTTGCGCTAC 3' R 5' CGGTACTCGTCCTGCTCAAA 3'	59 °C
MAPK12	F 5' AGTGGCTTTTACCGCCAGGA 3' R 5' GCTTCTTGATGGCCACCTTA 3'	59 °C
ARNT	F 5' GCAGGAAGTCAAGGGGATCT 3' R 5' ACACAGGGAACACACAGATG 3'	59 °C
MCM5	F 5' GGGTGGAATCAGAGACTTGC 3' R 5' TCACCAGAGCCATCTGTGTC 3'	59 °C
XBP1	F 5' TTCCTTACCAGCCTCCCTTT 3' R 5' GTGGCTGGATGAAAGCAGAT 3'	59 °C
ACO2	F 5' GGTCCTCACCCAAAGATGTG 3' R 5' CAAGTCCCCAAATGGAACAG 3'	59 °C
PRODH	F 5' CCACAGGATGCCTATGACAA 3' R 5' GGACACATGTGGCTGACAAG 3'	59 °C
LNX1	F 5' AGCAAAGGTGAGCCTTCTTG 3' R 5' GTGTGTCCACACGGAGTGTC 3'	59 °C
CYP3A4	F 5' GGAGTGTCTCACTCACTTTGATGC 3' R 5' TCTGGTGTTCTCAGGCACAG 3'	59 °C
CFHR5	F 5' CTCTGTATATGAAGCCCCTTGC 3' R 5' GGTGACCACCCAAATTGGTA 3'	59 °C
SAMD9	F 5' AGCAGAACATGCCTCAAGTG 3' R 5' TGGATCCCCTGGAATATCAC 3'	59 °C
NRD1	F 5' TGATGCAGCTGACCTACCTG 3' R 5' AATGGAAGCCCTAAGGCAGT 3'	59 °C
USP40	F 5' CAACAGAATCGACTCCCAAG 3' R 5' GGTTCCACCCTGATTTCTGA 3'	59 °C
PLEC	F 5' GTACCCTGCCTGCTGCCTTT 3' R 5' GATGCCTTCATTGGTGAGGT 3'	59 °C
COPZ2	F 5' GGGAGAACGAGCAAGAGGAC 3' R 5' CGAGTTCTGAGTTGGCTGCT 3'	59 °C
TNFAIP6	F 5' AGTGACAGCTGGAGGTTTCC 3' R 5' TGCAGTGGGATTATGAGAGG 3'	59 °C
CYBA	F 5' CACTCGGCCTGTTAGCTTTC 3' R 5' CCCCCTTCTCTACCTACTGTGG 3'	60 °C
SYTL2	F 5' GCTGTAGGGAAAACCACCAA 3' R 5' CTGGGAAGAAGCAAAGCAAC 3'	59 °C
ATRAID	F 5' GCCAGTATCCCCGAAAGAGG 3' R 5' CTTCAGCCCGACCTCAAACT 3'	59 °C

FASN	F 5' AGGCTGCCATTACCACACTT 3' R 5' AGATGGTGACTGTGTCCTTGG 3'	59 °C
ABCE1	F 5' TGTTCGCCTCCAACCTTAAG 3' R 5' TGCAGCCTTAGGAATCTGGT 3'	59 °C
SNX30	F 5' CCGTTAGACATTTGCTCCGC 3' R 5' CTTAACCCGAGGCCTTGCA 3'	59 °C
RNF19B	F 5' CGCTCCACATCGCTACATG 3' R 5' CTCCTCATCGTCGAACCCAG 3'	59 °C
LRPB1	F 5' TGTTTGTGCCAGGGACAATG 3' R 5' GTGCATCACTGTAAAAGATTCGG 3'	59 °C
ENOSF	F 5' GATCCCGACCAGTCCTGAC 3' R 5' CCAAGGCATGAAAAGCGAGT 3'	59 °C
SP7	F 5' CTAGAAGGGAGTGGTGGAGC 3' R 5' GCTCATCCGAACGAGTGAAC 3'	59 °C
RCSD1	F 5' AACTTGGAGATTTCAGGGCG 3' R 5' TGGCCTTTTCACCGTTCTTG 3'	59 °C
C2CD4A	F 5' AAGAAGCAGGGATGGACGAG 3' R 5' CTCCTTGTCCTCGTTCCCG 3'	59 °C
MESP2	F 5' GAGGAGAGTCTGCAGTGCC 3' R 5' CGACTGTATCTTGGGGCAGT 3'	59 °C
LINGO1	F 5' GACACTCATCCTGGACTCCA 3' R 5' TGAGACCAGGTGCTTTCGG 3'	59 °C
NRXN2	F 5' CTGAGGGAAGCCGGCATCT 3' R 5' CGTCGTCCAGGTAGAGCAG 3'	60 °C
GALNT12	F 5' ATCCGGGTCCTGGCCTCCA 3' R 5' CGGTCGCTGAGGTAGATGTTAAT 3'	59 °C
MLL2	F 5' CTGGTGGCCCCTTCCTTAAT 3' R 5' GGATTGCCACCTGTCCTAGA 3'	59 °C
RHOU	F 5' ATCGTTTCAAGGATGCTGGC 3' R 5' GACAGTGGCATCGAATCTCC 3'	58 °C
NQO2	F 5' CTATGCACACCAGGAACCCA 3' R 5' ACTTCTGGGATACTGCTGGTG 3'	59 °C
EPPK1	F 5' AGACTTCTGGCCTTCACCTC 3' R 5' GGGTCTCCTGGGTGATGATG 3'	59 °C
RETN	F 5' CTCCCAGCTCAGAGTCCAC 3' R 5' CAGGTTTATTTCCAGCTCCCC 3'	58 °C

265

7.2 Clinical information of the 35 glioblastoma cases used in chapter three

This table shows clinical features of 35 glioblastoma cases (STS=19 and LTS=16) used in the methylation array. Parameters include; gender of cases (male or female), age at diagnosis (range and median age), survival in years and median survival, case status (dead or alive), percentage score of KPS, radiotherapy and chemotherapy (yes or no), and IDH1 mutation status (mutated or wild-type). Number between brackets represents the percentage of total number of samples.

		STS Glioblastoma	LTS Glioblastoma
Gender	Male	12 (63%)	9 (56%)
	Female	7 (37%)	7 (44%)
Age at diagnosis	Range	38-82	33-72
	Median Age	63	52.5
Survival	Range (years)	0.54-0.97	3.06-11.11
	Median Survival	7 months	5 years
Status	Dead	19 (100%)	8 (50%)
	Alive	0 (0%)	8 (50%)
KPS %	Range	70-100	80-100
Radiotherapy	Yes	16 (84%)	16 (100%)
	No	3 (16%)	0 (0%)
Chemotherapy	Yes	10 (53%)	16 (100%)
	No	9 (47%)	0 (0%)
IDH1 mutation	Mutated	0 (0%)	5 (31%)
	Wild-Type	19 (100%)	11 (69%)

7.3 CIMP+ CpG loci

List of CIMP+ CpG loci with adjusted P value <0.05. The list contains gene names, adjusted P value, chromosomal location, accession number and relation to CpG island.

Gene NAME	Adj. Pval	CHR	ACCESSION Number	RELATION TO CpG
COL11A2	2.96501E-25	6	NM_001163771	N_Shore
KIAA0495	1.35115E-23	1	NM_207306	Island
ADAMTSL4	1.13621E-22	1	NM_019032	Island
NPM2	1.27923E-20	8	NM_182795	Island
GNMT	2.99914E-20	6	NM_018960	Island
TCEA3	1.31214E-18	1	NM_003196	Island
KCNB1	1.35837E-18	20	NM_004975	Island
CBLN3	1.36898E-18	14	NM_001039771	Island
NGB	3.212E-18	14	NM_021257	Island
CFTR	8.70933E-18	7	NM_000492	Open sea
MACROD1	1.01481E-17	11	NM_014067	Island
CBLN3	1.12008E-17	14	NM_001039771	Island
CLSTN1	1.27718E-17	1	NM_014944	Island
FRZB	1.31641E-17	2	NM_001463	Island
CMYA5	5.15879E-17	5	NM_153610	Open sea
TCEA3	5.41486E-17	1	NM_003196	Island
CBLN3	5.88516E-17	14	NM_001039771	S_Shore
CMYA5	1.6431E-16	5	NM_153610	Open sea
RHBDF2	1.65905E-16	17	NM_024599	Island
NTSR2	1.69212E-16	2	NM_012344	Island
TPPP3	1.69791E-16	16	NM_015964	Island
KIAA0495	2.10287E-16	1	NM_207306	Island
PRR15	3.3098E-16	7	NM_175887	Island
LRRC3B	8.27429E-16	3	NM_052953	N_Shore
GIPC2	8.86418E-16	1	NM_017655	Island

Gene	p-value	Chr	Accession	Region
DLGAP3	9.7562E-16	1	NM_001080418	Island
KIAA0495	1.24367E-15	1	NM_207306	Island
CNGA3	1.25263E-15	2	NM_001298	S_Shore
LIMS2	1.26021E-15	2	NM_001136037	Island
COL9A2	3.55036E-15	1	NM_001852	Island
HLF	3.87448E-15	17	NM_002126	S_Shore
FCHSD1	4.33341E-15	5	NM_033449	Island
TPPP3	4.66527E-15	16	NM_015964	Island
CBLN3	5.08916E-15	14	NM_001039771	Island
SLC22A15	5.14119E-15	1	NM_018420	Island
RFX8	6.11908E-15	2	NM_001145664	Island
MAGI2	6.20939E-15	7	NM_012301	Island
ACTA1	7.85378E-15	1	NM_001100	Island
COL11A2	7.91063E-15	6	NM_001163771	N_Shore
PNKD	8.01914E-15	2	NM_015488	Island
FCHSD1	1.19621E-14	5	NM_033449	Island
CLDN10	1.19698E-14	13	NM_006984	Island
ALS2CL	1.2111E-14	3	NM_147129	Island
SPTBN4	1.35016E-14	19	NM_020971	Island
TCIRG1	1.39217E-14	11	NM_006019	Island
AKR7A3	1.472E-14	1	NM_012067	Island
MCFD2	1.49614E-14	2	NM_001171508	S_Shore
MACROD1	2.03981E-14	11	NM_014067	Island
TTC12	2.0444E-14	11	NM_017868	N_Shore
CLSTN1	2.22592E-14	1	NM_014944	Island
SPEG	2.94825E-14	2	NM_005876	Island
NR2E1	3.64853E-14	6	NM_003269	Island
DNAJA4	4.33928E-14	15	NM_018602	Island
TUBA4B	4.99997E-14	2	NR_003063	Island

FCHSD1	5.1707E-14	5	NM_033449	Island
RHBDF2	5.28622E-14	17	NM_024599	S_Shore
CBLN3	5.28941E-14	14	NM_001039771	Island
PTPRU	5.81698E-14	1	NM_133178	Island
RARRES2	6.62572E-14	7	NM_002889	Island
GJA4	6.70831E-14	1	NM_002060	Island
PLD5	7.51899E-14	1	NM_152666	N_Shore
FGFRL1	7.56347E-14	4	NM_001004358	Island
G0S2	7.9631E-14	1	NM_015714	Island
LGALS3	8.82577E-14	14	NM_002306	Island
RARRES1	1.00482E-13	3	NM_206963	Island
ARL4A	1.03154E-13	7	NM_212460	Island
B3GNT7	1.03285E-13	2	NM_145236	Island
LGALS3	1.05926E-13	14	NM_002306	Island
TTC12	1.05948E-13	11	NM_017868	N_Shore
KCNB1	1.34636E-13	20	NM_004975	Island
SLC6A11	1.62366E-13	3	NM_014229	Island
LIMS2	3.19952E-13	2	NM_001136037	Island
EBF4	3.34346E-13	20	NM_001110514	Island
LRAT	3.41055E-13	4	NM_004744	Island
VSX1	3.43971E-13	20	NM_014588	Island
KCNB1	3.80101E-13	20	NM_004975	Island
ARL4A	4.69923E-13	7	NM_212460	Island
FABP5	4.76397E-13	8	NM_001444	Island
FAM132A	5.18265E-13	1	NM_001014980	Island
ANGPT4	5.84399E-13	20	NM_015985	N_Shore
THRB	6.17996E-13	3	NM_001128177	Island
CCNI2	6.23242E-13	5	NM_001039780	Island
CLSTN1	6.48294E-13	1	NM_014944	Island

Gene	p-value	Chr	Accession	Region
ACTA1	6.80792E-13	1	NM_001100	Island
ZDHHC8P	7.91967E-13	22	NR_003950	Open sea
TNFAIP8L3	8.12941E-13	15	NM_207381	Island
CBLN3	8.18742E-13	14	NM_001039771	Island
CLSTN1	9.97038E-13	1	NM_014944	Island
SRRM3	1.06839E-12	7	NM_001110199	Island
P2RY2	1.23083E-12	11	NM_176071	Island
MYRIP	1.23898E-12	3	NM_015460	Island
ANGPT4	1.25895E-12	20	NM_015985	Island
DNAJA4	1.32375E-12	15	NM_018602	Island
CMYA5	1.41724E-12	5	NM_153610	Open sea
COL18A1	1.41728E-12	21	NM_130445	Island
HIST1H4F	1.43376E-12	6	NM_003540	N_Shore
HSPA2	1.4873E-12	14	NM_021979	Island
G0S2	1.48934E-12	1	NM_015714	Island
KCNQ4	1.99214E-12	1	NM_004700	Island
GAL3ST2	2.98165E-12	2	NM_022134	Island
TMEM200B	3.20163E-12	1	NM_001003682	Island
GAL3ST2	3.6068E-12	2	NM_022134	Island
SLC6A11	3.67432E-12	3	NM_014229	Island
AKR7A3	4.14841E-12	1	NM_012067	Island
WWTR1	5.11334E-12	3	NM_001168278	Island
B3GNT7	5.30268E-12	2	NM_145236	Island
GLUL	6.89814E-12	1	NM_001033044	Island
NTN1	7.08416E-12	17	NM_004822	Island
TCEA3	8.80596E-12	1	NM_003196	Island
LGALS3	8.80784E-12	14	NM_002306	Island
KIAA0495	9.2286E-12	1	NM_207306	Island
KCNB1	9.58902E-12	20	NM_004975	Island

HHIPL1	9.68948E-12	14	NM_001127258	Island
HSD11B2	1.0161E-11	16	NM_000196	Island
ACTG1	1.09732E-11	17	NM_001614	Island
ACTA1	1.43315E-11	1	NM_001100	Island
CORO6	1.50925E-11	17	NM_032854	Island
FAM19A1	1.51584E-11	3	NM_213609	S_Shore
TMEM143	1.68337E-11	19	NM_018273	Island
ONECUT1	1.88424E-11	15	NM_004498	N_Shore
CNTN2	1.88947E-11	1	NM_005076	Open sea
VWDE	1.9411E-11	7	NM_001135924	Island
EFEMP1	2.1718E-11	2	NM_004105	Island
PYCARD	2.39917E-11	16	NM_145182	Island
GRIN1	2.43317E-11	9	NM_000832	Island
BMP8A	2.49937E-11	1	NM_181809	Island
TTC12	2.50293E-11	11	NM_017868	N_Shore
FCHSD1	2.56814E-11	5	NM_033449	Island
SRRM3	2.81192E-11	7	NM_001110199	Island
BRUNOL5	3.05977E-11	19	NM_021938	Island
TTC12	3.09297E-11	11	NM_017868	N_Shore
SRRM3	3.87383E-11	7	NM_001110199	Island
HCRT	3.99713E-11	17	NM_001524	Island
SOST	4.09479E-11	17	NM_025237	Island
TPPP3	4.25209E-11	16	NM_016140	Island
RARRES2	4.37675E-11	7	NM_002889	Island
VILL	4.3873E-11	3	NM_015873	S_Shore
PNKD	4.87514E-11	2	NM_015488	Island
RBP7	5.0358E-11	1	NM_052960	Island
PTPRU	5.11486E-11	1	NM_133178	Island
HSD11B2	5.58111E-11	16	NM_000196	Island

GNMT	7.07328E-11	6	NM_018960	Island
RBP1	8.20699E-11	3	NM_001130992	S_Shore
ZNF704	8.6219E-11	8	NM_001033723	Open sea
KCNJ3	8.92629E-11	2	NM_002239	Island
ZNF704	8.98827E-11	8	NM_001033723	Open sea
APBA2	9.18522E-11	15	NM_001130414	Island
CYP11A1	9.50842E-11	15	NM_001099773	Island
OPCML	9.62434E-11	11	NM_001012393	Island
CORO6	1.04147E-10	17	NM_032854	Island
AKR7A3	1.06038E-10	1	NM_012067	Island
ZAR1	1.25964E-10	4	NM_175619	Island
SYCP2L	1.35275E-10	6	NM_001040274	N_Shore
NTN1	1.38251E-10	17	NM_004822	Island
PNKD	1.43386E-10	2	NM_015488	Island
HSD11B2	1.48671E-10	16	NM_000196	Island
PARVB	1.52295E-10	22	NM_001003828	Island
XKR8	1.55847E-10	1	NM_018053	Island
ACTG1	1.5637E-10	17	NM_001614	Island
ONECUT2	1.58326E-10	18	NM_004852	S_Shore
AKR7A3	1.67551E-10	1	NM_012067	Island
TTC12	1.74635E-10	11	NM_017868	Island
NRXN1	1.75167E-10	2	NM_004801	Island
HIST1H4L	1.81862E-10	6	NM_003546	S_Shore
TTC12	1.82673E-10	11	NM_017868	N_Shore
VWDE	1.9201E-10	7	NM_001135924	Island
CPNE8	1.9218E-10	12	NM_153634	N_Shore
CRYAB	1.97386E-10	11	NM_001885	Open sea
THRB	1.97585E-10	3	NM_001128177	Island
KY	2.05855E-10	3	NM_178554	Island

RASGRP2	2.19512E-10	11	NM_001098670	Island
NEUROD1	2.25262E-10	2	NM_002500	N_Shore
KIAA0495	2.26462E-10	1	NM_207306	Island
DLGAP3	2.26749E-10	1	NM_001080418	Island
FABP5	2.3945E-10	8	NM_001444	Island
IQSEC1	2.81205E-10	3	NM_001134382	Island
VILL	2.99136E-10	3	NM_015873	Island
KIAA0495	3.18681E-10	1	NM_207306	Island
SOST	3.263E-10	17	NM_025237	Island
GRIN1	3.3099E-10	9	NM_000832	Island
COL24A1	3.53914E-10	1	NM_152890	Island
RAB38	3.67308E-10	11	NM_022337	S_Shore
PLLP	3.82989E-10	16	NM_015993	Island
TMEM200B	4.27041E-10	1	NM_001003682	Island
RGS22	4.49194E-10	8	NM_015668	Island
DLGAP3	4.56483E-10	1	NM_001080418	Island
GPR12	5.1627E-10	13	NM_005288	Island
ZAR1	5.65035E-10	4	NM_175619	Island
SPEG	6.00618E-10	2	NM_005876	Island
MAST1	6.28323E-10	19	NM_014975	Island
SIM2	6.83717E-10	21	NM_005069	Island
MBP	7.14126E-10	18	NM_001025101	Island
GNMT	7.48248E-10	6	NM_018960	Island
CPNE8	7.60311E-10	12	NM_153634	Island
ESRP2	7.73806E-10	16	NM_024939	Island
TNFAIP8L3	8.01785E-10	15	NM_207381	Island
TRH	9.12691E-10	3	NM_007117	Island
PYCARD	9.66985E-10	16	NM_145182	Island
TMED7-	9.9535E-10	5	NM_001164468	Island

ALS2CL	1.00115E-09	3	NM_147129	Island
PDE4A	1.08222E-09	19	NM_001111307	N_Shore
NIPAL4	1.09779E-09	5	NM_001099287	Island
BRUNOL4	1.11E-09	18	NM_001025088	Island
EBAG9	1.17677E-09	8	NM_004215	S_Shore
BRUNOL4	1.2267E-09	18	NM_001025088	Island
TRANK1	1.35257E-09	3	NM_014831	Island
G0S2	1.37846E-09	1	NM_015714	Island
SCN9A	1.38213E-09	2	NM_002977	S_Shore
GLB1L3	1.40565E-09	11	NM_001080407	Island
PPP1R14A	1.54065E-09	19	NM_033256	Island
COL24A1	1.62838E-09	1	NM_152890	Island
MAGI2	1.64031E-09	7	NM_012301	Island
RHOD	1.72418E-09	11	NM_014578	Island
FRZB	1.78851E-09	2	NM_001463	N_Shore
ME1	1.85672E-09	6	NM_002395	Island
CACNG8	2.0007E-09	19	NM_031895	Island
GRASP	2.02091E-09	12	NM_181711	Island
PRR15	2.04571E-09	7	NM_175887	Island
MBP	2.14649E-09	18	NM_001025100	Island
ACTA1	2.32912E-09	1	NM_001100	Island
FAM132A	2.37147E-09	1	NM_001014980	Island
PPP1R14A	2.45059E-09	19	NM_033256	Island
EMILIN2	2.46475E-09	18	NM_032048	Island
NEUROD1	2.46552E-09	2	NM_002500	Open sea
GRIN1	2.60576E-09	9	NM_000832	Island
ERBB2	2.61842E-09	17	NM_001005862	N_Shore
SYCP2L	2.99823E-09	6	NM_001040274	N_Shore
MYO18A	3.00675E-09	17	NM_078471	Island

ZNF704	3.26079E-09	8	NM_001033723	Open sea
GNMT	3.47673E-09	6	NM_018960	Island
OPCML	3.5346E-09	11	NM_001012393	Open sea
HLF	3.53973E-09	17	NM_002126	S_Shore
RBM24	3.54088E-09	6	NM_153020	S_Shore
AOX1	3.54543E-09	2	NM_001159	Island
THRB	3.74574E-09	3	NM_001128177	Island
COL9A3	3.76526E-09	20	NM_001853	Island
SIAH1	3.84749E-09	16	NM_001006610	Island
ARAP1	3.88609E-09	11	NM_001040118	Island
MAPT	3.95153E-09	17	NM_001123067	Island
LBXCOR1	3.95717E-09	15	NM_001031807	Island
FABP5	3.98792E-09	8	NM_001444	Island
LINGO3	4.05388E-09	19	NM_001101391	S_Shore
NEUROD1	4.27673E-09	2	NM_002500	Open sea
TOX2	4.37434E-09	20	NM_001098796	Island
KCNJ3	4.38721E-09	2	NM_002239	Island
GJA4	4.65156E-09	1	NM_002060	Island
GNMT	5.10955E-09	6	NM_018960	Island
CPNE8	5.12631E-09	12	NM_153634	N_Shore
MAPT	5.13135E-09	17	NM_001123067	Island
CPNE8	5.17443E-09	12	NM_153634	N_Shore
NID2	5.39083E-09	14	NM_007361	Island
CPNE8	5.41593E-09	12	NM_153634	Island
SULT1A1	5.72378E-09	16	NM_177536	Island
FAM19A1	6.9806E-09	3	NM_213609	S_Shore
COL24A1	7.70654E-09	1	NM_152890	Island
FAM59B	7.80868E-09	2	NM_001168241	Island
ZDHHC8P	8.08272E-09	22	NR_003950	Open sea

HAPLN3	8.75004E-09	15	NM_178232	Island
SST	8.94683E-09	3	NM_001048	Island
PPP1R14A	1.10818E-08	19	NM_033256	Island
NR4A1	1.17571E-08	12	NM_002135	Island
ADAM12	1.17842E-08	10	NM_021641	Island
NSD1	1.18483E-08	5	NM_172349	Island
IGSF21	1.20974E-08	1	NM_032880	Island
KCNIP3	1.28865E-08	2	NM_001034914	Island
EBAG9	1.31686E-08	8	NM_004215	Island
KCNIP3	1.31806E-08	2	NM_001034914	Island
SIM2	1.33973E-08	21	NM_005069	Island
DPYSL5	1.34012E-08	2	NM_020134	Island
ERBB2	1.34109E-08	17	NM_001005862	N_Shore
LRRC3B	1.3423E-08	3	NM_052953	Island
CACNG8	1.41268E-08	19	NM_031895	N_Shore
THRB	1.43165E-08	3	NM_001128177	Island
CPNE8	1.43917E-08	12	NM_153634	Island
TTC12	1.46281E-08	11	NM_017868	Island
SPEG	1.70576E-08	2	NM_005876	Island
PSD3	1.73387E-08	8	NM_015310	Island
CYP26B1	1.73812E-08	2	NM_019885	Island
LECT1	1.74085E-08	13	NM_001011705	Island
GPR153	1.8311E-08	1	NM_207370	Island
THRB	1.92323E-08	3	NM_001128177	Island
HLA-L	1.98493E-08	6	NR_027822	Island
HPDL	2.06905E-08	1	NM_032756	Island
MIR155HG	2.09788E-08	21	NR_001458	Island
GSX2	2.18268E-08	4	NM_133267	N_Shore
CLSTN1	2.18571E-08	1	NM_014944	S_Shore

MIR375	2.22701E-08	2	NR_029867	Island
ANGPT4	2.38933E-08	20	NM_015985	Island
SPIB	2.71796E-08	19	NM_003121	Island
CALHM2	2.86764E-08	10	NM_015916	Island
ONECUT2	2.95832E-08	18	NM_004852	S_Shore
VSX1	2.96362E-08	20	NM_014588	Island
SPINT2	2.96843E-08	19	NM_001166103	Island
UBE2E2	3.03837E-08	3	NM_152653	Island
SPEG	3.25285E-08	2	NM_005876	Island
SPTBN4	3.25679E-08	19	NM_020971	Island
CBLN3	3.28201E-08	14	NM_001039771	Island
MT1M	3.4623E-08	16	NM_176870	Island
ERBB2	3.88374E-08	17	NM_001005862	N_Shore
NGB	4.08781E-08	14	NM_021257	Island
PRR15	4.10773E-08	7	NM_175887	N_Shore
GRASP	4.1098E-08	12	NM_181711	Island
PDE4A	4.11383E-08	19	NM_006202	Open sea
SHANK1	4.29561E-08	19	NM_016148	Island
BNC1	4.45325E-08	15	NM_001717	Island
ZDHHC8P	4.62192E-08	22	NR_003950	Open sea
MDK	4.6908E-08	11	NM_001012334	N_Shore
MBP	5.11113E-08	18	NM_001025100	Island
CPLX1	5.11856E-08	4	NM_006651	Island
HPDL	5.70235E-08	1	NM_032756	Island
BRUNOL4	7.06902E-08	18	NM_001025088	Island
KY	7.33247E-08	3	NM_178554	Island
RHOD	7.37455E-08	11	NM_014578	Island
ARAP1	7.70264E-08	11	NM_001040118	Island
MIR155HG	7.71285E-08	21	NR_001458	N_Shore

SST	7.71638E-08	3	NM_001048	S_Shore
TOX2	7.76271E-08	20	NM_001098796	Island
EHBP1L1	7.85964E-08	11	NM_001099409	Island
ERBB2	7.8731E-08	17	NM_001005862	N_Shore
PDE4A	8.00506E-08	19	NM_001111308	N_Shore
NEUROD1	8.20579E-08	2	NM_002500	Open sea
GNMT	8.27018E-08	6	NM_018960	Island
LTBP3	8.40539E-08	11	NM_001164266	Island
RNF135	8.41256E-08	17	NM_197939	N_Shore
RHOD	8.49824E-08	11	NM_014578	Island
CNGA3	8.73701E-08	2	NM_001298	Island
SPINT2	8.86707E-08	19	NM_001166103	Island
CLDN5	8.90237E-08	22	NM_001130861	Island
GPR153	8.90836E-08	1	NM_207370	Island
FABP5	8.91126E-08	8	NM_001444	Island
KY	9.83017E-08	3	NM_178554	Island
RAB32	1.0256E-07	6	NM_006834	Island
COL9A2	1.07076E-07	1	NM_001852	S_Shore
ZBED3	1.13567E-07	5	NM_032367	Island
TMEM143	1.22034E-07	19	NM_018273	Island
NR4A1	1.23052E-07	12	NM_002135	Island
SOST	1.23681E-07	17	NM_025237	Island
ISLR2	1.2865E-07	15	NM_001130136	Island
MAGI2	1.31984E-07	7	NM_012301	Island
STL	1.32586E-07	6	NR_026876	S_Shore
MAGI2	1.42393E-07	7	NM_012301	Island
PARVB	1.48476E-07	22	NM_001003828	Island
TMED7-	1.51051E-07	5	NM_001164468	S_Shore
COL21A1	1.52715E-07	6	NM_030820	Island

CLSTN1	1.52732E-07	1	NM_014944	Island
VSIG10L	1.54101E-07	19	NM_001163922	Island
GAL3ST2	1.56171E-07	2	NM_022134	Island
GNMT	1.61567E-07	6	NM_018960	Island
CACNG8	1.61833E-07	19	NM_031895	Island
ZFHX3	1.63142E-07	16	NM_001164766	Island
FAM132A	1.68628E-07	1	NM_001014980	Island
PHLDB1	1.74419E-07	11	NM_001144758	Island
STL	1.74601E-07	6	NR_026876	S_Shore
EBAG9	1.99554E-07	8	NM_004215	S_Shore
LPCAT2	2.17308E-07	16	NM_017839	Island
SST	2.20068E-07	3	NM_001048	Island
NCRNA00085	2.20704E-07	19	NR_024330	Island
ESRP2	2.52006E-07	16	NM_024939	Island
NUDT16	2.70902E-07	3	NM_152395	Island
ISLR2	2.72085E-07	15	NM_001130136	Island
GRIN2C	2.72178E-07	17	NM_000835	Island
PGCP	2.74344E-07	8	NM_016134	Island
RBM24	2.76276E-07	6	NM_001143942	Island
ZBED3	2.79978E-07	5	NM_032367	Island
DNAJA4	2.89757E-07	15	NM_018602	Island
ACTG1	2.95119E-07	17	NM_001614	Island
NPHS2	2.96568E-07	1	NM_014625	Island
UNC13A	2.96933E-07	19	NM_001080421	Island
MAGI2	2.97706E-07	7	NM_012301	Island
RUNX1	3.00344E-07	21	NM_001754	Island
BRUNOL4	3.10917E-07	18	NM_001025088	Island
PXDN	3.16151E-07	2	NM_012293	N_Shore
MEGF11	3.3591E-07	15	NM_032445	Island

RPP25	3.40642E-07	15	NM_017793	N_Shore
NRXN1	3.42477E-07	2	NM_004801	S_Shore
DRD4	3.47937E-07	11	NM_000797	Island
ONECUT1	3.50417E-07	15	NM_004498	N_Shore
MAL	3.5258E-07	2	NM_022440	Island
PARVB	3.63603E-07	22	NM_001003828	Island
PLLP	3.76368E-07	16	NM_015993	Island
DERL3	3.76956E-07	22	NM_001135751	Island
CTHRC1	3.89523E-07	8	NM_138455	Island
RADIL	3.90003E-07	7	NM_018059	N_Shore
IQSEC1	4.03095E-07	3	NM_001134382	Island
PDLIM4	4.44369E-07	5	NM_001131027	Island
ATP5G2	4.48553E-07	12	NM_005176	N_Shore
TMEM200B	4.57192E-07	1	NM_001003682	Island
KCNIP3	4.57794E-07	2	NM_001034914	Island
CHPF	4.91546E-07	2	NM_024536	Island
ZAR1	5.52091E-07	4	NM_175619	Island
ANKS1B	5.86543E-07	12	NM_020140	S_Shore
EMILIN2	5.9624E-07	18	NM_032048	Island
RAPGEFL1	6.01712E-07	17	NM_016339	Island
GLB1L3	6.16649E-07	11	NM_001080407	Island
RFX8	6.28833E-07	2	NM_001145664	Island
CACNG8	6.45787E-07	19	NM_031895	Island
RAB32	6.50709E-07	6	NM_006834	Island
TMBIM1	6.92699E-07	2	NM_022152	Island
MEI1	6.94679E-07	22	NM_152513	Island
NSD1	7.06567E-07	5	NM_172349	Island
LPCAT2	7.20732E-07	16	NM_017839	Island
HTR6	7.24957E-07	1	NM_000871	Island

Gene	p-value	Chr	Accession	Region
PDLIM4	7.8451E-07	5	NM_001131027	Island
CPLX1	7.94866E-07	4	NM_006651	Island
KIAA1614	8.45867E-07	1	NM_020950	Island
MEGF11	8.63642E-07	15	NM_032445	Island
VAX2	8.6517E-07	2	NM_012476	Island
MEI1	9.33021E-07	22	NM_152513	Island
ADAM12	9.38538E-07	10	NM_021641	Island
GLUL	9.81009E-07	1	NM_001033044	Island
CPNE8	9.95754E-07	12	NM_153634	Island
PDE4A	1.09134E-06	19	NM_001111307	N_Shore
PNMT	1.09471E-06	17	NM_002686	Island
LTBP3	1.096E-06	11	NM_001164266	Island
RFX8	1.09681E-06	2	NM_001145664	Island
COL12A1	1.12587E-06	6	NM_004370	Island
SLC6A11	1.13379E-06	3	NM_014229	Island
B3GNT7	1.13871E-06	2	NM_145236	Island
RADIL	1.16017E-06	7	NM_018059	N_Shore
TMEM144	1.16888E-06	4	NM_018342	Island
RGNEF	1.20935E-06	5	NM_001080479	Island
GNMT	1.2825E-06	6	NM_018960	Island
LPCAT2	1.297E-06	16	NM_017839	Island
SGK1	1.30842E-06	6	NM_001143676	S_Shore
CENPV	1.32096E-06	17	NM_181716	Island
GNMT	1.34815E-06	6	NM_018960	Island
RAPGEFL1	1.35888E-06	17	NM_016339	Island
SNX31	1.44133E-06	8	NM_152628	Island
SIAH1	1.4494E-06	16	NM_001006610	Island
IER3	1.47747E-06	6	NM_003897	Island
CYP11A1	1.48575E-06	15	NM_000781	Island

CPNE8	1.51529E-06	12	NM_153634	N_Shore
GDNF	1.54643E-06	5	NM_000514	Island
NRG2	1.60925E-06	5	NM_013982	Island
CHPF	1.64908E-06	2	NM_024536	Island
CXCR4	1.67849E-06	2	NM_003467	N_Shore
NRG2	1.69043E-06	5	NM_013982	Island
ZBED3	1.75529E-06	5	NM_032367	Island
EFCAB4B	1.77213E-06	12	NM_001144959	Island
ANXA2	1.77852E-06	15	NM_001002857	Island
SPINT2	1.78155E-06	19	NM_001166103	Island
EBF4	1.81619E-06	20	NM_001110514	Island
VRK2	1.93206E-06	2	NM_006296	Island
BRUNOL4	1.94464E-06	18	NM_001025088	Island
DKK3	1.96118E-06	11	NM_001018057	S_Shore
CLDN10	1.9618E-06	13	NM_006984	Island
MEGF11	2.0221E-06	15	NM_032445	Island
ACRBP	2.02385E-06	12	NM_032489	Island
SRRM3	2.03009E-06	7	NM_001110199	Island
KCNB1	2.03625E-06	20	NM_004975	Island
TCEA3	2.08287E-06	1	NM_003196	Island
COL11A2	2.09805E-06	6	NM_001163771	Island
GNMT	2.26979E-06	6	NM_018960	Island
HPCA	2.36519E-06	1	NM_002143	N_Shore
BRUNOL4	2.40983E-06	18	NM_001025089	Island
MEI1	2.41711E-06	22	NM_152513	N_Shore
CLDN5	2.47482E-06	22	NM_001130861	Island
KY	2.65846E-06	3	NM_178554	Island
SGK1	2.86917E-06	6	NM_001143676	S_Shore
RHOD	2.8882E-06	11	NM_014578	Island

BMP8A	2.92664E-06	1	NM_181809	Island
CENPV	3.05283E-06	17	NM_181716	Island
PIM3	3.10189E-06	22	NM_001001852	Island
RAP1GAP	3.21204E-06	1	NM_002885	Island
UBE2E2	3.3199E-06	3	NM_152653	Island
HIST1H4F	3.35752E-06	6	NM_003540	N_Shore
ANXA2	3.44406E-06	15	NM_001002857	Island
MYO18A	3.55392E-06	17	NM_078471	Island
CDH3	3.59693E-06	16	NM_001793	Island
DRD4	3.63809E-06	11	NM_000797	Island
IER3	3.6949E-06	6	NM_003897	Island
NRG2	3.72201E-06	5	NM_013982	Island
TOX2	3.79777E-06	20	NM_001098796	Island
PDLIM4	3.90265E-06	5	NM_001131027	Island
RUNX1	4.004E-06	21	NM_001754	Island
ACTG1	4.0321E-06	17	NM_001614	Island
CACNG8	4.1338E-06	19	NM_031895	Island
DLG1	4.28734E-06	3	NM_004087	Island
RASGRP2	4.28924E-06	11	NM_153819	Island
COL11A2	4.49895E-06	6	NM_001163771	N_Shore
CHST6	4.53616E-06	16	NM_021615	Island
GNMT	4.62465E-06	6	NM_018960	Island
CYP26B1	4.62579E-06	2	NM_019885	Island
TTC12	4.84256E-06	11	NM_017868	N_Shore
ATP5G2	4.92458E-06	12	NM_005176	Island
ACRBP	5.25092E-06	12	NM_032489	Island
RFX4	5.32549E-06	12	NM_213594	Island
ISLR2	5.46393E-06	15	NM_001130136	Island
ZEB1	5.47581E-06	10	NM_030751	Island

PLLP	5.54378E-06	16	NM_015993	Island
FLJ23834	5.60889E-06	7	NM_152750	Island
BRUNOL4	5.61052E-06	18	NM_001025088	Island
NFKBIZ	5.68624E-06	3	NM_001005474	Island
TMEM22	5.69242E-06	3	NM_001097600	Island
LPCAT2	5.7984E-06	16	NM_017839	Island
LYPD1	5.81642E-06	2	NM_001077427	N_Shore
GPR12	6.34359E-06	13	NM_005288	Island
GDNF	6.37686E-06	5	NM_000514	Island
EFCAB4B	6.40282E-06	12	NM_001144959	Island
RAB38	6.574E-06	11	NM_022337	S_Shore
RAB33B	6.68713E-06	4	NM_031296	S_Shore
CPLX1	6.70691E-06	4	NM_006651	Island
HLA-J	6.86711E-06	6	NR_024240	Island
NIPAL4	7.09052E-06	5	NM_001099287	Island
LYN	7.22233E-06	8	NM_002350	N_Shore
MYO15B	7.37482E-06	17	NR_003587	Island
MACROD1	7.39016E-06	11	NM_014067	Island
NSUN7	7.45218E-06	4	NM_024677	Island
LINGO3	7.45908E-06	19	NM_001101391	Island
FRZB	7.55665E-06	2	NM_001463	N_Shore
KIF5C	7.56302E-06	2	NM_004522	S_Shore
SHANK1	7.64342E-06	19	NM_016148	S_Shore
IQSEC1	8.19722E-06	3	NM_001134382	Island
CLDN5	8.63791E-06	22	NM_001130861	Island
RASGRP2	8.64037E-06	11	NM_153819	Island
CFTR	8.81848E-06	7	NM_000492	Open sea
LYPD1	8.85836E-06	2	NM_001077427	N_Shore
GALNTL1	8.93099E-06	14	NM_001168368	Island

FABP5	9.1553E-06	8	NM_001444	S_Shore
SLC16A11	9.34485E-06	17	NM_153357	Island
DLGAP3	9.46213E-06	1	NM_001080418	Island
FLJ22536	9.50188E-06	6	NR_015410	S_Shore
UBE2E2	9.53926E-06	3	NM_152653	Island
KY	9.68867E-06	3	NM_178554	Island
TPPP3	9.90913E-06	16	NM_015964	Island
KBTBD11	1.01811E-05	8	NM_014867	Island
CNTD2	1.03826E-05	19	NM_024877	Island
ISLR2	1.0464E-05	15	NM_001130136	Island
GSX2	1.05811E-05	4	NM_133267	N_Shore
KLHL1	1.0705E-05	13	NM_020866	S_Shore
PDLIM4	1.08035E-05	5	NM_001131027	Island
EHBP1L1	1.10831E-05	11	NM_001099409	Island
EBF4	1.12927E-05	20	NM_001110514	Island
KCNAB3	1.13071E-05	17	NM_004732	Island
FGFRL1	1.14028E-05	4	NM_001004358	Island
MYO15B	1.14123E-05	17	NR_003587	Island
PYCARD	1.20804E-05	16	NM_013258	Island
NCRNA00092	1.29383E-05	9	NR_024129	Island
IER3	1.29585E-05	6	NM_003897	Island
LRAT	1.32656E-05	4	NM_004744	Island
KIAA1614	1.36678E-05	1	NM_020950	Island
MACROD1	1.3695E-05	11	NM_014067	Open sea
HPDL	1.44986E-05	1	NM_032756	Island
ALS2CL	1.46712E-05	3	NM_147129	Island
CRIP3	1.47846E-05	6	NM_206922	Island
ARHGAP24	1.57274E-05	4	NM_001025616	Open sea
EYA4	1.58661E-05	6	NM_172105	Island

Gene	p-value	Chr	Accession	Region
GRAMD2	1.59222E-05	15	NM_001012642	Island
RAPGEFL1	1.61086E-05	17	NM_016339	Island
GAMT	1.6159E-05	19	NM_138924	Island
ZNF296	1.66771E-05	19	NM_145288	Island
PITX3	1.68066E-05	10	NM_005029	Island
RASGRP2	1.70518E-05	11	NM_001098670	Island
MYO15B	1.71538E-05	17	NR_003587	Island
SIAH1	1.72744E-05	16	NM_001006610	Island
WWTR1	1.72852E-05	3	NM_001168278	Island
HIST1H4L	1.7318E-05	6	NM_003546	Island
EPS8L2	1.77488E-05	11	NM_022772	Island
TMED7-	1.80699E-05	5	NM_001164468	Island
OSBPL9	1.81543E-05	1	NM_148905	Island
IQSEC1	1.81716E-05	3	NM_001134382	Island
RASSF10	1.82075E-05	11	NM_001080521	Island
VILL	1.85366E-05	3	NM_015873	Island
CRADD	1.88018E-05	12	NM_003805	Open sea
GNMT	1.89112E-05	6	NM_018960	Island
TUBA4B	1.96087E-05	2	NR_003063	N_Shore
HLA-H	2.04392E-05	6	NR_001434	Island
DRD4	2.07047E-05	11	NM_000797	Island
NBL1	2.07453E-05	1	NM_005380	Island
CORO6	2.08242E-05	17	NM_032854	Island
HLA-H	2.11307E-05	6	NR_001434	Island
CCNI2	2.13917E-05	5	NM_001039780	N_Shore
TTC12	2.14197E-05	11	NM_017868	N_Shore
HHIPL1	2.17074E-05	14	NM_001127258	Island
DERL3	2.20237E-05	22	NM_001135751	Island
COL18A1	2.2513E-05	21	NM_130445	Island

Gene	p-value	Chr	Accession	Region
MYL12A	2.25229E-05	18	NM_006471	N_Shore
BCAT1	2.35399E-05	12	NM_005504	N_Shore
GRAMD2	2.4753E-05	15	NM_001012642	Island
RBP1	2.47974E-05	3	NM_002899	Island
UCP2	2.50183E-05	11	NM_003355	Island
NSUN7	2.54102E-05	4	NM_024677	Island
NTN1	2.54146E-05	17	NM_004822	Island
RNF135	2.55788E-05	17	NM_197939	N_Shore
PTPRN	2.67136E-05	2	NM_002846	Island
SRRM3	2.70039E-05	7	NM_001110199	Island
FAM123C	2.70277E-05	2	NM_001105194	N_Shore
ZC3HAV1L	2.71198E-05	7	NM_080660	Island
SNX31	2.74493E-05	8	NM_152628	Island
HPCAL1	2.75465E-05	2	NM_002149	Island
COL11A2	2.83082E-05	6	NM_001163771	Island
SCRN1	2.83615E-05	7	NM_014766	Island
PNMT	2.99557E-05	17	NM_002686	Island
HLA-L	3.00901E-05	6	NR_027822	Island
ANKRD34A	3.00976E-05	1	NM_001039888	Island
SFRP4	3.01268E-05	7	NM_003014	N_Shore
ARL4A	3.0146E-05	7	NM_212460	Island
NCRNA00085	3.02042E-05	19	NR_024330	S_Shore
CMYA5	3.10036E-05	5	NM_153610	Open sea
COL9A2	3.10378E-05	1	NM_001852	Island
RAB11FIP4	3.12377E-05	17	NM_032932	Island
CNTN2	3.1676E-05	1	NM_005076	Open sea
LZTS2	3.25539E-05	10	NM_032429	Island
ISLR2	3.33012E-05	15	NM_001130136	Island
TMEM22	3.38128E-05	3	NM_001097600	Island

GSX2	3.38135E-05	4	NM_133267	N_Shore
GJA1	3.42258E-05	6	NM_000165	Open sea
ZEB1	3.44758E-05	10	NM_030751	Island
NFKBIZ	3.4673E-05	3	NM_001005474	N_Shore
MAL	3.51821E-05	2	NM_022438	N_Shore
RAB27B	3.54795E-05	18	NM_004163	Island
SERINC2	3.54857E-05	1	NM_178865	Island
SLC22A16	3.65879E-05	6	NM_033125	Island
NIPAL4	3.69639E-05	5	NM_001099287	Island
PSD	3.69837E-05	10	NM_002779	Island
RASL12	3.78715E-05	15	NM_016563	Island
NAAA	3.94808E-05	4	NM_001042402	Island
ZC3HAV1L	3.94952E-05	7	NM_080660	Island
HSPA2	3.98801E-05	14	NM_021979	S_Shore
KIAA1614	3.99499E-05	1	NM_020950	Island
RASSF10	4.03091E-05	11	NM_001080521	Island
LRRC3B	4.08563E-05	3	NM_052953	N_Shore
ADAM12	4.13245E-05	10	NM_021641	Island
CYP11A1	4.13375E-05	15	NM_001099773	Island
DKK1	4.15904E-05	10	NM_012242	N_Shore
GDNF	4.23896E-05	5	NM_199231	Island
ATP5G2	4.3123E-05	12	NM_005176	Island
GIPC2	4.40169E-05	1	NM_017655	Island
RBP1	4.44154E-05	3	NM_002899	Island
SPINT2	4.44701E-05	19	NM_021102	Island
CACNG8	4.4541E-05	19	NM_031895	Island
PAX6	4.47243E-05	11	NM_001127612	Island
NRG2	4.49446E-05	5	NM_013982	Island
ARFGAP3	4.59309E-05	22	NM_014570	Island

GDNF	4.59681E-05	5	NM_199231	Island
TMEM178	4.598E-05	2	NM_152390	N_Shore
ADAM12	4.60443E-05	10	NM_021641	Island
ALOX5	4.83246E-05	10	NM_000698	Island
CDH3	4.83567E-05	16	NM_001793	Island
HSPB1	4.84963E-05	7	NM_001540	Island
UNC13A	4.88228E-05	19	NM_001080421	Island
RAPGEFL1	4.88728E-05	17	NM_016339	Island
CACNG8	4.91099E-05	19	NM_031895	N_Shore
ACVRL1	4.98738E-05	12	NM_000020	Island
LZTS2	5.17107E-05	10	NM_032429	Island
GALNT14	5.1901E-05	2	NM_024572	Island
MIR375	5.24404E-05	2	NR_029867	Island
BCL2	5.26713E-05	18	NM_000633	Island
COL9A2	5.38567E-05	1	NM_001852	Island
NCRNA00085	5.45087E-05	19	NR_024330	Island
GLUL	5.45287E-05	1	NM_001033044	Island
MYRIP	5.48504E-05	3	NM_015460	Island
SPTBN1	5.49219E-05	2	NM_178313	Island
RAP1GAP	5.51272E-05	1	NM_002885	Island
GRAMD2	5.78441E-05	15	NM_001012642	Island
PPP1CC	5.8921E-05	12	NM_002710	S_Shore
HIST1H4L	5.92954E-05	6	NM_003546	S_Shore
IGDCC3	5.95457E-05	15	NM_004884	Island
TBR1	5.96627E-05	2	NM_006593	N_Shore
MACROD1	6.03535E-05	11	NM_014067	Island
SGK1	6.03837E-05	6	NM_001143676	S_Shore
CFLAR	6.0443E-05	2	NM_001127184	Island
GALNT14	6.09087E-05	2	NM_024572	Island

Gene	p-value	Chr	RefSeq	Region
VWDE	6.2665E-05	7	NM_001135924	Island
MEGF10	6.36591E-05	5	NM_032446	Island
CCNI2	6.4448E-05	5	NM_001039780	N_Shore
NFATC1	6.48E-05	18	NM_006162	Island
RBP7	6.54137E-05	1	NM_052960	Island
KIF5C	6.62432E-05	2	NM_004522	Island
MEGF11	6.72694E-05	15	NM_032445	Island
PAX6	6.78534E-05	11	NM_001127612	Island
HMSD	6.79669E-05	18	NM_001123366	Open sea
SFRP4	6.79699E-05	7	NM_003014	Island
KBTBD11	6.80068E-05	8	NM_014867	Island
VILL	6.90776E-05	3	NM_015873	Island
PPP1CC	6.96391E-05	12	NM_002710	S_Shore
FRZB	6.96725E-05	2	NM_001463	N_Shore
ZBED3	6.99392E-05	5	NM_032367	Island
BCAT1	7.0823E-05	12	NM_005504	N_Shore
MCF2L2	7.19385E-05	3	NM_015078	S_Shore
LRAT	7.35241E-05	4	NM_004744	N_Shore
NRXN1	7.35685E-05	2	NM_004801	S_Shore
TOX2	7.41495E-05	20	NM_001098796	Island
VAX2	7.4417E-05	2	NM_012476	Island
SCGB3A1	7.51561E-05	5	NM_052863	Island
NSD1	7.51575E-05	5	NM_172349	Island
TGFB2	7.56556E-05	1	NM_003238	S_Shore
MAL	7.63432E-05	2	NM_022438	Island
DDN	7.67655E-05	12	NM_015086	Island
NR2E1	7.88711E-05	6	NM_003269	Island
HPDL	7.89306E-05	1	NM_032756	Island
LZTS2	7.9281E-05	10	NM_032429	Island

Gene	p-value	Chr	Accession	Region
OGDHL	8.07689E-05	10	NM_001143996	Island
BCL2	8.148E-05	18	NM_000657	Island
COL12A1	8.17385E-05	6	NM_004370	S_Shore
FKBP5	8.19039E-05	6	NM_001145777	Island
RASGRP2	8.27191E-05	11	NM_001098671	Island
PRDX1	8.27343E-05	1	NM_181696	Island
EMILIN3	8.29917E-05	20	NM_052846	Island
PGCP	8.59387E-05	8	NM_016134	Island
TXNRD1	8.5993E-05	12	NM_182729	Island
NPM2	8.62275E-05	8	NM_182795	N_Shore
RASL10B	8.68887E-05	17	NM_033315	Island
SSH3	8.79235E-05	11	NM_017857	N_Shore
KBTBD11	8.7929E-05	8	NM_014867	N_Shore
BRUNOL5	8.81939E-05	19	NM_021938	Island
HPCAL1	8.95337E-05	2	NM_002149	Island
RAB27B	8.9905E-05	18	NM_004163	Island
CHD5	8.99151E-05	1	NM_015557	Island
TMBIM1	9.00343E-05	2	NM_022152	Island
MYO15B	9.02204E-05	17	NR_003587	Island
NR4A1	9.04625E-05	12	NM_173157	Island
GNAL	9.09348E-05	18	NM_002071	Island
ZIC2	9.14732E-05	13	NM_007129	Island
HLA-L	9.1517E-05	6	NR_027822	Island
UNC13A	9.23218E-05	19	NM_001080421	Island
BCAT1	9.2372E-05	12	NM_005504	N_Shore
COL21A1	9.30333E-05	6	NM_030820	S_Shore
ZFHX3	9.7927E-05	16	NM_001164766	Island
LYPD1	9.91959E-05	2	NM_001077427	N_Shore
SYCE2	9.93054E-05	19	NM_001105578	Island

TRH	0.000100843	3	NM_007117	Island
MOSC2	0.000101356	1	NM_017898	Island
CCDC122	0.000101736	13	NM_144974	Island
RAB27B	0.000101912	18	NM_004163	Island
ZC3HAV1L	0.000102915	7	NM_080660	Island
RAB11FIP4	0.00010421	17	NM_032932	Island
BIRC3	0.000104841	11	NM_001165	Island
HLA-J	0.000107115	6	NR_024240	Island
KLHL35	0.000111784	11	NM_001039548	Island
DLGAP3	0.000112311	1	NM_001080418	N_Shore
RAB27B	0.000114046	18	NM_004163	Island
LOXL3	0.000114582	2	NM_032603	Island
CPLX1	0.000114756	4	NM_006651	Island
SYT7	0.000115329	11	NM_004200	Island
CHST6	0.000115423	16	NM_021615	Island
GALNT14	0.000117564	2	NM_024572	Island
RPP25	0.000120953	15	NM_017793	Island
GALNT14	0.00012399	2	NM_024572	Island
CYP11A1	0.000126031	15	NM_001099773	Island
NCAM2	0.000126057	21	NM_004540	Island
CFTR	0.000126119	7	NM_000492	Open sea
MOGAT3	0.00012801	7	NM_178176	Island
GDNF	0.000128351	5	NM_199231	Island
ADAM8	0.000128841	10	NM_001164489	Island
HLA-J	0.000129697	6	NR_024240	Island
FOXP1	0.000130286	3	NM_032682	Island
CLSTN1	0.000130611	1	NM_014944	Island
MCF2L2	0.000130667	3	NM_015078	Island
CBR1	0.000135243	21	NM_001757	Island

FBRSL1	0.000135372	12	NM_001142641	Island
HLA-L	0.000136094	6	NR_027822	Island
HSPB1	0.000136471	7	NM_001540	Island
BIRC3	0.000136545	11	NM_182962	Island
USP2	0.000137746	11	NM_004205	Island
RASSF10	0.000137761	11	NM_001080521	S_Shore
KY	0.000137814	3	NM_178554	Island
IER3	0.000138789	6	NM_003897	Island
BMPR2	0.000143074	2	NM_001204	Island
TMEM63A	0.00014463	1	NM_014698	Island
BCAT1	0.000144928	12	NM_005504	N_Shore
LMO2	0.000148203	11	NM_001142315	Island
CFLAR	0.000149419	2	NM_001127184	Island
CACNB4	0.000149527	2	NM_000726	Island
CBR1	0.000153952	21	NM_001757	Island
GPR12	0.000154089	13	NM_005288	Island
GRID2IP	0.000156405	7	NM_001145118	Island
FGFRL1	0.000156594	4	NM_001004358	Island
MACROD1	0.000157296	11	NM_014067	Open sea
DNAJA4	0.000158121	15	NM_018602	Island
FAM160A1	0.00015836	4	NM_001109977	Island
DPYSL5	0.000158927	2	NM_020134	Island
CHST8	0.000160048	19	NM_001127895	Island
NSUN7	0.000160526	4	NM_024677	Island
ISLR2	0.000160955	15	NM_001130136	Island
GDNF	0.000160979	5	NM_199231	Island
SSH3	0.000161026	11	NM_017857	Island
EVC2	0.000161093	4	NM_001166136	Island
HLA-J	0.000162642	6	NR_024240	Island

ISLR2	0.000164173	15	NM_001130136	Island
AEBP1	0.000164177	7	NM_001129	Open sea
IGDCC3	0.00016628	15	NM_004884	N_Shore
FAM132A	0.00016675	1	NM_001014980	Island
GPR12	0.000167887	13	NM_005288	Island
P2RY2	0.000167935	11	NM_176071	Island
GRAMD2	0.000169018	15	NM_001012642	Island
KLF16	0.000169879	19	NM_031918	Island
NFKBIZ	0.000170814	3	NM_001005474	Island
IGDCC3	0.00017192	15	NM_004884	Island
ADAM8	0.000176853	10	NM_001164490	Island
ITGB5	0.000176944	3	NM_002213	Island
RASL10B	0.000177054	17	NM_033315	Island
PRDX1	0.000177322	1	NM_002574	Island
TXNRD1	0.000178139	12	NM_182729	Island
KLHL35	0.0001786	11	NM_001039548	Island
RHBDF2	0.000179009	17	NM_024599	Island
NSD1	0.000179325	5	NM_022455	Island
BMP8A	0.000179329	1	NM_181809	Island
NGB	0.000180278	14	NM_021257	Island
BCAT1	0.000181158	12	NM_005504	N_Shore
ZNF575	0.000181242	19	NM_174945	Island
CALHM2	0.000181486	10	NM_015916	Island
TTBK1	0.000182945	6	NM_032538	Island
UBE2E2	0.000183072	3	NM_152653	Island
GALNT14	0.000183948	2	NM_024572	Island
PRDX1	0.000184226	1	NM_181696	S_Shore
XKR8	0.000185159	1	NM_018053	Island
CALHM2	0.000185203	10	NM_015916	Island

CRADD	0.000185913	12	NM_003805	Open sea
SOST	0.00018812	17	NM_025237	Island
QRFPR	0.000188558	4	NM_198179	Island
HLCS	0.000190231	21	NM_000411	Island
PAMR1	0.000190534	11	NM_001001991	S_Shore
SLC38A1	0.000192726	12	NM_001077484	Island
GRIN2C	0.000193939	17	NM_000835	Island
RBP1	0.000195343	3	NM_001130993	Island
NBEA	0.000196366	13	NM_015678	Island
EPHA2	0.000197708	1	NM_004431	Island
DDN	0.000198957	12	NM_015086	Island
PRKCZ	0.000199841	1	NM_001033582	Island
SHH	0.000200187	7	NM_000193	Island
NR4A1	0.000200228	12	NM_173157	Island
KCNAB3	0.000200753	17	NM_004732	Island
FKBP5	0.000200807	6	NM_001145777	S_Shore
BMP8B	0.00020095	1	NM_001720	Island
KCNAB3	0.000201472	17	NM_004732	Island
CLDN5	0.000201587	22	NM_001130861	Island
LZTS2	0.000201782	10	NM_032429	Island
CTHRC1	0.000202429	8	NM_138455	Island
CFLAR	0.00020347	2	NM_001127184	Island
SGK1	0.000204878	6	NM_001143676	S_Shelf
CACNB4	0.000208813	2	NM_000726	N_Shore
NFKBIZ	0.000210636	3	NM_001005474	Island
ATP5G2	0.000211745	12	NM_005176	Island
RASGRP2	0.000212867	11	NM_001098670	Island
LZTS2	0.000213195	10	NM_032429	Island
GRIN2B	0.000213739	12	NM_000834	N_Shore

KLF5	0.000214272	13	NM_001730	Island
SNX31	0.000215981	8	NM_152628	Island
GAMT	0.00021791	19	NM_138924	Island
CENPV	0.000218035	17	NM_181716	Island
TMEM143	0.000222313	19	NM_018273	S_Shore
COL12A1	0.000222314	6	NM_004370	Island
CYP27A1	0.000222846	2	NM_000784	Island
RNF39	0.000224261	6	NM_025236	Island
HIST1H4L	0.000231123	6	NM_003546	S_Shore
NR2E1	0.000232616	6	NM_003269	Island
LIMS2	0.000233226	2	NM_001136037	Island
SFRP2	0.000233279	4	NM_003013	Island
LYN	0.000240488	8	NM_002350	Island
RAB3D	0.000242775	19	NM_004283	Island
ISLR2	0.000242805	15	NM_001130136	Island
NPPB	0.000244791	1	NM_002521	Island
PLLP	0.000245261	16	NM_015993	Island
ME1	0.000247165	6	NM_002395	Island
PACS2	0.000249165	14	NM_015197	Island
LRAT	0.000250667	4	NM_004744	Island
PRR15	0.000250969	7	NM_175887	N_Shore
ADAM8	0.0002511	10	NM_001164489	Island
NBEA	0.000255155	13	NM_015678	Island
HLA-J	0.000257788	6	NR_024240	Island
FBRSL1	0.000257951	12	NM_001142641	Island
ACTG1	0.000258316	17	NM_001614	Island
SFRP4	0.000258593	7	NM_003014	Island
SIAH1	0.000261228	16	NM_001006610	Island
MYCBPAP	0.000263215	17	NM_032133	Island

SYT6	0.000266541	1	NM_205848	Island
HTR6	0.000267709	1	NM_000871	Island
SLC22A15	0.000275621	1	NM_018420	Island
ARL9	0.000275725	4	NM_206919	Island
RAB27B	0.000276821	18	NM_004163	Island
NRXN1	0.000276912	2	NM_004801	S_Shore
NR4A1	0.00027785	12	NM_002135	Island
ARL9	0.000278605	4	NM_206919	Island
PSD	0.000279501	10	NM_002779	Island
TGFB2	0.00028092	1	NM_003238	Island
SIAH1	0.000282675	16	NM_001006610	S_Shore
NCRNA00092	0.000285602	9	NR_024129	Island
RAB11FIP4	0.000291691	17	NM_032932	Island
RNF135	0.000291856	17	NM_197939	N_Shore
KLF5	0.000291973	13	NM_001730	Island
LPCAT2	0.000291986	16	NM_017839	Island
SRRM3	0.000292585	7	NM_001110199	Island
CD109	0.000295426	6	NM_001159588	Island
PDE4A	0.000295591	19	NM_001111308	S_Shore
RPP25	0.000299182	15	NM_017793	N_Shore
KBTBD11	0.000299954	8	NM_014867	Island
SSH3	0.000301119	11	NM_017857	Island
FBRSL1	0.000305063	12	NM_001142641	Island
LLGL2	0.000307948	17	NM_001031803	Island
SOCS3	0.000309082	17	NM_003955	Island
TMEM22	0.000309548	3	NM_001097600	Island
CXCR4	0.000309674	2	NM_003467	Island
PGCP	0.000311418	8	NM_016134	Island
IER3	0.000311625	6	NM_003897	Island

RFX4	0.000311864	12	NM_213594	Island
FAM123C	0.000313282	2	NM_001105194	N_Shore
TEKT3	0.000313681	17	NM_031898	N_Shore
KLF16	0.000315159	19	NM_031918	Island
TTC12	0.000316088	11	NM_017868	N_Shore
NPPB	0.000316247	1	NM_002521	Island
LOXL1	0.00031876	15	NM_005576	Island
LOXL1	0.000319256	15	NM_005576	Island
CTSZ	0.000322055	20	NM_001336	Island
ANKS1B	0.000324755	12	NM_020140	Island
RFX4	0.000325104	12	NM_213594	Island
NCAM2	0.00032515	21	NM_004540	Island
TMEM63A	0.000327762	1	NM_014698	Island
DKK1	0.000328324	10	NM_012242	N_Shore
MT1M	0.000338177	16	NM_176870	Island
RBP1	0.000343437	3	NM_001130992	S_Shore
GRAMD2	0.000345173	15	NM_001012642	Island
TET1	0.000345868	10	NM_030625	S_Shore
EFEMP2	0.000346636	11	NM_016938	Island
LPAR2	0.000349826	19	NM_004720	Island
CRADD	0.000358521	12	NM_003805	Open sea
HIST1H4L	0.000358669	6	NM_003546	Island
ZNF296	0.000358765	19	NM_145288	Island
TMBIM1	0.000359944	2	NM_022152	S_Shore
FAM132A	0.00036331	1	NM_001014980	S_Shore
ITGB5	0.000364907	3	NM_002213	Island
KIAA0495	0.000367181	1	NM_207306	Island
ACAA1	0.000367538	3	NR_024024	Island
BCAT1	0.000368659	12	NM_005504	N_Shore

RTBDN	0.000368752	19	NM_031429	S_Shore
RNF135	0.000369371	17	NM_197939	N_Shore
CBR1	0.000372177	21	NM_001757	Island
CFLAR	0.000372281	2	NM_001127184	Island
ADPRH	0.000372768	3	NM_001125	Island
GDNF	0.000373604	5	NM_199234	Island
VILL	0.000374342	3	NM_015873	S_Shore
KLHL35	0.000375239	11	NM_001039548	Island
NTN1	0.0003792	17	NM_004822	Island
MGST2	0.000389027	4	NM_002413	Open sea
LTBP4	0.000393103	19	NM_001042544	Island
RNF135	0.00039773	17	NM_032322	Island
RASSF10	0.000398322	11	NM_001080521	Island
GRASP	0.00039901	12	NM_181711	Island
CRIP3	0.000402684	6	NM_206922	Island
GRAMD2	0.000403322	15	NM_001012642	Island
VAX2	0.000408806	2	NM_012476	Island
CACNB4	0.000410292	2	NM_000726	N_Shore
MDK	0.000423663	11	NM_001012334	N_Shore
NBEA	0.000424288	13	NM_015678	Island
MT1M	0.000424987	16	NM_176870	Island
GNMT	0.000426799	6	NM_018960	Island
MACROD1	0.000428521	11	NM_014067	N_Shore
CHST8	0.000435957	19	NM_001127895	Island
PRDX1	0.000436502	1	NM_181696	Island
ANKRD34A	0.0004415	1	NM_001039888	Island
PRDX1	0.000441863	1	NM_181696	Island
TMEM51	0.000442834	1	NM_001136218	Island
RBM24	0.000443265	6	NM_153020	S_Shore

Gene	Value	Chr	Accession	Region
GRAMD2	0.000448741	15	NM_001012642	Island
KLHL35	0.000451141	11	NM_001039548	Island
CFLAR	0.000451305	2	NM_001127184	Island
COL9A3	0.000453017	20	NM_001853	Island
GRAMD2	0.000459478	15	NM_001012642	Island
ARAP1	0.000465461	11	NM_001040118	Island
FLOT1	0.000465682	6	NM_005803	Island
CRIP3	0.000469137	6	NM_206922	Island
LHX9	0.000469572	1	NM_020204	N_Shore
DPYSL5	0.000473049	2	NM_020134	Island
DLG1	0.00047539	3	NM_004087	Island
CACNG8	0.000490409	19	NM_031895	Island
ADPRH	0.000492172	3	NM_001125	Island
SERPINB6	0.000493704	6	NM_004568	S_Shore
FBRSL1	0.000502562	12	NM_001142641	Island
LAMP3	0.000510495	3	NM_014398	Island
FAM19A1	0.000512292	3	NM_213609	Island
FBRSL1	0.000517161	12	NM_001142641	Island
MACROD1	0.000518072	11	NM_014067	Open sea
LENG9	0.000519267	19	NM_198988	Island
HAPLN3	0.000522666	15	NM_178232	Island
PAMR1	0.000522928	11	NM_015430	Island
TET1	0.000529019	10	NM_030625	S_Shore
ALDH5A1	0.00052931	6	NM_001080	Island
TOM1L1	0.000535353	17	NM_005486	Island
TRIM36	0.000539535	5	NM_018700	Island
FBXO6	0.000539665	1	NM_018438	Island
MYCBPAP	0.000557761	17	NM_032133	Island
OGDHL	0.000567291	10	NM_001143996	Island

Gene	Value	Chr	Accession	Region
MDK	0.000572534	11	NM_001012334	N_Shore
PRDX1	0.00057254	1	NM_181696	Island
LIN28	0.000572653	1	NM_024674	N_Shore
ACVRL1	0.000576039	12	NM_000020	Island
ARL9	0.000576254	4	NM_206919	Island
MEOX2	0.000592	7	NM_005924	Open sea
SLC6A11	0.00059313	3	NM_014229	Island
NPHS2	0.000593292	1	NM_014625	Island
CRADD	0.000595679	12	NM_003805	Open sea
NFKBIZ	0.000596648	3	NM_001005474	Island
RBP1	0.000598092	3	NM_001130992	Island
CFLAR	0.00060256	2	NM_001127184	Island
SWAP70	0.000605984	11	NM_015055	N_Shore
HLA-J	0.00061398	6	NR_024240	Island
PANX2	0.000622033	22	NM_001160300	Island
HHIPL1	0.000625052	14	NM_001127258	Island
DERL3	0.00062629	22	NM_001135751	Island
ARL9	0.000626651	4	NM_206919	S_Shore
FBP1	0.000626978	9	NM_001127628	Island
KLF16	0.000644953	19	NM_031918	Island
NPM2	0.000645147	8	NM_182795	N_Shore
NIPAL4	0.000650733	5	NM_001099287	Island
DDN	0.000650933	12	NM_015086	Island
NXPH4	0.00065259	12	NM_007224	Island
TMEM144	0.00065777	4	NM_018342	N_Shore
SEC31B	0.000662913	10	NM_015490	Island
ZNF575	0.000667616	19	NM_174945	Island
GRIN1	0.000673344	9	NM_000832	N_Shore
CLDN5	0.000676274	22	NM_001130861	Island

MYRIP	0.000678628	3	NM_015460	S_Shore
SLC22A3	0.000678695	6	NM_021977	Island
PLA2R1	0.000680986	2	NM_001007267	S_Shore
MGST2	0.0006918	4	NM_002413	Open sea
ZNF575	0.000692797	19	NM_174945	Island
TEKT3	0.000699201	17	NM_031898	S_Shore
UCP2	0.000705363	11	NM_003355	Island
HLA-J	0.00070811	6	NR_024240	Island
ME1	0.000712432	6	NM_002395	Island
EFEMP1	0.000714572	2	NM_004105	Island
RHBDF2	0.00072036	17	NM_024599	Island
RAB27B	0.00072269	18	NM_004163	Island
PAK1	0.000725238	11	NM_002576	Island
KCNB1	0.00072562	20	NM_004975	Island
KCNAB3	0.000728718	17	NM_004732	Island
RBM24	0.000729163	6	NM_001143941	Island
LOXL1	0.000729845	15	NM_005576	Island
STAT4	0.000744629	2	NM_003151	Open sea
GJA1	0.000744755	6	NM_000165	Open sea
NID2	0.000749638	14	NM_007361	Island
PHF11	0.000755566	13	NM_001040444	N_Shore
ARL4A	0.000767595	7	NM_212460	Island
HLA-L	0.000773254	6	NR_027822	Island
SOCS3	0.000778414	17	NM_003955	Island
CLDN5	0.000778526	22	NM_001130861	Island
PIM3	0.000781687	22	NM_001001852	Island
PRKCZ	0.000781939	1	NM_001033582	Island
BMPR2	0.000787025	2	NM_001204	Island
CFLAR	0.000787095	2	NM_001127184	Island

Gene	p-value	Chr	Accession	Region
MOSC2	0.000804158	1	NM_017898	N_Shore
CACNG8	0.000817051	19	NM_031895	N_Shore
NFKBIZ	0.00082583	3	NM_001005474	Island
ISLR2	0.000826167	15	NM_001130136	Island
CHST6	0.000849933	16	NM_021615	Island
BMP8B	0.000850564	1	NM_001720	Island
RGS22	0.000851375	8	NM_015668	Island
KIAA0040	0.000859517	1	NM_001162893	Island
SSH3	0.000861295	11	NM_017857	Island
HPCA	0.000866829	1	NM_002143	Island
CYP11A1	0.000870469	15	NM_001099773	Island
HAPLN3	0.000870714	15	NM_178232	N_Shore
EYA4	0.000870775	6	NM_172105	N_Shore
COL9A2	0.000880496	1	NM_001852	Island
CDK5R2	0.000883543	2	NM_003936	N_Shore
PPP1R14A	0.000883631	19	NM_033256	Island
FBXO6	0.000895913	1	NM_018438	Island
CCNA1	0.000898514	13	NM_003914	S_Shore
MIR155HG	0.000901555	21	NR_001458	Island
ZIC2	0.000917402	13	NM_007129	Island
NXPH4	0.000919929	12	NM_007224	Island
PHLDB1	0.000922539	11	NM_001144758	Island
NPM2	0.000922649	8	NM_182795	Island
ZAR1	0.000924757	4	NM_175619	Island
DEDD2	0.000924899	19	NM_133328	Island
ZNF204P	0.000925361	6	NR_024553	Island
SEC31B	0.000925598	10	NM_015490	Island
HHIPL1	0.000925854	14	NM_032425	Island
CENPV	0.000932512	17	NM_181716	Island

CALHM2	0.00093809	10	NM_015916	Island
TMEM144	0.000938365	4	NM_018342	Island
CFLAR	0.000949	2	NM_001127184	N_Shore
MGST2	0.000949334	4	NM_002413	Open sea
SLC13A5	0.000949502	17	NM_177550	Island
TBR1	0.000952431	2	NM_006593	Island
SIAH1	0.000954333	16	NM_001006610	S_Shore
RNLS	0.000954731	10	NM_001031709	Island
KIAA0495	0.000964229	1	NM_207306	Island
SOCS3	0.000965236	17	NM_003955	Island
SST	0.00097032	3	NM_001048	Island
TMBIM1	0.000970676	2	NM_022152	Island
THRB	0.000979577	3	NM_001128177	Island
IQSEC1	0.000980822	3	NM_001134382	Island
SLC38A1	0.000982642	12	NM_001077484	Island
TXNRD1	0.000986673	12	NM_182729	Island
FBXO6	0.001002185	1	NM_018438	Island
MGST2	0.001012574	4	NM_002413	Open sea
KCNJ3	0.001013161	2	NM_002239	Island
RBP7	0.001013165	1	NM_052960	Island
DNAJA4	0.00101521	15	NM_018602	Island
GRID2IP	0.001019732	7	NM_001145118	Island
HLA-J	0.001027075	6	NR_024240	Island
GRM8	0.001037613	7	NM_001127323	N_Shore
MAL	0.001049848	2	NM_022438	Island
KLC2	0.00105142	11	NM_001134776	N_Shore
ADPRH	0.001053024	3	NM_001125	Island
MAST1	0.001055483	19	NM_014975	Island
RAP1GAP	0.001056076	1	NM_002885	Island

WWTR1	0.001061327	3	NM_001168278	S_Shore
RUNX1	0.00106174	21	NM_001754	Island
HIST1H4L	0.001062652	6	NM_003546	S_Shore
ADPRH	0.001085218	3	NM_001125	Island
EPS8L2	0.0010854	11	NM_022772	Island
BRUNOL5	0.001086274	19	NM_021938	N_Shore
DKK1	0.001087141	10	NM_012242	N_Shore
RHBDF2	0.001090169	17	NM_024599	Island
TBR1	0.001095059	2	NM_006593	Island
RBP7	0.001097744	1	NM_052960	Island
EHBP1L1	0.001103609	11	NM_001099409	Island
GPR12	0.00110507	13	NM_005288	Island
PANX2	0.001113958	22	NM_001160300	Island
CYB5R1	0.00112142	1	NM_016243	Island
CYP27A1	0.00113471	2	NM_000784	Island
RILP	0.001142421	17	NM_031430	Island
PRKCZ	0.001145758	1	NM_002744	N_Shore
PPP1R14A	0.001150891	19	NM_033256	Island
PLLP	0.001161565	16	NM_015993	Island
GRASP	0.001169359	12	NM_181711	Island
CTSZ	0.001172774	20	NM_001336	Island
ADPRH	0.001173162	3	NM_001125	Island
SLC22A3	0.001173438	6	NM_021977	Island
LAMP3	0.001178345	3	NM_014398	Island
ESAM	0.001183081	11	NM_138961	N_Shore
PRKCZ	0.001184579	1	NM_001033582	Island
ADPRH	0.001194517	3	NM_001125	Island
TGFB2	0.0011947	1	NM_001135599	N_Shore
TGFB2	0.001213378	1	NM_001135599	Island

HIST1H4L	0.001215286	6	NM_003546	S_Shore
COL12A1	0.001217205	6	NM_004370	Island
DLC1	0.001217558	8	NM_182643	Island
MEGF10	0.001222943	5	NM_032446	Island
HSPA2	0.001223619	14	NM_021979	Island
ZNF204P	0.00123662	6	NR_024553	N_Shore
NXPH4	0.001236833	12	NM_007224	Island
ALOX5	0.001238671	10	NM_000698	Island
SPAG1	0.00124268	8	NM_003114	N_Shore
HPCA	0.001243993	1	NM_002143	Island
SLC25A20	0.001248383	3	NM_000387	N_Shore
CRIP3	0.001258951	6	NM_206922	S_Shore
MGST2	0.001283859	4	NM_002413	Open sea
KIF5C	0.001286796	2	NM_004522	Island
ACTA1	0.001301267	1	NM_001100	Island
TNXB	0.001309436	6	NM_019105	S_Shore
PIM3	0.001312774	22	NM_001001852	Island
ACAA1	0.001313501	3	NR_024024	Island
MYO18A	0.001315034	17	NM_078471	Island
ANKRD34A	0.001316708	1	NM_001039888	Island
RAB27B	0.001326379	18	NM_004163	N_Shore
ESAM	0.001337996	11	NM_138961	Island
MOSC2	0.001361007	1	NM_017898	N_Shore
RBM24	0.001366241	6	NM_001143941	S_Shore
EFEMP1	0.001366859	2	NM_004105	Island
ONECUT2	0.001372006	18	NM_004852	Open sea
SLC38A1	0.001372325	12	NM_001077484	Island
ARHGAP24	0.001376991	4	NM_001025616	Open sea
MAP7	0.001384252	6	NM_003980	Island

Gene	Value	Chr	Accession	Location
GIPR	0.001398656	19	NM_000164	Island
RGNEF	0.001399873	5	NM_001080479	Island
MEOX2	0.001424693	7	NM_005924	Open sea
ZFHX3	0.001429172	16	NM_001164766	Island
NUDT16	0.001434994	3	NM_152395	Island
CD8A	0.001446664	2	NM_001145873	Island
KLF11	0.00144784	2	NM_003597	Island
FBXO6	0.001482306	1	NM_018438	Island
FZD6	0.001494476	8	NM_001164616	Island
PAMR1	0.001496345	11	NM_001001991	S_Shore
GRIN2B	0.00150214	12	NM_000834	N_Shore
ERBB2	0.001502237	17	NM_001005862	Island
CDH3	0.001502821	16	NM_001793	Island
GDNF	0.001513562	5	NM_000514	Island
FBRSL1	0.001551126	12	NM_001142641	N_Shore
DDIT4L	0.001561465	4	NM_145244	S_Shore
NCRNA00092	0.001568671	9	NR_024129	Island
RBP1	0.001572528	3	NM_001130993	Island
EFNB2	0.001584001	13	NM_004093	Island
LECT1	0.001591838	13	NM_001011705	Island
MAL	0.001592404	2	NM_022440	Island
ACAA1	0.001608885	3	NR_024024	Island
LGALS3	0.001624665	14	NM_002306	S_Shore
MIR375	0.001637917	2	NR_029867	Island
MYRIP	0.001647145	3	NM_015460	S_Shore
HLF	0.001669242	17	NM_002126	S_Shore
TMEM26	0.001678889	10	NM_178505	S_Shore
TGFB2	0.001679117	1	NM_003238	S_Shore
GNAL	0.001679871	18	NM_002071	Island

GPR19	0.001681485	12	NM_006143	Island
BIRC3	0.001708525	11	NM_182962	N_Shore
GPR19	0.001711497	12	NM_006143	Island
CTHRC1	0.00171258	8	NM_138455	Island
FBXO6	0.001724421	1	NM_018438	Island
NCRNA00085	0.001725088	19	NR_024330	Island
PTEN	0.001733654	10	NM_000314	Island
ZMYND10	0.0017499	3	NM_015896	Island
RNLS	0.001763236	10	NM_001031709	Island
ARL9	0.001793677	4	NM_206919	Island
EVC2	0.001812681	4	NM_001166136	Island
GSX2	0.00181827	4	NM_133267	Island
ALOX5	0.001830098	10	NM_000698	Island
BCAT1	0.001835018	12	NM_005504	Island
RASGRP2	0.00185628	11	NM_001098671	Island
RNF39	0.001867537	6	NM_025236	Island
FAM160A1	0.001883402	4	NM_001109977	Island
GNAL	0.001888588	18	NM_002071	Island
GSX2	0.001890511	4	NM_133267	Island
ARHGAP24	0.001891343	4	NM_001025616	Open sea
SSH3	0.001894589	11	NM_017857	Island
IGDCC3	0.001901522	15	NM_004884	Island
MOSC2	0.001910491	1	NM_017898	Island
NTSR2	0.001910701	2	NM_012344	Island
CPLX1	0.001911221	4	NM_006651	Island
RHOD	0.001914777	11	NM_014578	Island
SHH	0.001922714	7	NM_000193	Island
USP4	0.001922992	3	NM_199443	S_Shore
CHST8	0.001927017	19	NM_001127895	N_Shore

BCL6	0.001938216	3	NM_001706	Island
CALHM2	0.001950818	10	NM_015916	N_Shore
OSBPL9	0.0019942	1	NM_148904	N_Shore
TMEM200B	0.001995014	1	NM_001003682	Island
EYA4	0.001996214	6	NM_172105	N_Shore
EFCAB4B	0.001996492	12	NM_001144959	Island
PRR15	0.002021163	7	NM_175887	Island
QRFPR	0.002028012	4	NM_198179	Island
ISLR2	0.002028415	15	NM_001130136	Island
LYPD1	0.002037283	2	NM_001077427	Island
GAMT	0.00204182	19	NM_138924	Island
FOXP1	0.002051394	3	NM_032682	Island
EFNB2	0.002067769	13	NM_004093	Island
TMEM26	0.002075307	10	NM_178505	S_Shore
ALDH3A1	0.002080392	17	NM_001135167	Island
HDAC4	0.002083153	2	NM_006037	S_Shore
RBP1	0.002087149	3	NM_001130992	S_Shore
NIPAL4	0.002114898	5	NM_001099287	Island
APBA2	0.002115113	15	NM_001130414	Island
RASSF10	0.002125105	11	NM_001080521	Island
DERL3	0.002132223	22	NM_198440	Island
KLF5	0.002132477	13	NM_001730	S_Shore
HPCA	0.002137349	1	NM_002143	Island
COL19A1	0.002141188	6	NM_001858	N_Shore
MOSC2	0.00214662	1	NM_017898	Island
BCL2	0.002153084	18	NM_000657	Island
GDNF	0.002156056	5	NM_000514	Island
USP4	0.002157948	3	NM_003363	Island
TOM1L1	0.002158372	17	NM_005486	Island

Gene	p-value	Chr	RefSeq	Region
MEGF10	0.002162942	5	NM_032446	Island
PRDX1	0.002163387	1	NM_002574	Island
TNXB	0.002168941	6	NM_019105	Island
FBXO6	0.002176619	1	NM_018438	Island
HMSD	0.002185351	18	NM_001123366	Open sea
SSH3	0.002211447	11	NM_017857	Island
SERINC2	0.00222372	1	NM_178865	Island
PACS2	0.00223506	14	NM_015197	S_Shore
TCEA3	0.002248886	1	NM_003196	Island
ALOX5	0.002254226	10	NM_000698	S_Shore
FBRSL1	0.002254532	12	NM_001142641	Island
PHF11	0.002273777	13	NM_001040444	Island
SLC16A11	0.002286188	17	NM_153357	Island
PAX6	0.002298184	11	NM_001127612	Island
RADIL	0.002302914	7	NM_018059	Open sea
RTBDN	0.002318174	19	NM_031429	S_Shore
HSD11B2	0.002318522	16	NM_000196	Island
MACROD1	0.002322742	11	NM_014067	N_Shore
COL19A1	0.002346021	6	NM_001858	N_Shore
MMP14	0.002352248	14	NM_004995	Island
LYN	0.002355554	8	NM_002350	N_Shore
CCNI2	0.002373385	5	NM_001039780	Island
ELL3	0.002397895	15	NM_025165	Island
PSD	0.002398658	10	NM_002779	Island
WWTR1	0.002402441	3	NM_001168278	S_Shore
FBRSL1	0.002405435	12	NM_001142641	Island
TET1	0.002410655	10	NM_030625	S_Shore
LGALS3	0.002420096	14	NM_002306	Island
CABYR	0.002426493	18	NM_153768	Island

ENC1	0.002439162	5	NM_003633	S_Shore
USP4	0.002440027	3	NM_003363	Island
LTBP4	0.002445648	19	NM_001042544	Island
FLJ22536	0.002452639	6	NR_015410	S_Shore
RASL12	0.002453232	15	NM_016563	Island
SERHL2	0.002468876	22	NM_014509	Island
SLC16A11	0.002499049	17	NM_153357	Island
MFSD2B	0.00250459	2	NM_001080473	Island
APBA2	0.002510097	15	NM_001130414	Island
LTBP4	0.002511049	19	NM_001042545	Island
LYPD1	0.002518794	2	NM_001077427	Island
ALOX5	0.002524516	10	NM_000698	Island
PXDN	0.002533684	2	NM_012293	Island
NFATC4	0.002534768	14	NM_001136022	Island
SPIB	0.002535883	19	NM_003121	Island
RGNEF	0.00254265	5	NM_001080479	Island
TBR1	0.002548825	2	NM_006593	N_Shore
KCNIP3	0.002553237	2	NM_001034914	Island
MAPT	0.002555742	17	NM_001123067	Island
BRUNOL5	0.002556894	19	NM_021938	S_Shore
RASGRP2	0.002561108	11	NM_001098671	Island
FOXP1	0.002562383	3	NM_032682	Island
CYP26B1	0.002587926	2	NM_019885	Island
HIST1H4L	0.002590172	6	NM_003546	S_Shore
MACROD1	0.002612887	11	NM_014067	
CTSZ	0.002635721	20	NM_001336	Island
CCDC122	0.002641161	13	NM_144974	Island
FGFRL1	0.002642105	4	NM_001004356	Island
HLA-L	0.002645666	6	NR_027822	Island

Gene	Value	Chr	Accession	Region
SLC13A5	0.002655897	17	NM_001143838	Island
CCDC122	0.002669253	13	NM_144974	N_Shore
RAB11FIP4	0.00267758	17	NM_032932	N_Shore
CBR1	0.002694326	21	NM_001757	Island
SHANK1	0.002708705	19	NM_016148	Island
SERPINB6	0.002709904	6	NM_004568	Island
MAST1	0.002725751	19	NM_014975	Island
SHANK1	0.002726397	19	NM_016148	Island
EPHA2	0.002742936	1	NM_004431	Island
PGCP	0.002748735	8	NM_016134	Island
ARHGAP24	0.00275653	4	NM_001025616	Open sea
MACROD1	0.002769756	11	NM_014067	Open sea
CHAD	0.00277002	17	NM_001267	Island
MGST2	0.002772892	4	NM_002413	Open sea
SYT6	0.002773259	1	NM_205848	Island
ANKRD34A	0.002773859	1	NM_001039888	Island
HLA-J	0.002786534	6	NR_024240	Island
LLGL2	0.002797133	17	NM_001015002	Island
LIN28	0.002805311	1	NM_024674	N_Shore
IGSF21	0.002816568	1	NM_032880	Island
LY75	0.002816571	2	NM_002349	Island
ALDH3A1	0.002817648	17	NM_001135167	Island
SLC25A20	0.002821121	3	NM_000387	Island
PTEN	0.002828289	10	NM_000314	Island
PANX2	0.002832555	22	NM_001160300	Island
ALOX5	0.002838385	10	NM_000698	Island
VRK2	0.002869284	2	NM_001130483	Island
RBM24	0.002874362	6	NM_153020	Island
SLC38A1	0.002879349	12	NM_001077484	Island

THBS4	0.002886494	5	NM_003248	Island
LAMA3	0.002886898	18	NM_198129	Island
HAPLN3	0.002906967	15	NM_178232	Island
PICK1	0.002914599	22	NM_001039583	N_Shore
HDAC4	0.002925786	2	NM_006037	Island
CALHM2	0.002944701	10	NM_015916	Island
HDAC4	0.002953807	2	NM_006037	Island
CHPF	0.002954416	2	NM_024536	Island
RASL12	0.00297363	15	NM_016563	Island
USP4	0.002989883	3	NM_003363	Island
CAMTA1	0.003002334	1	NM_015215	Open sea
FLJ23834	0.003034206	7	NM_152750	Island
BRUNOL4	0.003059428	18	NM_001025089	Island
ACTA1	0.0030596	1	NM_001100	Island
XKR8	0.003061132	1	NM_018053	Island
EFEMP1	0.003063539	2	NM_004105	Island
NR1I2	0.003069309	3	NM_033013	Island
SHH	0.003069903	7	NM_000193	Island
BCL2	0.003070722	18	NM_000633	Island
CFLAR	0.003071274	2	NM_001127184	Island
FAM19A1	0.003071786	3	NM_213609	S_Shore
EMILIN3	0.003080442	20	NM_052846	Island
SPTBN1	0.003085514	2	NM_178313	Island
NFKBIZ	0.003088809	3	NM_001005474	Island
ACTA1	0.003090134	1	NM_001100	Island
TMEM26	0.003104673	10	NM_178505	S_Shore
COL21A1	0.003110007	6	NM_030820	S_Shore
PTEN	0.003111448	10	NM_000314	Island
SEC31B	0.003145388	10	NM_015490	Island

FKBP5	0.003151793	6	NM_001145777	Island
HSPB1	0.003151805	7	NM_001540	Island
FGFRL1	0.003152768	4	NM_001004356	Island
SLC22A16	0.003157786	6	NM_033125	Island
PAMR1	0.003168586	11	NM_015430	Island
MCFD2	0.003176594	2	NM_001171511	S_Shore
RUSC1	0.003177141	1	NM_001105205	Island
ACADS	0.003177309	12	NM_000017	Island
VSIG10L	0.003180668	19	NM_001163922	Island
SPAG1	0.00318508	8	NM_003114	N_Shore
RAB32	0.003190224	6	NM_006834	Island
BMP8B	0.003232972	1	NM_001720	Island
CAMK2B	0.003236688	7	NM_172081	Island
KBTBD11	0.003242599	8	NM_014867	S_Shore
HLA-J	0.00324821	6	NR_024240	Island
KIAA0040	0.003252647	1	NM_001162893	Island
RASGRP2	0.003256739	11	NM_153819	Island
UCP2	0.003265327	11	NM_003355	S_Shore
RNF39	0.003265709	6	NM_025236	S_Shore
ARFGAP3	0.00326654	22	NM_014570	Island
LLGL2	0.003293407	17	NM_001015002	Island
GRASP	0.003293927	12	NM_181711	Island
DEDD2	0.003296941	19	NM_133328	Island
ACTA1	0.003308264	1	NM_001100	Island
GPR12	0.003323214	13	NM_005288	Island
RGNEF	0.003337815	5	NM_001080479	Island
NAV2	0.003341296	11	NM_182964	S_Shore
OPCML	0.003403093	11	NM_001012393	Open sea
RGS22	0.003425685	8	NM_015668	Island

USP4	0.00344677	3	NM_003363	Island
THBS4	0.003490245	5	NM_003248	Island
GPR150	0.003495326	5	NM_199243	Island
PRR15	0.003514088	7	NM_175887	Island
PDPN	0.003535427	1	NM_001006624	Island
DLG1	0.003539151	3	NM_004087	Island
RADIL	0.003554246	7	NM_018059	Open sea
MIR548H4	0.003569329	15	NR_031680	S_Shore
FAM5B	0.003604725	1	NM_021165	Island
CD2AP	0.003617623	6	NM_012120	N_Shore
KCNB1	0.003622917	20	NM_004975	Island
TEKT3	0.003637355	17	NM_031898	N_Shore
ADAM8	0.003638814	10	NM_001164490	Island
GPR19	0.003679264	12	NM_006143	Island
GRID2IP	0.003688207	7	NM_001145118	Island
GJA1	0.003705472	6	NM_000165	Open sea
SOCS3	0.003721762	17	NM_003955	Island
MCFD2	0.003749857	2	NM_001171508	S_Shore
TNFAIP8L3	0.003753	15	NM_207381	Island
PNMT	0.003764244	17	NM_002686	Island
PANX2	0.003774904	22	NM_001160300	Island
SLC16A11	0.003775124	17	NM_153357	Island
PDPN	0.003790456	1	NM_001006624	Island
GSX2	0.003798511	4	NM_133267	N_Shore
RUSC1	0.003843649	1	NM_001105203	Island
PTEN	0.003854538	10	NM_000314	Island
PSD3	0.003856029	8	NM_015310	Island
GPR19	0.003931881	12	NM_006143	Island
COL21A1	0.003939097	6	NM_030820	Island

CLDN5	0.003951431	22	NM_001130861	S_Shore
RNLS	0.003956223	10	NM_001031709	Island
LOXL3	0.003961595	2	NM_032603	Island
TOM1L1	0.003974287	17	NM_005486	Island
LIN28	0.003994067	1	NM_024674	N_Shore
STAT4	0.004012254	2	NM_003151	Open sea
VWA1	0.004030376	1	NM_199121	N_Shore
KIF19	0.004042354	17	NM_153209	Island
GALNT14	0.004104764	2	NM_024572	Island
EMILIN3	0.004117449	20	NM_052846	Island
MYCBPAP	0.004143458	17	NM_032133	Island
TLX1	0.004147705	10	NM_005521	N_Shore
FBRSL1	0.004151893	12	NM_001142641	Island
GRM8	0.004163725	7	NM_001127323	N_Shore
ACRBP	0.004203739	12	NM_032489	Island
SEC31B	0.004263225	10	NM_015490	Island
MFSD2B	0.004276866	2	NM_001080473	Island
VRK2	0.00428493	2	NM_006296	Island
SLC22A16	0.004297289	6	NM_033125	Island
BIRC3	0.004300432	11	NM_182962	N_Shore
CALHM2	0.004334772	10	NM_015916	Island
ZMYND10	0.004338362	3	NM_015896	Island
PRDX1	0.004352991	1	NM_181696	Island
ALOX5	0.00440621	10	NM_000698	N_Shore
TRANK1	0.004407048	3	NM_014831	Island
BCAT1	0.004412162	12	NM_005504	Island
CCDC122	0.004421851	13	NM_144974	Island
TUBA4B	0.004447664	2	NR_003063	N_Shore
AOX1	0.004449394	2	NM_001159	Island

ANKRD34A	0.004457647	1	NM_001039888	Island
RADIL	0.004483495	7	NM_018059	Open sea
CYP27A1	0.004508567	2	NM_000784	Island
LBXCOR1	0.004508635	15	NM_001031807	Island
CYP27A1	0.004509907	2	NM_000784	N_Shore
PRSS27	0.004512352	16	NM_031948	N_Shore
NBL1	0.004513287	1	NM_005380	Island
PRSS27	0.004567002	16	NM_031948	Island
KIAA0040	0.004567514	1	NM_001162893	Island
VSX1	0.004597061	20	NM_199425	Island
COL24A1	0.004598473	1	NM_152890	Island
RNF135	0.004609756	17	NM_197939	N_Shore
SYT6	0.004609815	1	NM_205848	Island
CHAD	0.004629591	17	NM_001267	Island
ADPRH	0.004666175	3	NM_001125	Island
PLA2R1	0.004732611	2	NM_001007267	S_Shore
FLJ23834	0.004733441	7	NM_152750	S_Shore
STARD9	0.004741068	15	NM_020759	Island
ZC3HAV1L	0.004747218	7	NM_080660	Island
LBXCOR1	0.004748853	15	NM_001031807	Island
RAPGEFL1	0.004762787	17	NM_016339	Island
RASSF10	0.004824149	11	NM_001080521	Island
PTEN	0.00486326	10	NM_000314	N_Shore
GRASP	0.004865585	12	NM_181711	Island
NCAM2	0.004865729	21	NM_004540	Island
HLA-J	0.004867014	6	NR_024240	Island
ENC1	0.004890461	5	NM_003633	Island
CACNB4	0.004934731	2	NM_000726	Island
PXDN	0.004940753	2	NM_012293	Island

317

DOK7	0.004942283	4	NM_173660	Island
CCDC122	0.004948573	13	NM_144974	Island
WWTR1	0.00495493	3	NM_001168278	Island
VRK2	0.004956768	2	NM_006296	Island
RAP1GAP	0.004970622	1	NM_002885	Island
SEC31B	0.004986795	10	NM_015490	Island
PYCARD	0.005019774	16	NM_013258	Island
CNTD2	0.005035828	19	NM_024877	Island
CNGA3	0.005037631	2	NM_001298	Island
UPK3A	0.005039328	22	NM_006953	Island
GRIN2C	0.005040825	17	NM_000835	Island
RTBDN	0.005075345	19	NM_031429	S_Shore
RASL12	0.005080122	15	NM_016563	Island
NGB	0.005080711	14	NM_021257	Island
SHH	0.005082027	7	NM_000193	Island
CD2AP	0.005101133	6	NM_012120	N_Shore
PDPN	0.005103112	1	NM_001006624	Island
PDPN	0.005167789	1	NM_001006624	Island
DEDD2	0.005171406	19	NM_133328	Island
SWAP70	0.005199175	11	NM_015055	S_Shore
DLGAP3	0.00520215	1	NM_001080418	Island
PAK1	0.005214291	11	NM_002576	Island
SCGB3A1	0.005218745	5	NM_052863	Island
PRSS27	0.005231043	16	NM_031948	Island
ITGB5	0.005237953	3	NM_002213	Island
UPK3A	0.005238347	22	NM_006953	Island
GPR153	0.005254085	1	NM_207370	S_Shore
TLX1	0.005281648	10	NM_005521	Island
MT1E	0.005282497	16	NM_175617	Island

ERBB2	0.005288029	17	NM_001005862	S_Shore
PPP1R14A	0.005288706	19	NM_033256	Island
RPP25	0.005296227	15	NM_017793	N_Shore
GJA4	0.005367512	1	NM_002060	N_Shore
RNLS	0.005376423	10	NM_001031709	Island
XKR8	0.005418923	1	NM_018053	Island
RAB3D	0.005440149	19	NM_004283	Island
COL19A1	0.005440711	6	NM_001858	N_Shore
NFATC1	0.005536708	18	NM_006162	Island
DRD4	0.005556921	11	NM_000797	Island
COL12A1	0.005562696	6	NM_004370	Island
PPM1L	0.005567282	3	NM_139245	S_Shore
CNTD2	0.005610581	19	NM_024877	Island
EPAS1	0.005618624	2	NM_001430	Island
MT1E	0.00562495	16	NM_175617	S_Shore
HNRNPF	0.005666443	10	NM_001098207	Island
PPM1L	0.005690258	3	NM_139245	S_Shore
ANKS1B	0.005725518	12	NM_020140	Island
SULT1A1	0.005725875	16	NM_177536	S_Shore
GPR150	0.005736724	5	NM_199243	Island
CCDC122	0.005799884	13	NM_144974	Island
KLC2	0.005813253	11	NM_001134776	N_Shore
B3GNT7	0.005886979	2	NM_145236	Island
PDPN	0.005889797	1	NM_001006624	Island
FBRSL1	0.005896362	12	NM_001142641	Island
SLC35D3	0.005896742	6	NM_001008783	N_Shore
NFATC4	0.005929024	14	NM_001136022	N_Shore
AOX1	0.005936555	2	NM_001159	Island
ENC1	0.005936672	5	NM_003633	S_Shore

Gene	Value	Chr	Accession	Region
PHF11	0.005973904	13	NM_001040443	Island
AOX1	0.005986724	2	NM_001159	Island
HLA-J	0.006017269	6	NR_024240	Island
HSPA2	0.0061106	14	NM_021979	Island
LPAR2	0.006118951	19	NM_004720	Island
HSPB1	0.00616837	7	NM_001540	Island
CLDN10	0.006177309	13	NM_006984	Island
FBP1	0.006183594	9	NM_000507	Island
CD8A	0.006190793	2	NM_001145873	Island
USP4	0.00620579	3	NM_199443	S_Shore
PGCP	0.00620612	8	NM_016134	Island
PPM1L	0.006256048	3	NM_139245	S_Shore
ADCY1	0.006286994	7	NM_021116	Island
ALS2CL	0.006340027	3	NM_147129	Island
EPAS1	0.00634919	2	NM_001430	Island
KIAA0040	0.006398056	1	NM_001162893	S_Shore
TMEM144	0.006428801	4	NM_018342	Island
FAM5B	0.006471441	1	NM_021165	Island
MOSC2	0.006473556	1	NM_017898	Island
FLOT1	0.006490307	6	NM_005803	Island
MFSD2B	0.006523961	2	NM_001080473	Island
KLF5	0.006527989	13	NM_001730	S_Shore
CHRNB1	0.006530839	17	NM_000747	Island
EPAS1	0.006536629	2	NM_001430	Island
ENC1	0.006538363	5	NM_003633	Island
GABRA1	0.006584328	5	NM_001127645	Open sea
EFNB2	0.00664018	13	NM_004093	Island
IMPACT	0.006651793	18	NM_018439	Island
SHANK1	0.00668739	19	NM_016148	Island

Gene	Value	Chr	Accession	Region
HPCAL1	0.006772969	2	NM_002149	Island
RASL10B	0.006795269	17	NM_033315	Island
FLJ23834	0.006800997	7	NM_152750	Island
SEC31B	0.006805669	10	NM_015490	Island
RNLS	0.00683059	10	NM_001031709	Island
GDNF	0.006831621	5	NM_000514	Island
RNF135	0.006880849	17	NM_032322	Island
CHAD	0.006924727	17	NM_001267	Island
CHAD	0.006942221	17	NM_001267	Island
SYT6	0.006975512	1	NM_205848	Island
MT1A	0.006989134	16	NM_005946	Island
VILL	0.006996234	3	NM_015873	N_Shore
HOXC4	0.007025624	12	NM_014620	N_Shore
KCNQ4	0.007025724	1	NM_004700	Island
LIN28	0.00703373	1	NM_024674	N_Shore
COL9A3	0.007046562	20	NM_001853	Island
EFEMP2	0.007072463	11	NM_016938	Island
CXCR4	0.007102259	2	NM_003467	N_Shore
MLLT10	0.007113888	10	NM_004641	S_Shore
SIM2	0.007258543	21	NM_005069	Island
VRK2	0.007273479	2	NM_001130483	Island
FLOT1	0.007276459	6	NM_005803	Island
PTPRU	0.007330693	1	NM_133178	Island
RNF39	0.007392462	6	NM_025236	Island
HIST1H4L	0.007420721	6	NM_003546	S_Shore
CFTR	0.007422369	7	NM_000492	Open sea
HTR6	0.007430951	1	NM_000871	Island
GRB7	0.007449648	17	NM_005310	Open sea
SSH3	0.007558297	11	NM_017857	Island

NFATC1	0.007581332	18	NM_006162	Island
FGR	0.007604531	1	NM_005248	Island
HSD11B2	0.007679246	16	NM_000196	Island
TET1	0.007689163	10	NM_030625	S_Shore
SPAG1	0.007720732	8	NM_003114	N_Shore
CXCR4	0.007724038	2	NM_003467	N_Shore
VSX1	0.007725756	20	NM_014588	Island
FLJ40330	0.007737828	2	NR_015424	Island
WWTR1	0.007751772	3	NM_001168278	S_Shore
SCGB3A1	0.007767691	5	NM_052863	Island
RNLS	0.007797474	10	NM_001031709	Island
XKR8	0.007808935	1	NM_018053	Island
RAB32	0.00780926	6	NM_006834	Island
NPY5R	0.007831341	4	NM_006174	Island
TBR1	0.007835749	2	NM_006593	N_Shore
FAM59B	0.007845899	2	NM_001168241	Island
MFSD2B	0.007869988	2	NM_001080473	Island
USP4	0.007891386	3	NM_003363	Island
SHH	0.007893356	7	NM_000193	Island
RGNEF	0.007905393	5	NM_001080479	Island
PAMR1	0.007937377	11	NM_001001991	S_Shore
KIAA0146	0.007974009	8	NM_001080394	Open sea
FAM5B	0.008061346	1	NM_021165	Island
EYA4	0.008068786	6	NM_172105	Island
HFE	0.008068881	6	NM_139010	Open sea
COL21A1	0.008073495	6	NM_030820	Island
SPTBN1	0.008104432	2	NM_178313	Island
CRYAB	0.00821677	11	NM_001885	Open sea
FRZB	0.008228262	2	NM_001463	N_Shore

THBS4	0.008317269	5	NM_003248	Island
PPP1R14A	0.008442834	19	NM_033256	Island
MOSC2	0.008473765	1	NM_017898	Island
CNTN2	0.008481817	1	NM_005076	Open sea
EPS8L2	0.008484551	11	NM_022772	Island
HLA-J	0.00851929	6	NR_024240	Island
CTHRC1	0.008600562	8	NM_138455	S_Shore
TOM1L1	0.008656979	17	NM_005486	Island
GRID2IP	0.00867619	7	NM_001145118	Island
TMEM178	0.00869922	2	NM_152390	N_Shore
RAPGEFL1	0.008722953	17	NM_016339	Island
SERPINB6	0.008744454	6	NM_004568	Island
NR1I2	0.008760442	3	NM_033013	Island
MGST2	0.008792528	4	NM_002413	Open sea
LGR5	0.008816959	12	NM_003667	S_Shore
HSPB1	0.008881158	7	NM_001540	Island
SSH3	0.008901183	11	NM_017857	N_Shore
TMEM26	0.008928069	10	NM_178505	S_Shore
LPAR2	0.008945844	19	NM_004720	Island
FGR	0.008951717	1	NM_005248	Island
MFSD2B	0.008969807	2	NM_001080473	Island
IER3	0.008997907	6	NM_003897	Island
VAX2	0.009013311	2	NM_012476	Island
CHST8	0.009036995	19	NM_022467	Island
NR5A2	0.009037388	1	NM_205860	N_Shore
SLC22A15	0.009037778	1	NM_018420	Island
COL21A1	0.009049268	6	NM_030820	N_Shore
OSBPL1A	0.00907019	18	NM_080597	Island
SPATS1	0.00908249	6	NM_145026	Open sea

Gene	Value		RefSeq	Region
DLGAP3	0.009131369	1	NM_001080418	Island
LOXL3	0.009161581	2	NM_032603	N_Shore
SERINC2	0.009211597	1	NM_178865	Island
TTBK1	0.009211892	6	NM_032538	N_Shore
PDE4A	0.009220866	19	NM_001111308	Island
NR3C2	0.009242603	4	NM_000901	N_Shore
TMEM51	0.009242936	1	NM_001136218	Island
GPR19	0.009254774	12	NM_006143	Island
ARFGAP3	0.009320882	22	NM_014570	Island
P2RY2	0.009330767	11	NM_002564	Island
FAM59B	0.009331966	2	NM_001168241	Island
FGFRL1	0.00933691	4	NM_001004358	Island
MFSD2B	0.009366822	2	NM_001080473	Island
RAB32	0.009367517	6	NM_006834	Island
EFEMP2	0.009416179	11	NM_016938	Island
THBS4	0.009426863	5	NM_003248	Island
RNF39	0.009437857	6	NM_025236	Island
ME1	0.00945316	6	NM_002395	N_Shore
PAMR1	0.009482203	11	NM_015430	Island
SPAG1	0.00951145	8	NM_003114	N_Shore
ALS2CL	0.009533276	3	NM_147129	Island
MT1E	0.009755986	16	NM_175617	Island
KIAA0146	0.009799638	8	NM_001080394	Open sea
HLA-J	0.009851689	6	NR_024240	Island
ACTA1	0.009854989	1	NM_001100	Island
VPS37B	0.009947739	12	NM_024667	Island
COL1A2	0.009949633	7	NM_000089	Open sea
NAALAD2	0.009979418	11	NM_005467	Open sea
NFKBIZ	0.010034748	3	NM_001005474	Island

Gene	Value	Chr	Accession	Region
LZTS2	0.010079024	10	NM_032429	N_Shore
COL18A1	0.010098924	21	NM_130445	Island
TEKT3	0.010150287	17	NM_031898	S_Shore
RNF135	0.01015033	17	NM_197939	N_Shore
PACS2	0.010150688	14	NM_015197	Island
MT1E	0.010197556	16	NM_175617	Island
HHIPL1	0.01037231	14	NM_032425	S_Shore
ADAMTSL4	0.010391012	1	NM_019032	Island
RCCD1	0.010458274	15	NM_033544	Island
EPHA2	0.010469907	1	NM_004431	Island
MYL12A	0.010580165	18	NM_006471	Island
LGR5	0.010580565	12	NM_003667	S_Shore
KIAA0040	0.010598752	1	NM_001162893	Island
TMEM26	0.010656965	10	NM_178505	S_Shore
CABYR	0.010670367	18	NM_153768	Island
TRH	0.01069254	3	NM_007117	Island
LINGO3	0.010703433	19	NM_001101391	Island
LBXCOR1	0.010708694	15	NM_001031807	Island
SYCP2L	0.010709324	6	NM_001040274	N_Shore
LBXCOR1	0.01074912	15	NM_001031807	Island
FZD6	0.010851146	8	NM_003506	Island
TXNRD1	0.010874437	12	NM_182729	N_Shore
MLLT10	0.010893675	10	NM_004641	Island
SWAP70	0.01090179	11	NM_015055	Island
MMP14	0.010952897	14	NM_004995	Island
TRIM36	0.011000632	5	NM_018700	Island
RAB38	0.011002544	11	NM_022337	S_Shore
CD8A	0.011039345	2	NM_001145873	Island
CAMTA1	0.011051768	1	NM_015215	Open sea

OSBPL1A	0.011052691	18	NM_080597	S_Shore
LENG9	0.011053113	19	NM_198988	Island
LHX9	0.011099756	1	NM_001014434	Island
SCRN1	0.011115745	7	NM_014766	N_Shore
CD2AP	0.011134366	6	NM_012120	Island
NSUN7	0.011212405	4	NM_024677	Island
VILL	0.011279357	3	NM_015873	S_Shore
ME1	0.011311285	6	NM_002395	Island
ZNF575	0.01132956	19	NM_174945	S_Shore
TGFB2	0.011350774	1	NM_003238	Island
CXCR4	0.011373484	2	NM_003467	Island
CMTM7	0.011386783	3	NM_181472	Island
GRIN1	0.011447338	9	NM_000832	Island
TTBK1	0.011461455	6	NM_032538	N_Shore
CUL7	0.011508819	6	NM_001168370	Island
HLA-J	0.011516565	6	NR_024240	Island
GSX2	0.011518104	4	NM_133267	Island
MAST1	0.011689324	19	NM_014975	Island
ACRBP	0.011740179	12	NM_032489	Island
ZNF389	0.011769011	6	NM_001145129	Open sea
VWA1	0.011845145	1	NM_199121	Island
EPS8L2	0.011863541	11	NM_022772	Island
TUBA8	0.011874498	22	NM_018943	Island
TMEM22	0.01187812	3	NM_025246	Island
BRUNOL5	0.011977953	19	NM_021938	Island
SYCE2	0.012059447	19	NM_001105578	Island
LHX9	0.012070062	1	NM_020204	Island
OSBPL1A	0.01213845	18	NM_080597	Island
GALNT14	0.012162591	2	NM_024572	Island

NSUN7	0.012165968	4	NM_024677	Island
LBXCOR1	0.012288839	15	NM_001031807	Island
TLR2	0.012416751	4	NM_003264	Island
LTBP4	0.012481156	19	NM_001042545	Island
OGDHL	0.012518598	10	NM_018245	Island
PICK1	0.012553412	22	NM_001039583	Island
BCL2	0.012557292	18	NM_000657	Island
ACADS	0.012557547	12	NM_000017	Island
ZNF296	0.01262203	19	NM_145288	Island
MOSC2	0.012651971	1	NM_017898	Island
RARRES1	0.012665941	3	NM_206963	N_Shore
ALOX5	0.01270606	10	NM_000698	Island
WWTR1	0.012851172	3	NM_001168278	Island
NR3C2	0.012896216	4	NM_000901	N_Shore
TMEM51	0.012899152	1	NM_001136218	Island
CCNA1	0.012959251	13	NM_001111047	N_Shore
KCNIP3	0.012964434	2	NM_013434	Island
BIRC3	0.013017519	11	NM_182962	Island
NR3C2	0.013035843	4	NM_000901	N_Shore
FKBP5	0.013108302	6	NM_001145777	Island
ACADS	0.013125498	12	NM_000017	N_Shore
NTSR2	0.013147385	2	NM_012344	Island
SCRN1	0.013153927	7	NM_014766	N_Shore
GRB7	0.013175577	17	NM_001030002	Open sea
LOXL1	0.013190068	15	NM_005576	Island
FAM123C	0.013257618	2	NM_001105194	N_Shore
SST	0.01340295	3	NM_001048	N_Shore
CYP27A1	0.013403319	2	NM_000784	Island
MAST1	0.013437739	19	NM_014975	Island

HFE	0.013461392	6	NM_139010	Open sea
GPR19	0.013516055	12	NM_006143	Island
SLC35D3	0.013606884	6	NM_001008783	Island
CD302	0.013617459	2	NM_014880	S_Shore
BCL6	0.013637367	3	NM_001706	Island
MMP14	0.013642349	14	NM_004995	Island
LAMA3	0.013694904	18	NM_198129	Island
CHST8	0.013712362	19	NM_022467	Island
GIPC2	0.013739528	1	NM_017655	Island
HSPA2	0.013742167	14	NM_021979	Island
AOX1	0.013765728	2	NM_001159	Island
SLC12A8	0.013799611	3	NM_024628	S_Shore
ZNF3	0.01379979	7	NM_017715	S_Shore
MFSD2B	0.013839421	2	NM_001080473	Island
FGFRL1	0.013839911	4	NM_001004358	Island
RARRES1	0.013843252	3	NM_206963	Island
CYB5R2	0.013926358	11	NM_016229	Island
ZNF296	0.013937042	19	NM_145288	Island
KLHL1	0.013957804	13	NM_020866	Island
CHRNB1	0.013965699	17	NM_000747	Island
SERHL2	0.013970837	22	NM_014509	Island
SERHL2	0.013992254	22	NM_014509	N_Shore
STAT4	0.014015768	2	NM_003151	Open sea
CAMTA1	0.014292638	1	NM_015215	N_Shore
RGNEF	0.01434726	5	NM_001080479	Island
DKK1	0.014347293	10	NM_012242	N_Shore
EMILIN2	0.014361769	18	NM_032048	Island
VSX1	0.014368502	20	NM_199425	Island
ACTA1	0.01442471	1	NM_001100	Island

Gene	Value	Chr	Accession	Region
LOXL1	0.014425972	15	NM_005576	Island
RNLS	0.014443498	10	NM_001031709	Island
THBS1	0.014462304	15	NM_003246	Island
FLJ40330	0.014697374	2	NR_015424	S_Shore
RUSC1	0.014715358	1	NM_001105203	Island
IGSF21	0.014767259	1	NM_032880	N_Shore
SCRN1	0.014821969	7	NM_014766	Island
HNRNPF	0.014842882	10	NM_001098207	N_Shore
PICK1	0.014908869	22	NM_001039583	Island
KIAA0146	0.014912968	8	NM_001080394	Open sea
FZD6	0.01491404	8	NM_001164616	Island
LGALS8	0.014926042	1	NM_201544	Island
ONECUT1	0.014980162	15	NM_004498	Island
NR1I2	0.015048105	3	NM_033013	Island
DOK7	0.015048968	4	NM_173660	Island
KLC2	0.015095906	11	NM_001134776	N_Shore
EVI5	0.015112758	1	NM_005665	S_Shore
MYADM	0.015134761	19	NM_001020818	Island
CTSZ	0.015135047	20	NM_001336	Island
FBP1	0.015136043	9	NM_000507	Island
QRFPR	0.015137423	4	NM_198179	S_Shore
FAM123C	0.015164152	2	NM_001105194	N_Shore
CRYAB	0.015166706	11	NM_001885	Open sea
RARRES1	0.015167255	3	NM_206963	Island
SYT7	0.0151975	11	NM_004200	Island
TUBA8	0.015231272	22	NM_018943	Island
KIF5C	0.015252215	2	NM_004522	Island
CNTD2	0.015266675	19	NM_024877	Island
EVI5	0.015311134	1	NM_005665	S_Shore

MAST1	0.015313197	19	NM_014975	Island
SYCP2L	0.015342482	6	NM_001040274	N_Shore
LTBP3	0.015355493	11	NM_001164266	Island
RAB3D	0.015437932	19	NM_004283	Island
USP4	0.015598936	3	NM_199443	S_Shore
THBS4	0.015691629	5	NM_003248	Island
GABRA1	0.015718802	5	NM_001127645	Open sea
DOK7	0.01572927	4	NM_173660	Island
TUBA8	0.015833887	22	NM_018943	Island
TLX1	0.0158665	10	NM_005521	Island
LGALS8	0.015959512	1	NM_201544	Island
NAAA	0.015978811	4	NM_001042402	Island
P2RY2	0.015999723	11	NM_002564	Island
SLCO2A1	0.016023037	3	NM_005630	Island
MOGAT3	0.016055692	7	NM_178176	Island
VPS37B	0.016107147	12	NM_024667	Island
PHLDB1	0.016110534	11	NM_001144758	Island
LGR5	0.016139245	12	NM_003667	S_Shore
SFRP2	0.01621487	4	NM_003013	Island
GRIN2B	0.016215383	12	NM_000834	N_Shore
TRIP4	0.016215957	15	NM_016213	S_Shore
MYADM	0.016277339	19	NM_001020821	Island
CYB5R2	0.016322573	11	NM_016229	Island
DLC1	0.016468008	8	NM_182643	N_Shore
IGSF21	0.016476201	1	NM_032880	Island
SLC35D3	0.016486286	6	NM_001008783	Island
CUL7	0.016590317	6	NM_001168370	Island
NAAA	0.01666965	4	NM_001042402	Island
COL1A2	0.016685448	7	NM_000089	Open sea

ZNF3	0.016695968	7	NM_017715	S_Shore
ZMYND10	0.016781311	3	NM_015896	Island
RUSC1	0.016794402	1	NM_001105205	Island
TLR2	0.016803755	4	NM_003264	S_Shore
GIPR	0.016803862	19	NM_000164	Island
SWAP70	0.017030935	11	NM_015055	N_Shore
MCF2L2	0.017036003	3	NM_015078	N_Shore
PLEC1	0.01705783	8	NM_000445	Island
LMO7	0.017118508	13	NM_005358	S_Shore
PTPRN	0.017125554	2	NM_002846	Island
NPY2R	0.017242474	4	NM_000910	Island
SRRM3	0.017267652	7	NM_001110199	Island
AOX1	0.017376679	2	NM_001159	Island
PITX3	0.017393865	10	NM_005029	Island
SPTBN4	0.017401161	19	NM_025213	S_Shore
RAB33B	0.017453846	4	NM_031296	Island
MAP7	0.017463734	6	NM_003980	Island
EFEMP2	0.017580644	11	NM_016938	Island
MYADM	0.017604071	19	NM_001020818	Island
TNXB	0.017678436	6	NM_019105	Island
HNRNPF	0.017747848	10	NM_001098205	Island
ACADS	0.017763689	12	NM_000017	Island
RAP1GAP	0.017835144	1	NM_002885	Island
KCNIP3	0.017854327	2	NM_013434	Island
CRIP1	0.017868088	14	NM_001311	Island
HOXC4	0.017870978	12	NM_014620	N_Shore
MT1E	0.017888471	16	NM_175617	Island
NSUN7	0.017913085	4	NM_024677	Island
TEKT3	0.018013601	17	NM_031898	Island

DKK3	0.018028274	11	NM_001018057	S_Shore
NAALAD2	0.018029092	11	NM_005467	Open sea
KLF11	0.018061524	2	NM_003597	Island
NID2	0.018080985	14	NM_007361	Island
SFRP2	0.018084242	4	NM_003013	Island
DEDD2	0.018140911	19	NM_133328	Island
MYADM	0.018275176	19	NM_001020821	Island
CD109	0.018313843	6	NM_133493	N_Shore
ZMYND10	0.01846122	3	NM_015896	Island
RNF39	0.018505125	6	NM_025236	Island
PLLP	0.018585366	16	NM_015993	Island
SYCE2	0.01869128	19	NM_001105578	Island
UPK3A	0.018692227	22	NM_006953	Island
IER3	0.018722916	6	NM_003897	S_Shore
MT1A	0.01876313	16	NM_005946	Island
RAB33B	0.018809065	4	NM_031296	Island
HLA-J	0.018875396	6	NR_024240	Island
ATP5G2	0.018878388	12	NM_001002031	N_Shore
RNF39	0.018895143	6	NM_025236	Island
HLA-J	0.01889951	6	NR_024240	Island
PTPRN	0.019014863	2	NM_002846	Island
TMEM63A	0.019088085	1	NM_014698	Island
ZDHHC8P	0.019112683	22	NR_003950	Open sea
ZNF296	0.019309225	19	NM_145288	Island
GPR153	0.01934068	1	NM_207370	S_Shore
MCF2L2	0.019422509	3	NM_015078	S_Shore
LTBP4	0.019438339	19	NM_001042545	Island
NPY2R	0.019482173	4	NM_000910	Island
MYADM	0.019553752	19	NM_001020818	N_Shore

RNF135	0.019565041	17	NM_197939	N_Shore
HMSD	0.019610167	18	NM_001123366	Open sea
STL	0.019610621	6	NR_026876	Island
ZC3HAV1L	0.019692058	7	NM_080660	Island
CCNA1	0.019729939	13	NM_001111047	N_Shore
BMPR2	0.019737665	2	NM_001204	Island
DEDD2	0.019934959	19	NM_133328	Island
MCF2L2	0.019997167	3	NM_015078	Island
FCHSD1	0.020038895	5	NM_033449	N_Shore
MYADM	0.020111441	19	NM_001020818	Island
SYT6	0.020113169	1	NM_205848	N_Shore
MYADM	0.020115117	19	NM_001020818	Island
MYADM	0.020134158	19	NM_001020818	Island
MYL12A	0.020178622	18	NM_006471	Island
TOM1L1	0.020227635	17	NM_005486	N_Shore
MYADM	0.020274321	19	NM_001020818	Island
THBS1	0.020337745	15	NM_003246	Island
DEDD2	0.020488942	19	NM_133328	S_Shore
ELL3	0.020580789	15	NM_025165	S_Shore
MEGF10	0.020625545	5	NM_032446	Island
FAM160A1	0.02088686	4	NM_001109977	Island
MYADM	0.020972959	19	NM_001020818	Island
FAM59B	0.021019566	2	NM_001168241	Island
CRIP1	0.02107233	14	NM_001311	Island
IQSEC1	0.021151178	3	NM_001134382	S_Shore
RARRES1	0.021194227	3	NM_206963	Island
PHLDB1	0.02119447	11	NM_001144758	Island
LENG9	0.021276947	19	NM_198988	Island
ZNF296	0.021343579	19	NM_145288	Island

NAALAD2	0.021348928	11	NM_005467	Open sea
TMEM26	0.021404247	10	NM_178505	S_Shore
PRKCZ	0.02156063	1	NM_001033582	Island
LGALS8	0.021651744	1	NM_201544	Island
TUBA8	0.021703463	22	NM_018943	Island
RCCD1	0.021810415	15	NM_033544	Island
MYADM	0.021866911	19	NM_001020818	S_Shore
KIF19	0.021870862	17	NM_153209	Island
MYL12A	0.021907684	18	NM_006471	Island
DERL3	0.021980782	22	NM_001002862	Island
ELL3	0.022062262	15	NM_025165	Island
RAP1GAP	0.022087816	1	NM_001145658	Island
FAM160A1	0.022089445	4	NM_001109977	Island
PICK1	0.022229728	22	NM_001039583	N_Shore
CMTM7	0.02224274	3	NM_138410	Island
PTPRU	0.022272995	1	NM_133178	Island
UPK3A	0.022293832	22	NM_006953	Island
MYL12A	0.022303692	18	NM_006471	Island
LGALS8	0.022315416	1	NM_201544	Island
RAB27B	0.022439882	18	NM_004163	Island
CRIP1	0.02244055	14	NM_001311	Island
MFSD2B	0.022445027	2	NM_001080473	Island
UCP2	0.022507475	11	NM_003355	Island
LGALS1	0.02254589	22	NM_002305	N_Shore
DPYSL5	0.02258023	2	NM_020134	Island
CDK5R2	0.022588717	2	NM_003936	N_Shore
DEDD2	0.022617072	19	NM_133328	Island
LINGO3	0.02268067	19	NM_001101391	S_Shore
PTRF	0.022707145	17	NM_012232	Island

Gene	Value	Chr	Accession	Region
UNC13A	0.022729423	19	NM_001080421	Island
HCRT	0.022732905	17	NM_001524	Island
MEOX2	0.022737285	7	NM_005924	Open sea
LTBP3	0.02284279	11	NM_001164266	Island
NFATC4	0.022859738	14	NM_001136022	Island
PYCARD	0.022886261	16	NM_013258	N_Shore
USP2	0.02292712	11	NM_171997	Island
HCRT	0.022929918	17	NM_001524	Island
ZNF204P	0.02294259	6	NR_024553	Island
MLLT10	0.022955527	10	NM_001009569	N_Shore
SCN9A	0.023011394	2	NM_002977	S_Shore
ZC3HAV1L	0.023195996	7	NM_080660	Island
CCNA1	0.023230729	13	NM_001111047	N_Shore
CHD5	0.02323348	1	NM_015557	Island
PPP1CC	0.023251574	12	NM_002710	S_Shore
CENPV	0.023272656	17	NM_181716	Island
MCFD2	0.023402211	2	NM_001171508	Island
DEDD2	0.023507496	19	NM_133328	Island
HNRNPF	0.023588699	10	NM_001098207	Island
NAV2	0.023592599	11	NM_001111018	N_Shore
MCF2L2	0.023605007	3	NM_015078	N_Shore
PTRF	0.023616273	17	NM_012232	Island
KCNQ4	0.023621525	1	NM_004700	Island
MYADM	0.0237004	19	NM_001020818	Island
MYADM	0.023703489	19	NM_001020818	Island
PAX6	0.023709861	11	NM_001127612	Island
TRIP4	0.023800469	15	NM_016213	Island
KDELC2	0.023828231	11	NM_153705	Island
PICK1	0.023831896	22	NM_001039583	N_Shore

Gene	Value	Chr	Accession	Region
LGALS8	0.023862106	1	NM_201544	Island
ACTC1	0.02391709	15	NM_005159	Open sea
TMEM22	0.023951004	3	NM_001097599	Island
BCL6	0.023951921	3	NM_001706	Island
PLEC1	0.024067792	8	NM_000445	Island
ESAM	0.024100828	11	NM_138961	Island
AOX1	0.024135876	2	NM_001159	Island
CXCR4	0.024139288	2	NM_003467	Island
CHAD	0.024209057	17	NM_001267	Island
SCGB3A1	0.024217484	5	NM_052863	Island
COL9A2	0.024309777	1	NM_001852	Island
LMO7	0.024324092	13	NM_005358	Island
UCP2	0.024390746	11	NM_003355	Island
FAS	0.024393857	10	NR_028034	N_Shore
BCL6	0.024398273	3	NM_001706	Island
PLEC1	0.024408883	8	NM_000445	Island
TNXB	0.024414708	6	NM_019105	Open sea
HSPA2	0.02441604	14	NM_021979	N_Shore
HLCS	0.024464904	21	NM_000411	Island
EFEMP1	0.024544817	2	NM_001039349	Island
HPCA	0.024627859	1	NM_002143	Island
AOX1	0.024663987	2	NM_001159	Island
LTBP4	0.024696108	19	NM_001042545	Island
PAK1	0.024812177	11	NM_002576	Island
PICK1	0.024816763	22	NM_001039583	Island
FLJ22536	0.024824834	6	NR_015410	Island
FAS	0.02482765	10	NM_000043	Island
NEAT1	0.024833894	11	NR_028272	Island
STARD9	0.024859831	15	NM_020759	Island

LTBP4	0.024861683	19	NM_001042545	Island
SCN9A	0.024880906	2	NM_002977	Island
TGFBI	0.025041637	5	NM_000358	Island
CRIP1	0.02504291	14	NM_001311	Island
TCIRG1	0.025070439	11	NM_006019	Island
AOX1	0.025071967	2	NM_001159	N_Shore
PICK1	0.02508851	22	NM_001039583	Island
SPATS1	0.025111489	6	NM_145026	Open sea
SLC12A8	0.025112074	3	NM_024628	S_Shore
CCND2	0.025153821	12	NM_001759	Open sea
ONECUT1	0.025206395	15	NM_004498	Island
LENG9	0.025261285	19	NM_198988	Island
PTRF	0.025262356	17	NM_012232	Island
EVC2	0.025276757	4	NM_001166136	Island
TNXB	0.025301379	6	NM_019105	Open sea
CMTM7	0.025484689	3	NM_138410	Island
ZIC2	0.025609715	13	NM_007129	N_Shore
PHLDB1	0.025667054	11	NM_001144758	Island
NRXN1	0.025672898	2	NM_004801	S_Shore
ZNF788	0.025715285	19	NR_027049	Island
GRB7	0.025715921	17	NM_001030002	Open sea
RNLS	0.025748381	10	NM_018363	S_Shore
ADAM12	0.025770661	10	NM_003474	Island
IQSEC1	0.025800791	3	NM_001134382	S_Shore
DEDD2	0.025831217	19	NM_133328	Island
NSUN7	0.025884777	4	NM_024677	Island
KDELC2	0.025911538	11	NM_153705	Island
DKFZp761E1	0.025931192	11	NM_138368	Island
HPCAL1	0.025933387	2	NM_134421	Island

LRRC3B	0.025989116	3	NM_052953	Island
HLCS	0.02599245	21	NM_000411	Island
ZNF389	0.025995733	6	NM_001145129	Open sea
ZNF788	0.026015013	19	NR_027049	Island
ZNF788	0.026245465	19	NR_027049	Island
RPP25	0.026265018	15	NM_017793	Island
KIAA0495	0.02627697	1	NM_207306	Island
UCP2	0.026307852	11	NM_003355	Island
SYNPO2L	0.026366832	10	NM_024875	S_Shore
PAX6	0.026430185	11	NM_001127612	Island
CDK5R2	0.026437375	2	NM_003936	N_Shore
LY6G5C	0.026448068	6	NM_025262	N_Shore
CD302	0.026450819	2	NM_014880	N_Shore
SIAH3	0.026488059	13	NM_198849	Open sea
DDIT4L	0.026495839	4	NM_145244	Island
LGALS1	0.026587252	22	NM_002305	N_Shore
CYB5R1	0.026590601	1	NM_016243	Island
TTBK1	0.026642071	6	NM_032538	N_Shore
UPK3A	0.026648401	22	NM_006953	Island
RCCD1	0.02671361	15	NM_033544	Island
ALDH5A1	0.0268075	6	NM_001080	Island
NR2E1	0.026811657	6	NM_003269	Island
PAX6	0.026856485	11	NM_001127612	Island
SPATS1	0.02689941	6	NM_145026	Open sea
RPP25	0.026955326	15	NM_017793	Island
RUSC1	0.02697766	1	NM_001105205	Island
FAM5B	0.026982127	1	NM_021165	N_Shore
EVC2	0.026983323	4	NM_001166136	Island
RUSC1	0.02699564	1	NM_001105204	N_Shore

Gene	Value	Chr	Accession	Region
SEC31B	0.027001575	10	NM_015490	Island
TTBK1	0.027126621	6	NM_032538	N_Shore
SCGB3A1	0.027212999	5	NM_052863	Island
GPR150	0.027213048	5	NM_199243	Island
CHRNB1	0.027214267	17	NM_000747	Island
DKFZp761E1	0.027216974	11	NM_138368	Island
DKFZp761E1	0.027301975	11	NM_138368	Island
RILP	0.027343324	17	NM_031430	Island
LECT1	0.027373117	13	NM_001011705	N_Shore
GRB7	0.027377028	17	NM_001030002	Open sea
CD302	0.027404805	2	NM_014880	Island
PHLDB1	0.027406958	11	NM_001144758	Island
NAV2	0.027407124	11	NM_001111018	N_Shore
NAV1	0.027408705	1	NM_001167738	S_Shore
GRM8	0.027408941	7	NM_001127323	Island
CABYR	0.027410963	18	NM_153768	Island
FGFRL1	0.027457394	4	NM_001004358	Island
TMEM144	0.027459876	4	NM_018342	S_Shore
SCGB3A1	0.027473174	5	NM_052863	Island
CLDN10	0.027473532	13	NM_006984	Island
TRIM36	0.027564408	5	NM_018700	Island
LIN28	0.027565416	1	NM_024674	N_Shore
PICK1	0.027566687	22	NM_001039583	N_Shore
SLC25A20	0.027629015	3	NM_000387	Island
IMPACT	0.027658014	18	NM_018439	Island
KLF16	0.027743575	19	NM_031918	Island
GLB1L3	0.027882287	11	NM_001080407	Island
AEBP1	0.027904547	7	NM_001129	Open sea
GALNTL1	0.027904929	14	NM_001168368	Island

JAG1	0.027906836	20	NM_000214	Island
PLEC1	0.028021143	8	NM_000445	Island
MT1A	0.02802782	16	NM_005946	Island
TRANK1	0.028262931	3	NM_014831	Island
HFE	0.028279278	6	NM_139008	Open sea
RUSC1	0.028348334	1	NM_001105203	Island
MDK	0.028371855	11	NM_002391	Island
RCCD1	0.028378944	15	NM_033544	Island
PTRF	0.028444346	17	NM_012232	Island
CHAD	0.028447074	17	NM_001267	Island
SYNPO2L	0.028450753	10	NM_024875	Island
GPR150	0.028480274	5	NM_199243	Island
CFTR	0.02849952	7	NM_000492	Open sea
HNRNPF	0.028524828	10	NM_001098205	Island
CUL7	0.028551898	6	NM_001168370	Island
RILP	0.028740943	17	NM_031430	Island
JAG1	0.028793603	20	NM_000214	Island
PICK1	0.028803794	22	NM_001039583	Island
KDELC2	0.028813336	11	NM_153705	Island
COL9A2	0.02885032	1	NM_001852	Island
ADAM12	0.028863564	10	NM_003474	Island
NEUROG1	0.028886246	5	NM_006161	Island
HNRNPF	0.028911938	10	NM_001098205	S_Shore
ANXA2	0.028994013	15	NM_001002857	Island
TCIRG1	0.029067492	11	NM_006019	Island
TMEM178	0.029073518	2	NM_152390	N_Shore
SYNPO2L	0.029156949	10	NM_024875	Island
SORBS3	0.029292157	8	NM_005775	Island
TGFBI	0.029380056	5	NM_000358	Island

NMRAL1	0.029531341	16	NM_020677	N_Shore
NEAT1	0.029714679	11	NR_028272	Island
SPATS1	0.029826171	6	NM_145026	Open sea
CD109	0.029830197	6	NM_001159588	N_Shore
RGNEF	0.029900267	5	NM_001080479	Island
PLD5	0.029928573	1	NM_152666	Island
LGALS1	0.029990469	22	NM_002305	Island
SLC13A5	0.030045868	17	NM_001143838	Island
MIR548H4	0.030193691	15	NR_031680	Island
SPIB	0.03021723	19	NM_003121	Open sea
CCNA1	0.030394976	13	NM_001111047	N_Shore
PRSS27	0.030536678	16	NM_031948	N_Shore
SPATS1	0.030737483	6	NM_145026	Open sea
ZNF788	0.030917919	19	NR_027049	Island
SLC22A3	0.031137764	6	NM_021977	Island
TGFBI	0.031227664	5	NM_000358	Island
TRIP6	0.031415569	7	NM_003302	S_Shore
NBL1	0.031431515	1	NM_182744	Island
DKFZp761E1	0.03165343	11	NM_138368	Island
ONECUT2	0.031714847	18	NM_004852	Island
RNF39	0.031717533	6	NM_025236	Island
SORBS3	0.03177639	8	NM_005775	N_Shelf
TCTEX1D1	0.031780281	1	NM_152665	N_Shore
ESRP2	0.031837637	16	NM_024939	Island
CCNA1	0.031842484	13	NM_001111047	N_Shore
NBL1	0.031905781	1	NM_005380	Island
MIR548H4	0.031924981	15	NR_031680	Island
SORBS3	0.03195139	8	NM_005775	Island
IMPACT	0.032018447	18	NM_018439	Island

Gene	Value	Chr	Accession	Location
TRIM36	0.032155973	5	NM_018700	Island
BNC1	0.032168949	15	NM_001717	Island
SLC12A8	0.032169322	3	NM_024628	Island
RASSF10	0.032218303	11	NM_001080521	Island
DKFZp761E1	0.032218539	11	NM_138368	Island
EVI5	0.032292705	1	NM_005665	S_Shore
GABRA1	0.032293273	5	NM_000806	Open sea
CAMK2B	0.032364052	7	NM_172081	Island
KDM2B	0.032392007	12	NM_032590	S_Shore
USP2	0.032409213	11	NM_171997	Island
STARD9	0.032411859	15	NM_020759	Island
B3GALT4	0.032418275	6	NM_003782	Island
SYNPO2L	0.032568048	10	NM_024875	Island
MAP1LC3A	0.032749416	20	NM_032514	Island
AEBP1	0.032902132	7	NM_001129	Open sea
MYL12A	0.033038018	18	NM_006471	Island
CHD5	0.03304102	1	NM_015557	Island
FLJ22536	0.033145882	6	NR_015410	Island
NRXN1	0.033181334	2	NM_004801	Island
NEAT1	0.03318839	11	NR_028272	Island
FGR	0.033427338	1	NM_005248	Island
TRIP6	0.033508793	7	NM_003302	S_Shore
SIAH3	0.033543072	13	NM_198849	Open sea
PHF11	0.03358628	13	NM_001040444	N_Shore
CTHRC1	0.033637258	8	NM_138455	Island
DKK3	0.033637576	11	NM_001018057	Island
TGFBI	0.033638199	5	NM_000358	Island
OSBPL9	0.033824281	1	NM_148904	Island
CXCR4	0.033998966	2	NM_003467	N_Shore

LGALS1	0.03402873	22	NM_002305	N_Shore
FAM188B	0.034097753	7	NM_032222	Island
KDM2B	0.034154004	12	NM_032590	S_Shore
THBS1	0.03415467	15	NM_003246	Island
TCTEX1D1	0.034155791	1	NM_152665	Island
FAS	0.034243462	10	NM_000043	Island
NAV1	0.034366691	1	NM_001167738	Island
NMRAL1	0.034377427	16	NM_020677	Island
CYB5R2	0.034405481	11	NM_016229	Island
LY75	0.03441098	2	NM_002349	Island
FGFRL1	0.034427509	4	NM_001004358	Island
LHX9	0.034455877	1	NM_001014434	Island
PHLDB1	0.034486785	11	NM_001144758	Island
THBS1	0.034775632	15	NM_003246	Island
XKR8	0.034775821	1	NM_018053	Island
DRD4	0.034866531	11	NM_000797	Island
MYRIP	0.034960823	3	NM_015460	Island
FAM188B	0.03496466	7	NM_032222	Island
KDM2B	0.035066555	12	NM_032590	S_Shore
NPHS2	0.035084226	1	NM_014625	Island
NMRAL1	0.03509607	16	NM_020677	Island
THBS1	0.035224454	15	NM_003246	Island
HLA-H	0.035239402	6	NR_001434	Island
ACTC1	0.035294134	15	NM_005159	Open sea
PHLDB1	0.035346332	11	NM_001144758	Island
CENPV	0.035355148	17	NM_181716	Island
CCND2	0.035385599	12	NM_001759	Island
SORBS3	0.035418958	8	NM_005775	Island
ZNF3	0.035484171	7	NM_017715	S_Shore

KLHL1	0.035608103	13	NM_020866	S_Shore
LHX9	0.035652361	1	NM_001014434	S_Shore
NID2	0.035761759	14	NM_007361	Island
CCND2	0.035862423	12	NM_001759	Island
FAM123C	0.035869237	2	NM_001105194	N_Shore
CHD5	0.035904924	1	NM_015557	Island
DDIT4L	0.035907052	4	NM_145244	Island
SORBS3	0.035913079	8	NM_005775	S_Shore
OSBPL1A	0.036136719	18	NM_080597	Island
HPCA	0.036178171	1	NM_002143	N_Shore
AOX1	0.036196094	2	NM_001159	Island
NPPB	0.036241266	1	NM_002521	S_Shore
SPIB	0.03636725	19	NM_003121	Open sea
B3GALT4	0.036425238	6	NM_003782	Island
MT1E	0.036667606	16	NM_175617	Island
NMRAL1	0.036746357	16	NM_020677	Island
ACVRL1	0.036849822	12	NM_000020	Island
FLJ22536	0.037047991	6	NR_015410	Island
NAAA	0.037059317	4	NM_014435	Island
MYADM	0.037135149	19	NM_001020818	Island
KDM2B	0.037236938	12	NM_032590	S_Shore
RAB11FIP4	0.037279554	17	NM_032932	S_Shelf
CAMK2B	0.037285482	7	NM_172081	Island
MAP7	0.037293336	6	NM_003980	Island
NAV1	0.037447763	1	NM_001167738	Island
VSIG10L	0.037462042	19	NM_001163922	Island
TRIP6	0.03750223	7	NM_003302	S_Shore
ZIC2	0.037524247	13	NM_007129	N_Shore
LINGO3	0.037620714	19	NM_001101391	Island

Gene	Value	Chr	Accession	Region
KDELC2	0.037653929	11	NM_153705	Island
MCF2L2	0.037729669	3	NM_015078	N_Shore
ACTC1	0.037858394	15	NM_005159	Open sea
RPP25	0.03793694	15	NM_017793	Island
COL9A2	0.038166417	1	NM_001852	Island
SYNPO2L	0.038224847	10	NM_024875	S_Shore
IQSEC1	0.038251409	3	NM_001134382	Island
MAP1LC3A	0.038322076	20	NM_032514	Island
DLC1	0.038332005	8	NM_006094	Island
CAMK2B	0.038343873	7	NM_172081	Island
MOGAT3	0.038607218	7	NM_178176	Island
NEAT1	0.038681251	11	NR_028272	S_Shelf
TRH	0.03877206	3	NM_007117	Island
SIAH3	0.038931732	13	NM_198849	Open sea
OSBPL9	0.039138539	1	NM_148904	Island
KCNJ3	0.039169498	2	NM_002239	Island
RAB38	0.039251347	11	NM_022337	S_Shore
WRB	0.039305287	21	NM_004627	N_Shore
MAP1LC3A	0.03935296	20	NM_032514	Island
AOX1	0.039491094	2	NM_001159	Island
FBRSL1	0.039633642	12	NM_001142641	Island
PLEC1	0.039637477	8	NM_000445	Island
ADCY1	0.039726848	7	NM_021116	Island
TRANK1	0.039897182	3	NM_014831	Island
NSD1	0.040112796	5	NM_022455	Island
TRIP4	0.040142003	15	NM_016213	Island
WRB	0.040190242	21	NM_004627	N_Shore
LY75	0.040191949	2	NM_002349	Island
IQSEC1	0.04019227	3	NM_001134382	S_Shore

NPPB	0.040519383	1	NM_002521	Island
NID2	0.040611738	14	NM_007361	Island
NPY5R	0.040626362	4	NM_006174	Island
FBRSL1	0.040667161	12	NM_001142641	Island
MAL	0.040684253	2	NM_022439	Island
GALNTL1	0.040690044	14	NM_020692	Island
TCTEX1D1	0.040763342	1	NM_152665	Island
FLJ40330	0.040982653	2	NR_015424	Island
RBM24	0.041101742	6	NM_001143942	Island
SEC31B	0.041323044	10	NM_015490	S_Shore
KLF11	0.041458707	2	NM_003597	Island
LHX9	0.041521249	1	NM_001014434	Island
NAALAD2	0.041521309	11	NM_005467	Open sea
VWA1	0.041637246	1	NM_022834	Island
SLCO2A1	0.041683258	3	NM_005630	Island
KIF19	0.041798198	17	NM_153209	Island
CHRNB1	0.041828656	17	NM_000747	Island
NEUROG1	0.04196653	5	NM_006161	Island
PLD5	0.042021741	1	NM_152666	Island
CYB5R1	0.042067815	1	NM_016243	S_Shore
GIPR	0.042093969	19	NM_000164	Island
RNF39	0.042126057	6	NM_170769	Island
ADCY1	0.042175649	7	NM_021116	Island
HOXC4	0.042355927	12	NM_014620	N_Shore
ZEB1	0.042372362	10	NM_030751	Island
NID2	0.042400258	14	NM_007361	Island
WRB	0.04247073	21	NM_004627	Island
FLJ22536	0.042634262	6	NR_015410	Island
TUBA8	0.04273971	22	NM_018943	Island

NBL1	0.042821548	1	NM_182744	Island
SORBS3	0.042828359	8	NM_005775	Island
SLCO2A1	0.042950676	3	NM_005630	S_Shore
VPS37B	0.043024986	12	NM_024667	Island
SULT1A1	0.043081365	16	NM_177536	Island
TUBA8	0.043202197	22	NM_018943	Island
NEUROG1	0.043236593	5	NM_006161	Island
LMO7	0.043277508	13	NM_005358	Island
ZNF389	0.043341043	6	NM_001145129	Open sea
PSD3	0.043583917	8	NM_206909	Open sea
FAM188B	0.043880862	7	NM_032222	N_Shore
NAV1	0.04393138	1	NM_001167738	Island
LY6G5C	0.043938139	6	NM_025262	Island
NR5A2	0.0440449	1	NM_205860	Island
XKR8	0.044164379	1	NM_018053	S_Shore
NEUROG1	0.044183502	5	NM_006161	Island
TEKT3	0.04420307	17	NM_031898	Island
RCCD1	0.044204031	15	NM_033544	Island
RCCD1	0.044328602	15	NM_033544	N_Shore
CTHRC1	0.044728997	8	NM_138455	Island
VWDE	0.044788631	7	NM_001135924	Island
THBS1	0.044805988	15	NM_003246	Island
FAM123C	0.044820048	2	NM_001105194	N_Shore
LMO7	0.044986377	13	NM_005358	Island
NR5A2	0.044992077	1	NM_205860	Island
TLR2	0.045007297	4	NM_003264	Island
SPAG1	0.045486872	8	NM_003114	Island
RNF39	0.045555848	6	NM_170769	Island
TOM1L1	0.045568219	17	NM_005486	Island

Gene	Value	Chr	Accession	Region
LY6G5C	0.045608223	6	NM_025262	Island
FRZB	0.04571933	2	NM_001463	Island
TMED7-	0.045720937	5	NM_001164468	Island
SYT6	0.045783744	1	NM_205848	Island
CXCR4	0.045838783	2	NM_003467	N_Shore
CHAD	0.045940318	17	NM_001267	Island
PTRF	0.045940391	17	NM_012232	Island
NID2	0.046117715	14	NM_007361	Island
LMO2	0.046205901	11	NM_001142315	Island
WRB	0.046527858	21	NM_004627	Island
NEAT1	0.047189624	11	NR_028272	Island
ADAMTSL4	0.047301952	1	NM_025008	Island
LRAT	0.047306593	4	NM_004744	Island
CCNA1	0.047310401	13	NM_001111047	N_Shore
JAG1	0.047409865	20	NM_000214	Island
RPP25	0.047442861	15	NM_017793	Island
TRH	0.04758056	3	NM_007117	Island
NUDT16	0.048044609	3	NM_152395	Island
PLA2R1	0.048051855	2	NM_007366	N_Shore
COL9A2	0.048063926	1	NM_001852	Island
COL1A2	0.048078848	7	NM_000089	Open sea
CCND2	0.048080104	12	NM_001759	Island
PITX3	0.048102464	10	NM_005029	Island
B3GALT4	0.048102772	6	NM_003782	Island
USP2	0.04834295	11	NM_171997	Island
ACADS	0.049037697	12	NM_000017	Island
COL9A2	0.049343116	1	NM_001852	Island
FZD6	0.049364781	8	NM_001164616	Island
TGFBI	0.049437455	5	NM_000358	Island

MYADM	0.049737974	19	NM_001020818	Island
SLC25A20	0.049846704	3	NM_000387	Island
BNC1	0.049932515	15	NM_001717	Island

7.4 Cancer genes associated with the 535 CIMP gene list

Category	Functions Annotation	p-Value	Molecules	# Molecules
Cancer	metastasis of cancer cells	<0.0001	CBR1,COL18A1,CXCR4,NID2,PTEN,THRB	6
Cancer	metastasis	<0.0001	BMPR2,CBR1,CCNA1,CFLAR,COL18A1,CXCR4,DKK1,EFEMP1,EPAS1,EPHA2, ERBB2,FABP5,FGR,FOXP1,GJA1,GLUL,JAG1,LAMA3,LGALS1,LGALS3,LGALS8,MMP14, NID2,NR4A1,PTEN,SHH,SLC12A8,SYNPO2L,TGFB2,THBS1,THRB,TLX1,TUBA8,ZEB1	35
Cancer	development of tumor	0.0001	ALOX5,ANXA2,BCL2,BCL6,COL18A1,CXCR4,EFEMP1,EPHA2,ERBB2,FAS,GDNF,IER3, ITGB5,KLF5,LGALS1,MDK,MMP14,PTEN,RUNX1,THBS1,WWTR1	21
Cancer	carcinoma	0.0001	ACADS,ACTC1,ACTG1,ADAM12,ADAM8,ALDH3A1,ALDH5A1,ALOX5,ANXA2,APBA2,BCL2, BMPR2,BNC1,CABYR,CCNA1,CCND2,CDH3,CDK5R2,CFTR,CHRNB1,CHST8,CLDN5, CLSTN1,COL18A1,COL1A2,CRIP1,CXCR4,DKK1,DLC1,EBF4,EFEMP1,EFNB2,ENC1, EPAS1,EPHA2,ERBB2,FABP5,FAS,FRZB,GDNF,GJA1,GLUL,GNMT,GRB7,GRIN2B,HDAC4, HIST1H4A,HLCS,HLF,HOXC4,HSD11B2,IER3,ITGB5,JAG1,KCNB1,KLF5,LAMA3,LAMP3, LGALS1,LGALS3,LGALS8,LIN28A,LMO7,LPCAT2,LTBP4,LYN,MAL,MAPT,MEOX2, MIR155HG,MMP14,MT1E,MYL12A,NAAA,NEUROG1,NFATC1,NR5A2,OPCML,PAK1,PDPN ,PLEC,PLLP,PPM1L,PRDX1,PRKCZ,PTEN,PXDN,PYCARD,RAP1GAP,RARRES1,RASL12, RILP,SFRP2,SFRP4,SGK1,SHH,SIM2,SLC12A8,SOCS3,SPIB,SPINT2,SPTBN1, SULT1A1,SYNPO2L,TGFB2,TGFBI,THBS1,THBS4,THRB,TLR2,TMEM200B,TNXB,TOM1L1, TUBA8,USP2,USP4,VRK2,WWTR1,ZIC2	121
Cancer	epithelial tumor	0.0003	ACADS,ACTC1,ACTG1,ADAM12,ADAM8,AEBP1,ALDH3A1,ALDH5A1,ALOX5,ANGPT4, ANXA2,APBA2,BCL2,BMPR2,BNC1,CABYR,CCNA1,CCND2,CDH3,CDK5R2,CFTR,CHRNB1, CHST8,CLDN5,CLSTN1,COL18A1,COL1A2,CRIP1,CXCR4,DKK1,DLC1,EBF4,EFEMP1, EFNB2,ENC1,EPAS1,EPHA2,ERBB2,FABP5,FAS,FRZB,GDNF,GJA1,GLUL,GNMT,GRB7, GRIN2B,HDAC4,HIST1H4A,HLCS,HLF,HOXC4,HSD11B2,IER3,ITGB5,JAG1,KCNB1,KLF5, LAMA3,LAMP3,LGALS1,LGALS3,LGALS8,LIN28A,LMO7,LPCAT2,LTBP4,LYN,MAL,MAPT, MEOX2,MIR155HG,MMP14,MT1E,MYL12A,NAAA,NEUROG1,NFATC1,NR3C2,NR5A2, OPCML,PAK1,PDPN,PLEC,PLLP,PPM1L,PRDX1,PRKCZ,PTEN,PXDN,PYCARD,RAP1GAP, RARRES1,RARRES2,RASL12,RILP,SFRP2,SFRP4,SGK1,SHH,SIM2,SLC12A8,SLC22A3, SOCS3,SPIB,SPINT2,SPTBN1,SULT1A1,SYNPO2L,TGFB2,TGFBI,THBS1,THBS4,THRB, TLR2,TMEM200B,TNXB,TOM1L1,TUBA8,USP2,USP4,VRK2,WWTR1,ZEB1,ZIC2	125
Cancer	solid tumor	0.0003	ACADS,ACTC1,ACTG1,ADAM12,ADAM8,ALDH3A1,ALDH5A1,ALOX5,ANXA2,APBA2,BCL2, BMPR2,BNC1,CABYR,CCNA1,CCND2,CDH3,CDK5R2,CFTR,CHRNB1,CHST8,CLDN5, CLSTN1,COL18A1,COL1A2,CRIP1,CXCR4,DKK1,DLC1,EBF4,EFEMP1,EFNB2,ENC1, EPAS1,EPHA2,ERBB2,FABP5,FAS,FRZB,GDNF,GJA1,GLUL,GNMT,GRB7,GRIN2B,HDAC4, HIST1H4A,HLCS,HLF,HOXC4,HSD11B2,IER3,ITGB5,JAG1,KCNB1,KLF5,LAMA3,LAMP3, LGALS1,LGALS3,LGALS8,LIN28A,LMO7,LPCAT2,LTBP4,LYN,MAL,MAPT,MEOX2, MIR155HG,MMP14,MT1E,MYL12A,NAAA,NEUROG1,NFATC1,NR5A2,OPCML,PAK1,PDPN, PLEC,PLLP,PPM1L,PRDX1,PRKCZ,PTEN,PXDN,PYCARD,RAP1GAP,RARRES1,RASL12, RILP,RUNX1,SFRP2,SFRP4,SGK1,SHH,SIM2,SLC12A8,SLC22A3,SOCS3,SPIB,SPINT2, SPTBN1,SULT1A1,SYNPO2L,TGFB2,TGFBI,THBS1,THBS4,THRB,TLR2,TMEM200B,TNXB, TOM1L1,TUBA8,USP2,USP4,VRK2,WWTR1,ZEB1,ZIC2	122
Cancer	metastasis of tumor cells	0.0003	CBR1,COL18A1,CXCR4,NID2,PTEN,THRB,ZEB1	7
Cancer	angiogenesis of tumor	0.0003	ANXA2,BCL2,COL18A1,EPHA2,ITGB5,KLF5,LGALS1,MMP14,THBS1	9

Category	Functions Annotation	p-Value	Molecules	# Molecules
Cancer	head and neck tumor	0.0014	ADAM12,ADAM8,AEBP1,ALOX5,ANGPT4,ANXA2,BCL2,CCND2,CHRNB1,CXCR4,EFNB2,ENC1,EPHA2,ERBB2,FABP5,FRZB,GJA1,KLF5,LGALS1,LGALS3,LYN,MMP14,PDPN,PLA2R1,PTEN,PYCARD,RAP1GAP,SFRP2,SFRP4,SHH,SOCS3,THBS1,THRB,TLR2,TUBA8,VRK2	36
Cancer	bladder cancer	0.0019	BCL2,BIRC3,CCND2,CHRNB1,ERBB2,FAS,KDM2B,MBP,PTEN,RBP1,THBS1	11
Cancer	neoplasia of breast cancer cell lines	0.0019	CXCR4,DLC1,EPAS1,ERBB2,FABP5,LOXL1	6
Cancer	non-Hodgkin's disease	0.0021	BCL2,BCL6,BIRC3,CFLAR,COL12A1,CXCR4,EYA4,FAS,HDAC4,IER3,NR1I2,PRDX1,PRKCZ,PTEN,SGK1,SHH,TLR2,TUBA8,ZEB1	19
Cancer	cancer of secretory structure	0.0023	ACTC1,ANXA2,BCL2,BMPR2,CCNA1,CDH3,CLSTN1,CXCR4,ERBB2,FAS,HOXC4,IER3,LGALS8,mir375,MMP14,PDLIM4,PRKCZ,PTEN,PYCARD,RARRES1,SGK1,SHH,SIM2,SLC22A3,SOCS3,TLR2,TUBA8,USP2,ZFHX3	29
Cancer	endometrial carcinoma	0.0024	ANXA2,BCL2,CLDN5,DKK1,ERBB2,FRZB,GRB7,HLF,KCNB1,LAMP3,LTBP4,MAL,MEOX2,PLLP,PTEN,RASL12,SFRP4,SHH,SPINT2,SULT1A1,THBS4,TMEM200B,TNXB,TUBA8,ZEB1	25
Cancer	lymphomagenesis	0.0024	BCL2,BCL6,BIRC3,CD8A,CFLAR,COL12A1,CXCR4,EYA4,FAS,HDAC4,IER3,NR1I2,PRDX1,PRKCZ,PTEN,RUNX1,SGK1,SHH,SPTBN1,TGFBI,THBS1,TLR2,TLX1,TUBA8,ZEB1	25
Cancer	tumorigenesis of mammary adenocarcinoma	0.0025	ERBB2,PRDX1,PTEN,THBS1	4
Cancer	thyroid cancer	0.0026	ALOX5,CCND2,CXCR4,EFNB2,ERBB2,LGALS3,PLA2R1,PTEN,RAP1GAP,SFRP2,SOCS3,THBS1,THRB	13
Cancer	germ cell tumor	0.0029	BCL2,GDNF,LOXL3,PDPN,PTEN,TUBA8	6
Cancer	prostatic carcinoma	0.0032	ACTC1,ANXA2,BCL2,CDH3,CLSTN1,ERBB2,FAS,HOXC4,IER3,PTEN,PYCARD,RARRES1,SGK1,SIM2,SLC22A3,SOCS3,USP2	17
Cancer	metastasis of lung	0.0035	EPAS1,ERBB2,NID2,PTEN,THRB	5
Cancer	tumorigenesis of hepatocellular carcinoma	0.0039	BCL2,GNMT,PRDX1,PTEN,SOCS3,SPTBN1	6
Cancer	colorectal tumor	0.0040	ACADS,ADAM12,APBA2,BCL2,CCND2,CLSTN1,CRIP1,CXCR4,DKK1,DLC1,EBF4,EFEMP1,ENC1,EPAS1,ERBB2,FAS,GJA1,GLUL,HIST1H4A,HLCS,HSD11B2,JAG1,LAMA3,LGALS3,LGALS8,LPCAT2,LTBP4,MT1E,NAAA,NEUROG1,PAK1,PDPN,PLEC,PPM1L,PTEN,RAP1GAP,RILP,SFRP4,SGK1,SLC12A8,SPIB,SYNPO2L,TGFB2,THBS4,TLR2,USP4	46
Cancer	prostate cancer	0.0042	ACTC1,ANXA2,BCL2,BMPR2,CCNA1,CDH3,CLSTN1,CXCR4,ERBB2,FAS,HOXC4,IER3,LGALS8,mir375,PDLIM4,PRKCZ,PTEN,PYCARD,RARRES1,SGK1,SHH,SIM2,SLC22A3,SOCS3,TLR2,TUBA8,USP2,ZFHX3	28
Cancer	hyperplasia of alveolar epithelium	0.0045	ERBB2,PTEN	2
Cancer	invasion of mammary tumor	0.0045	ERBB2,NFATC1	2
Cancer	metastasis of melanoma cells	0.0045	COL18A1,NID2	2
Cancer	onset of tumorigenesis of mammary tumor	0.0045	ERBB2,ITGB5	2
Cancer	tumorigenesis of salivary gland tumor	0.0045	ERBB2,PTEN	2

Category	Functions Annotation	p-Value	Molecules	# Molecules
Cancer	Cancer	0.0003	ACAA1,ACADS,ACTC1,ACTG1,ADAM12,ADAM8,AEBP1,ALDH3A1,ALDH5A1,ALOX5, ANGPT4,ANXA2,APBA2,BCAT1,BCL2,BCL6,BIRC3,BMPR2,BNC1,CABYR,CBR1,CCNA1, CCND2,CD8A,CDH3,CDK5R2,CFLAR,CFTR,CHRNB1,CHST8,CLDN5,CLSTN1,COL12A1, COL18A1,COL1A2,CRIP1,CRYAB,CXCR4,DKK1,DLC1,EBF4,EFEMP1,EFNB2,ENC1,EPAS1, EPHA2,ERBB2,EYA4,FABP5,FAS,FGR,FOXP1,FRZB,GDNF,GJA1,GLUL,GNMT,GRB7, GRIN1,GRIN2B,GRIN2C,HDAC4,HIST1H4A,HLCS,HLF,HOXC4,HSD11B2,HSPB1,IER3,ITGB5, JAG1,KCNB1,KDM2B,KLF11,KLF5,LAMA3,LAMP3,LGALS1,LGALS8,LIN28A,LMO2, LMO7,LOX3,LPCAT2,LTBP4,LYN,MAL,MAPT,MBP,MDK,MEOX2,MGST2,mir375,MIR155HG ,MLLT10,MMP14,MT1E,MYL12A,NAAA,NEUROD1,NEUROG1,NFATC1,NID2,NR1I2,NR4A1, NR5A2,NSD1,OPCML,PAK1,PDE4A,PDLIM4,PDPN,PLA2R1,PLEC,PLLP,PPM1L,PRDX1, PRKCZ,PSD3,PTEN,PTRF,PXDN,PYCARD,RAP1GAP,RARRES1,RASL12,RBP1,RILP, RUNX1,SFRP2,SFRP4,SGK1,SHH,SIAH1,SIM2,SLC12A8,SLC22A3,SOCS3,SPIB,SPINT2, SPTBN1,SULT1A1,SYNPO2L,TGFB2,TGFBI,THBS1,THBS4,THRB,TLR2,TLX1,TMEM200B, TNX8,TOM1L1,TUBA8,USP2,USP4,VRK2,WWTR1,ZEB1,ZFHX3,ZIC2	162
Cancer	endocrine gland tumor	0.0004	AEBP1,ALOX5,ANXA2,BCL2,BMPR2,CCND2,CDH3,CDK5R2,CFTR,CHRNB1,CXCR4,DKK1, EFNB2,EPHA2,ERBB2,FRZB,GDNF,LGALS1,LGALS3,MIR155HG,MMP14,NR5A2,PLA2R1, PTEN,PXDN,RAP1GAP,SFRP2,SFRP4,SHH,SOCS3,THBS1,THRB	32
Cancer	digestive organ tumor	0.0004	ACADS,ACTG1,ADAM12,ALOX5,ANXA2,APBA2,BCL2,CABYR,CCND2,CDK5R2,CFTR ,CHRNB1,CLSTN1,COL18A1,COL1A2,CRIP1,CXCR4,DKK1,DLC1,EBF4,EFEMP1,ENC1, EPAS1,EPHA2,ERBB2,FAS,GJA1,GLUL,GNMT,GRIN2B,HIST1H4A,HLCS,HSD11B2,JAG1, KLF5,LAMA3,LGALS1,LGALS3,LGALS8,LPCAT2,LTBP4,MAL,MDK,MIR155HG,MMP14, MT1E,NAAA,NEUROG1,NR5A2,PAK1,PDPN,PLEC,PPM1L,PRDX1,PTEN,PXDN,RAP1GAP, RBP1,RILP,SFRP2,SFRP4,SGK1,SHH,SIAH1,SLC12A8,SOCS3,SPIB,SPTBN1,SYNPO2L, TGFB2,THBS1,THBS4,TLR2,TUBA8,USP4,VRK2	77
Cancer	metastasis of tumor	0.0006	CCNA1,CFLAR,CXCR4,EPAS1,EPHA2,ERBB2,FOXP1,GJA1,MMP14,SHH,TLX1	11
Cancer	uterine serous papillary cancer	0.0006	ANXA2,BCL2,CLDN5,DKK1,ERBB2,FRZB,GRB7,HLF,KCNB1,LAMP3,MAL,MEOX2,PLLP,SF RP4,SPINT2,THBS4,TNX8,ZEB1	18
Cancer	gastrointestinal tract cancer	0.0007	ACADS,ADAM12,APBA2,BCL2,CCND2,CHRNB1,CLSTN1,COL18A1,COL1A2,CRIP1,CXCR4, DKK1,DLC1,EBF4,EFEMP1,ENC1,EPAS1,ERBB2,FAS,GJA1,GLUL,HIST1H4A (includes others),HLCS,HSD11B2,JAG1,LAMA3,LGALS3,LGALS8,LPCAT2,LTBP4,MAL,MDK,MT1E,NA AA,NEUROG1,PAK1,PDPN (includes EG:10630),PLEC,PPM1L,PTEN,RAP1GAP,RILP,SFRP4,SGK1,SHH,SIAH1,SLC12A8,SPIB,S YNPO2L,TGFB2,THBS4,TLR2,TUBA8,USP4,VRK2	55
Cancer	endometrial cancer	0.0009	ANXA2,BCL2,CLDN5,DKK1,ERBB2,FRZB,GRB7,HLF,KCNB1,LAMP3,LTBP4,MAL,MEOX2, PLLP,PTEN,RASL12,SFRP4,SPINT2,SULT1A1,THBS4,TMEM200B,TNX8,TUBA8,ZEB1	24
Cancer	growth of tumor	0.0012	BCL2,CDH3,COL18A1,DKK1,EFNB2,EPAS1,EPHA2,ERBB2,FAS,GJA1,GJA4,GNMT,ITGB5, JAG1,LGALS1,LGALS3,NEUROD1,PTEN,RUNX1,SHH,THBS1,TLR2	22
Cancer	adenocarcinoma	0.0013	ANXA2,BCL2,BNC1,CCNA1,CLDN5,CXCR4,DKK1,DLC1,EFNB2,EPAS1,ERBB2,FAS,FRZB, GJA1,GRB7,GRIN2B,HLCS,HLF,KCNB1,LAMP3,LGALS3,LMO7,LTBP4,MAL,MEOX2,MMP14 ,PLLP,PPM1L,PRDX1,PTEN,RAP1GAP,RASL12,SFRP4,SHH,SPINT2,SPTBN1,SULT1A1, TGFBI,THBS1,THBS4,THRB,TMEM200B,TNX8,TUBA8,USP2,USP4,ZEB1	47
Cancer	lymphatic node tumor	0.0014	BCL2,BCL6,BIRC3,CD8A,CFLAR,COL12A1,CXCR4,EYA4,FAS,HDAC4,IER3,NR1I2,PRDX1, PRKCZ,PTEN,SGK1,SHH,TLR2,TUBA8,ZEB1	20
Cancer	testicular cancer	0.0014	CCND2,ERBB2,FAS,GDNF,LIN28A,PDPN,TUBA8	7
Cancer	germ cell and embryonal neoplasm	0.0014	BCL2,CXCR4,ERBB2,GDNF,LIN28A,LOXL3,MDK,NEUROD1,PDPN,PTEN,SHH,TUBA8	12

Category	Functions Annotation	p-Value	Molecules	# Molecules
Cancer	tumorigenesis of adenocarcinoma	0.0046	ERBB2,LMO7,MMP14,PRDX1,PTEN,SPTBN1,TGFBI,THBS1	8
Cancer	prostatic intraepithelial neoplasia	0.0048	ACTC1,ANXA2,BCL2,CDH3,ERBB2,IER3,PTEN,SIM2,SOCS3	9
Cancer	head and neck cancer	0.0048	ADAM12,ADAM8,AEBP1,ANGPT4,ANXA2,BCL2,CHRNB1,CXCR4,EFNB2,ENC1,EPHA2, ERBB2,FABP5,FRZB,GJA1,KLF5,LGALS1,LGALS3,LYN,MMP14,PDPN,PTEN,PYCARD, RAP1GAP,SFRP2,SFRP4,SHH,THBS1,THRB,TLR2,TUBA8,VRK2	32

7.5 Clinical information of the 28 non-metastatic breast tumour cases from TCGA data portal used in chapter four

This table shows clinical features of 28 non-metastatic breast tumours from TCGA data portal (batch 96) used in the methylation array. Parameters include; age at diagnosis, metastasis stage code, tumour stage code, estrogen receptor status (negative or positive), progesterone receptor status (negative, positive, or intermediate), Her2 receptor status (negative, positive, or equivocal), and case status (dead or alive).

Breast Cancer Patient Code	Age at initial pathologic diagnosis	Cancer metastasis stage code	Tumor stage code	Estrogen receptor status	Progesterone receptor status	Her2 receptor status	Vital status
TCGA-A7-A13E	62	M0	T2	Positive	Negative	Equivocal	LIVING
TCGA-A7-A13F	44	M0	T3	Positive	Positive	Equivocal	LIVING
TCGA-BH-A0AU	45	M0	T2	Positive	Positive	Positive	LIVING
TCGA-BH-A0AZ	47	M0	T3	Positive	Positive	Negative	LIVING
TCGA-BH-A0B2	56	M0	T2	Positive	Positive	Negative	LIVING
TCGA-BH-A0BC	55	M0	T3	Positive	Positive	Negative	LIVING
TCGA-BH-A0BF	55	M0	T3	Positive	Positive	Negative	LIVING
TCGA-BH-A0BS	56	M0	T1c	Positive	Positive	[Not Available]	LIVING
TCGA-BH-A0BT	56	M0	T1c	Positive	Positive	Negative	LIVING
TCGA-BH-A0BZ	59	M0	T3	Positive	Positive	Negative	LIVING
TCGA-BH-A0C3	47	M0	T1c	Positive	Negative	Negative	LIVING
TCGA-BH-A0DG	58	M0	T2	Positive	Negative	Negative	LIVING
TCGA-BH-A0DH	30	M0	T2	Positive	Positive	Negative	LIVING
TCGA-BH-A0DI	30	M0	T2	Positive	Positive	Negative	LIVING
TCGA-BH-A0DQ	63	M0	T2	Positive	Positive	Negative	LIVING
TCGA-BH-A0DV	54	M0	T2	Positive	Positive	Negative	LIVING
TCGA-BH-A0HA	31	M0	T1c	Positive	Positive	[Not Available]	LIVING
TCGA-BH-A1EO	68	M0	T1c	Positive	Positive	Negative	DECEASED
TCGA-BH-A1ES	35	M0	T2	Positive	Positive	Negative	DECEASED
TCGA-BH-A1ET	55	M0	T1	Positive	Positive	Negative	DECEASED
TCGA-BH-A1EU	83	M0	T1	Positive	Positive	Positive	DECEASED
TCGA-BH-A1EV	45	M0	T3	Positive	Positive	Positive	DECEASED
TCGA-BH-A1EW	38	M0	T2	Negative	Negative	Negative	DECEASED
TCGA-BH-A1F0	80	M0	T1a	Negative	Indeterminate	Negative	DECEASED
TCGA-E2-A15I	44	M0	T2	Positive	Positive	Equivocal	LIVING
TCGA-E2-A15K	58	M0	T2	Positive	Positive	Equivocal	LIVING
TCGA-E2-A1B5	46	M0	T2	Positive	Positive	Negative	LIVING
TCGA-E2-A1BC	63	M0	T1c	Positive	Positive	Negative	LIVING

7.6 CoBRA digestion results of the hypermethylated genes in BCBM

CoBRA digestion results are shown below for selected methylated and unmethylated samples of brain metastasis from breast cancer.
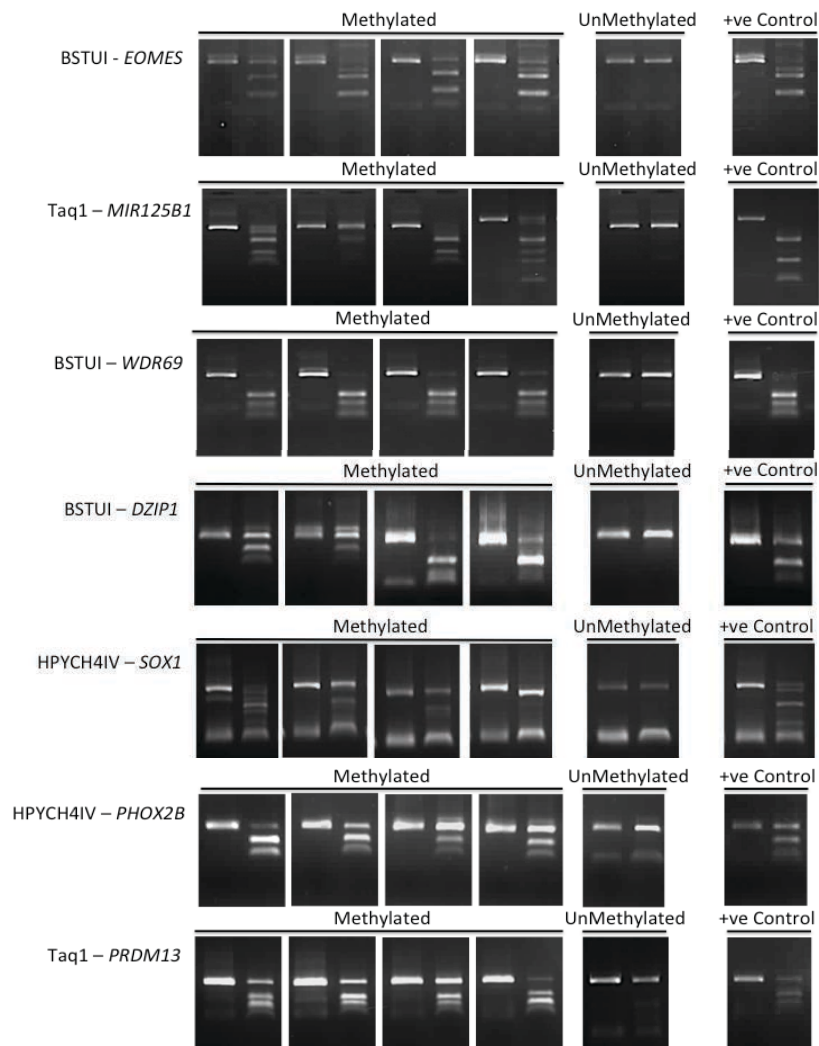
7.7 Data and results for LOH and MLPA in Schwannoma samples used in chapter five

Sample ID	Gender	NF2 mutation	LOH 22q	Loss/Gain of NF2 by MLPA
S14	Female	Yes	No	-
S21	Male	Yes	No	-
S36	Male	No	No	-
S37	Male	No	No	-
S44	Female	Yes	No	-
Lp152	Male	Yes	No	-
Lp272	Female	No	No	-
Lp303	Female	No	No	-

Chapter Eight: References

- Agrawal, Nishant, Mitchell J. Frederick, Curtis R. Pickering, Chetan Bettegowda, Kyle Chang, Ryan J. Li, Carole Fakhry, et al. 2011. "Exome Sequencing of Head and Neck Squamous Cell Carcinoma Reveals Inactivating Mutations in NOTCH1." *Science* 333 (6046): 1154–57. doi:10.1126/science.1206923.

- Ahmadi, Rezvan, Florian Stockhammer, Natalia Becker, Katarina Hohlen, Martin Misch, Arne Christians, Christine Dictus, et al. 2012. "No Prognostic Value of IDH1 Mutations in a Series of 100 WHO Grade II Astrocytomas." *Journal of Neuro-Oncology* 109 (1): 15–22. doi:10.1007/s11060-012-0863-y.

- Arai, Eri, Suenori Chiku, Taisuke Mori, Masahiro Gotoh, Tohru Nakagawa, Hiroyuki Fujimoto, and Yae Kanai. 2012. "Single-CpG-Resolution Methylome Analysis Identifies Clinicopathologically Aggressive CpG Island Methylator Phenotype Clear Cell Renal Cell Carcinomas." *Carcinogenesis* 33 (8): 1487–93. doi:10.1093/carcin/bgs177.

- Ars, Elisabet, Eduard Serra, Judit García, Helena Kruyer, Antonia Gaona, Conxi Lázaro, and Xavier Estivill. 2000. "Mutations Affecting mRNA Splicing Are the Most Common Molecular Defects in Patients with Neurofibromatosis Type 1." *Human Molecular Genetics* 9 (2): 237–47. doi:10.1093/hmg/9.2.237.

- Atreya, I., C. C Schimanski, C. Becker, S. Wirtz, H. Dornhoff, E. Schnurer, M. R Berger, P. R Galle, W. Herr, and M. F Neurath. 2007. "The T-Box Transcription Factor Eomesodermin Controls CD8 T Cell Activity and Lymph Node Metastasis

in Human Colorectal Cancer." *Gut* 56 (11): 1572–78. doi:10.1136/gut.2006.117812.

- Avvakumov, N., and J. Côté. 2007. "The MYST Family of Histone Acetyltransferases and Their Intimate Links to Cancer." *Oncogene* 26 (37): 5395–5407. doi:10.1038/sj.onc.1210608.

- Bady, Pierre, Davide Sciuscio, Annie-Claire Diserens, Jocelyne Bloch, Martin J. van den Bent, Christine Marosi, Pierre-Yves Dietrich, et al. 2012. "MGMT Methylation Analysis of Glioblastoma on the Infinium Methylation BeadChip Identifies Two Distinct CpG Regions Associated with Gene Silencing and Outcome, Yielding a Prediction Model for Comparisons across Datasets, Tumor Grades, and CIMP-Status." *Acta Neuropathologica* 124 (4): 547–60. doi:10.1007/s00401-012-1016-2.

- Baek, Sung Hee. 2011. "When Signaling Kinases Meet Histones and Histone Modifiers in the Nucleus." *Molecular Cell* 42 (3): 274–84. doi:10.1016/j.molcel.2011.03.022.

- Baeza, Nathalie, Michael Weller, Yasuhiro Yonekawa, Paul Kleihues, and Hiroko Ohgaki. 2003. "PTEN Methylation and Expression in Glioblastomas." *Acta Neuropathologica* 106 (5): 479–85. doi:10.1007/s00401-003-0748-4.

- Bakker, Janine L., Eswary Thirthagiri, Saskia E. van Mil, Muriel A. Adank, Hideyuki Ikeda, Henk M. W. Verheul, Hanne Meijers-Heijboer, Johan P. de Winter, Shyam K. Sharan, and Quinten Waisfisz. 2014. "A Novel Splice Site Mutation in the Noncoding Region of BRCA2: Implications for Fanconi Anemia

and Familial Breast Cancer Diagnostics." *Human Mutation* 35 (4): 442–46. doi:10.1002/humu.22505.

- Balss, Jörg, Jochen Meyer, Wolf Mueller, Andrey Korshunov, Christian Hartmann, and Andreas von Deimling. 2008. "Analysis of the IDH1 Codon 132 Mutation in Brain Tumors." *Acta Neuropathologica* 116 (6): 597–602. doi:10.1007/s00401-008-0455-2.

- Bamford, S, E Dawson, S Forbes, J Clements, R Pettett, A Dogan, A Flanagan, et al. 2004. "The COSMIC (Catalogue of Somatic Mutations in Cancer) Database and Website." *British Journal of Cancer* 91 (2): 355–58. doi:10.1038/sj.bjc.6601894.

- Banerji, Shantanu, Kristian Cibulskis, Claudia Rangel-Escareno, Kristin K. Brown, Scott L. Carter, Abbie M. Frederick, Michael S. Lawrence, et al. 2012. "Sequence Analysis of Mutations and Translocations across Breast Cancer Subtypes." *Nature* 486 (7403): 405–9. doi:10.1038/nature11154.

- Bannister, A. J., and T. Kouzarides. 1996. "The CBP Co-Activator Is a Histone Acetyltransferase." *Nature* 384 (6610): 641–43. doi:10.1038/384641a0.

- Barbus, Sebastian, Björn Tews, Daniela Karra, Meinhard Hahn, Bernhard Radlwimmer, Nicolas Delhomme, Christian Hartmann, et al. 2011. "Differential Retinoic Acid Signaling in Tumors of Long- and Short-Term Glioblastoma Survivors." *Journal of the National Cancer Institute* 103 (7): 598–606. doi:10.1093/jnci/djr036.

- Baudot, Anaïs, Victor de la Torre, and Alfonso Valencia. 2010. "Mutated Genes, Pathways and Processes in Tumours." *EMBO Reports* 11 (10): 805–10. doi:10.1038/embor.2010.133.

- Baylin, Stephen B., and Peter A. Jones. 2011. "A Decade of Exploring the Cancer Epigenome — Biological and Translational Implications." *Nature Reviews Cancer* 11 (10): 726–34. doi:10.1038/nrc3130.

- Bedford, M. T., and P. D. van Helden. 1987. "Hypomethylation of DNA in Pathological Conditions of the Human Prostate." *Cancer Research* 47 (20): 5274–76.

- Bennett, Kristi L., Todd Romigh, and Charis Eng. 2009. "AP-2alpha Induces Epigenetic Silencing of Tumor Suppressive Genes and Microsatellite Instability in Head and Neck Squamous Cell Carcinoma." *PloS One* 4 (9): e6931. doi:10.1371/journal.pone.0006931.

- Berdasco, María, Santiago Ropero, Fernando Setien, Mario F. Fraga, Pablo Lapunzina, Régine Losson, Miguel Alaminos, Nai-Kong Cheung, Nazneen Rahman, and Manel Esteller. 2009. "Epigenetic Inactivation of the Sotos Overgrowth Syndrome Gene Histone Methyltransferase NSD1 in Human Neuroblastoma and Glioma." *Proceedings of the National Academy of Sciences of the United States of America* 106 (51): 21830–35. doi:10.1073/pnas.0906831106.

- Bereshchenko, Oksana R., Wei Gu, and Riccardo Dalla-Favera. 2002. "Acetylation Inactivates the Transcriptional Repressor BCL6." *Nature Genetics* 32 (4): 606–13. doi:10.1038/ng1018.

- Berglund, Eva C., Anna Kiialainen, and Ann-Christine Syvänen. 2011. "Next-Generation Sequencing Technologies and Applications for Human Genetic History and Forensics." *Investigative Genetics* 2: 23. doi:10.1186/2041-2223-2-23.

- Bernards, René, and Robert A. Weinberg. 2002. "Metastasis Genes: A Progression Puzzle." *Nature* 418 (6900): 823–823. doi:10.1038/418823a.

- Bianco, Roberto, Davide Melisi, Fortunato Ciardiello, and Giampaolo Tortora. 2006. "Key Cancer Cell Signal Transduction Pathways as Therapeutic Targets." *European Journal of Cancer* 42 (3): 290–94. doi:10.1016/j.ejca.2005.07.034.

- Bian, Liu-guan, Qing-fang Sun, Wuttipong Tirakotai, Wei-guo Zhao, Jian-kang Shen, Qi-zhong Luo, and Helmut Bertalanffy. 2005. "Loss of Heterozygosity on Chromosome 22 in Sporadic Schwannoma and Its Relation to the Proliferation of Tumor Cells." *Chinese Medical Journal* 118 (18): 1517–24.

- Bibikova, Marina, Bret Barnes, Chan Tsan, Vincent Ho, Brandy Klotzle, Jennie M. Le, David Delano, et al. 2011. "High Density DNA Methylation Array with Single CpG Site Resolution." *Genomics* 98 (4): 288–95. doi:10.1016/j.ygeno.2011.07.007.

- Bick, David, and David Dimmock. 2011. "Whole Exome and Whole Genome Sequencing:" *Current Opinion in Pediatrics* 23 (6): 594–600. doi:10.1097/MOP.0b013e32834b20ec.

- Bishop, J. Michael. 1985. "Viral Oncogenes." *Cell* 42 (1): 23–38. doi:10.1016/S0092-8674(85)80098-2.

- Bos, Paula D., Xiang H.-F. Zhang, Cristina Nadal, Weiping Shu, Roger R. Gomis, Don X. Nguyen, Andy J. Minn, et al. 2009. "Genes That Mediate Breast Cancer Metastasis to the Brain." *Nature* 459 (7249): 1005–9. doi:10.1038/nature08021.

- Brennan, Cameron W., Roel G. W. Verhaak, Aaron McKenna, Benito Campos, Houtan Noushmehr, Sofie R. Salama, Siyuan Zheng, et al. 2013. "The Somatic Genomic Landscape of Glioblastoma." *Cell* 155 (2): 462–77. doi:10.1016/j.cell.2013.09.034.

- Brokinkel, Benjamin, Bernhard R. Fischer, Susanne Peetz-Dienhart, Heinrich Ebel, Abolghassem Sepehrnia, Burckhard Rama, Friedrich K. Albert, Walter Stummer, Werner Paulus, and Martin Hasselblatt. 2010. "MGMT Promoter Methylation Status in Anaplastic Meningiomas." *Journal of Neuro-Oncology* 100 (3): 489–90. doi:10.1007/s11060-010-0202-0.

- Buccoliero, A. M., F. Castiglione, D. Rossi Degl'Innocenti, M. Paglierani, V. Maio, C. F. Gheri, F. Garbini, et al. 2008. "O6-Methylguanine-DNA-Methyltransferase in Recurring Anaplastic Ependymomas: PCR and Immunohistochemistry." *Journal of Chemotherapy (Florence, Italy)* 20 (2): 263–68. doi:10.1179/joc.2008.20.2.263.

- Bush, William S., and Jason H. Moore. 2012. "Chapter 11: Genome-Wide Association Studies." *PLoS Computational Biology* 8 (12). doi:10.1371/journal.pcbi.1002822.

- Cadieux, Benoît, Tsui-Ting Ching, Scott R. VandenBerg, and Joseph F. Costello. 2006. "Genome-Wide Hypomethylation in Human Glioblastomas Associated with

Specific Copy Number Alteration, Methylenetetrahydrofolate Reductase Allele Status, and Increased Proliferation." *Cancer Research* 66 (17): 8469–76. doi:10.1158/0008-5472.CAN-06-1547.

- Cao, Ru, and Yi Zhang. 2004. "SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex." *Molecular Cell* 15 (1): 57–67. doi:10.1016/j.molcel.2004.06.020.

- Celis-Aguilar, Erika, Luis Lassaletta, Torres-Mart&#237, Miguel N, F. Yuri Rodrigues, Manuel Nistal, Javier S. Castresana, Javier Gavilan, and Juan A. Rey. 2012. "The Molecular Biology of Vestibular Schwannomas and Its Association with Hearing Loss: A Review." *Genetics Research International* 2012 (February). doi:10.1155/2012/856157.

- Chang, Vivian Y, Giuseppe Basso, Kathleen M Sakamoto, and Stanley F Nelson. 2013. "Identification of Somatic and Germline Mutations Using Whole Exome Sequencing of Congenital Acute Lymphoblastic Leukemia." *BMC Cancer* 13 (1): 55. doi:10.1186/1471-2407-13-55.

- Chen, Jian, Renée M McKay, and Luis F Parada. 2012. "Malignant Glioma: Lessons from Genomics, Mouse Models, and Stem Cells." *Cell* 149 (1): 36–47. doi:10.1016/j.cell.2012.03.009.

- Chen, Yi-an, Mathieu Lemire, Sanaa Choufani, Darci T. Butcher, Daria Grafodatskaya, Brent W. Zanke, Steven Gallinger, Thomas J. Hudson, and Rosanna Weksberg. 2013. "Discovery of Cross-Reactive Probes and Polymorphic

CpGs in the Illumina Infinium HumanMethylation450 Microarray." *Epigenetics* 8 (2): 203–9. doi:10.4161/epi.23470.

- Christensen, Brock C., Ashley A. Smith, Shichun Zheng, Devin C. Koestler, E. Andres Houseman, Carmen J. Marsit, Joseph L. Wiemels, et al. 2011. "DNA Methylation, Isocitrate Dehydrogenase Mutation, and Survival in Glioma." *Journal of the National Cancer Institute* 103 (2): 143–53. doi:10.1093/jnci/djq497.

- Clark, Victoria E., E. Zeynep Erson-Omay, Akdes Serin, Jun Yin, Justin Cotney, Koray Özduman, Timuçin Avşar, et al. 2013. "Genomic Analysis of Non-NF2 Meningiomas Reveals Mutations in TRAF7, KLF4, AKT1, and SMO." *Science* 339 (6123): 1077–80. doi:10.1126/science.1233009.

- Costello, Joseph F., Michael C. Frühwald, Dominic J. Smiraglia, Laura J. Rush, Gavin P. Robertson, Xin Gao, Fred A. Wright, et al. 2000. "Aberrant CpG-Island Methylation Has Non-Random and Tumour-Type–specific Patterns." *Nature Genetics* 24 (2): 132–38. doi:10.1038/72785.

- Dalgliesh, Gillian L., Kyle Furge, Chris Greenman, Lina Chen, Graham Bignell, Adam Butler, Helen Davies, et al. 2010. "Systematic Sequencing of Renal Carcinoma Reveals Inactivation of Histone Modifying Genes." *Nature* 463 (7279): 360–63. doi:10.1038/nature08672.

- Dallol, Ashraf, Dietmar Krex, Luke Hesson, Charis Eng, Eamonn R. Maher, and Farida Latif. 2003. "Frequent Epigenetic Inactivation of the SLIT2 Gene in Gliomas." *Oncogene* 22 (29): 4611–16. doi:10.1038/sj.onc.1206687.

- Davidson, B., R. Reich, I. Goldberg, W. H. Gotlieb, J. Kopolovic, A. Berner, G. Ben-Baruch, M. Bryne, and J. M. Nesland. 2001. "Ets-1 Messenger RNA Expression Is a Novel Marker of Poor Survival in Ovarian Carcinoma." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 7 (3): 551–57.

- Dawson, Mark A., and Tony Kouzarides. 2012. "Cancer Epigenetics: From Mechanism to Therapy." *Cell* 150 (1): 12–27. doi:10.1016/j.cell.2012.06.013.

- Dedeurwaerder, Sarah, Matthieu Defrance, Martin Bizet, Emilie Calonne, Gianluca Bontempi, and François Fuks. 2013. "A Comprehensive Overview of Infinium HumanMethylation450 Data Processing." *Briefings in Bioinformatics*, August, bbt054. doi:10.1093/bib/bbt054.

- Dedeurwaerder, Sarah, Matthieu Defrance, Emilie Calonne, Hélène Denis, Christos Sotiriou, and François Fuks. 2011. "Evaluation of the Infinium Methylation 450K Technology." *Epigenomics* 3 (6): 771–84. doi:10.2217/epi.11.105.

- De Pontual, Loïc, Delphine Trochet, Franck Bourdeaut, Sophie Thomas, Heather Etchevers, Agnes Chompret, Véronique Minard, et al. 2007. "Methylation-Associated PHOX2B Gene Silencing Is a Rare Event in Human Neuroblastoma." *European Journal of Cancer* 43 (16): 2366–72. doi:10.1016/j.ejca.2007.07.016.

- Di Vinci, Angela, Ida Casciano, Elena Marasco, Barbara Banelli, Gian Luigi Ravetti, Luana Borzì, Claudio Brigati, et al. 2012. "Quantitative Methylation Analysis of HOXA3, 7, 9, and 10 Genes in Glioma: Association with Tumor

WHO Grade and Clinical Outcome." *Journal of Cancer Research and Clinical Oncology* 138 (1): 35–47. doi:10.1007/s00432-011-1070-5.

- Du, Pan, Warren A. Kibbe, and Simon M. Lin. 2008. "Lumi: A Pipeline for Processing Illumina Microarray." *Bioinformatics (Oxford, England)* 24 (13): 1547–48. doi:10.1093/bioinformatics/btn224.

- D'Urso, Pietro Ivo, Oscar Fernando D'Urso, Carlo Storelli, Giuseppe Catapano, Cosimo Damiano Gianfreda, Antonio Montinaro, Antonella Muscella, and Santo Marsigliante. 2011. "Retrospective Protein Expression and Epigenetic Inactivation Studies of CDH1 in Patients Affected by Low-Grade Glioma." *Journal of Neuro-Oncology* 104 (1): 113–18. doi:10.1007/s11060-010-0481-5.

- Eberhard, David A., Bruce E. Johnson, Lukas C. Amler, Audrey D. Goddard, Sherry L. Heldens, Roy S. Herbst, William L. Ince, et al. 2005. "Mutations in the Epidermal Growth Factor Receptor and in KRAS Are Predictive and Prognostic Indicators in Patients with Non-Small-Cell Lung Cancer Treated with Chemotherapy Alone and in Combination with Erlotinib." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 23 (25): 5900–5909. doi:10.1200/JCO.2005.02.857.

- Ebinger, Martin, Leonore Senf, Olga Wachowski, and Wolfram Scheurlen. 2004. "Promoter Methylation Pattern of Caspase-8, P16INK4A, MGMT, TIMP-3, and E-Cadherin in Medulloblastoma." *Pathology Oncology Research: POR* 10 (1): 17–21. doi:PAOR.2004.10.1.0017.

- Ehrlich, Melanie. 2009. "DNA Hypomethylation in Cancer Cells." *Epigenomics* 1 (2): 239–59. doi:10.2217/epi.09.33.

- Ehrlich, M., C. B. Woods, M. C. Yu, L. Dubeau, F. Yang, M. Campan, D. J. Weisenberger, et al. 2006. "Quantitative Analysis of Associations between DNA Hypermethylation, Hypomethylation, and DNMT RNA Levels in Ovarian Tumors." *Oncogene* 25 (18): 2636–45. doi:10.1038/sj.onc.1209145.

- Esteller, Manel, Carlos Cordon-Cardo, Paul G. Corn, Steve J. Meltzer, Kamal S. Pohar, D. Neil Watkins, Gabriel Capella, et al. 2001. "p14ARF Silencing by Promoter Hypermethylation Mediates Abnormal Intracellular Localization of MDM2." *Cancer Research* 61 (7): 2816–21.

- Esteller, Manel, Jesus Garcia-Foncillas, Esther Andion, Steven N. Goodman, Oscar F. Hidalgo, Vicente Vanaclocha, Stephen B. Baylin, and James G. Herman. 2000. "Inactivation of the DNA-Repair Gene MGMT and the Clinical Response of Gliomas to Alkylating Agents." *New England Journal of Medicine* 343 (19): 1350–54. doi:10.1056/NEJM200011093431901.

- Esteller, Manel, Jose Maria Silva, Gemma Dominguez, Felix Bonilla, Xavier Matias-Guiu, Enrique Lerma, Elena Bussaglia, et al. 2000. "Promoter Hypermethylation and BRCA1 Inactivation in Sporadic Breast and Ovarian Tumors." *Journal of the National Cancer Institute* 92 (7): 564–69. doi:10.1093/jnci/92.7.564.

- Esteller, Manel, Silvia Tortola, Minoru Toyota, Gabriel Capella, Miguel Angel Peinado, Stephen B. Baylin, and James G. Herman. 2000. "Hypermethylation-

Associated Inactivation of p14ARF Is Independent of p16INK4a Methylation and p53 Mutational Status." *Cancer Research* 60 (1): 129–33.

- Esteller, M., S. R. Hamilton, P. C. Burger, S. B. Baylin, and J. G. Herman. 1999. "Inactivation of the DNA Repair Gene O6-Methylguanine-DNA Methyltransferase by Promoter Hypermethylation Is a Common Event in Primary Human Neoplasia." *Cancer Research* 59 (4): 793–97.

- Evans, D. Gareth R. 2009. "Neurofibromatosis Type 2 (NF2): A Clinical and Molecular Review." *Orphanet Journal of Rare Diseases* 4: 16. doi:10.1186/1750-1172-4-16.

- Evans, D. G. R., E. R. Maher, and M. E. Baser. 2005. "Age Related Shift in the Mutation Spectra of Germline and Somatic NF2 Mutations: Hypothetical Role of DNA Repair Mechanisms." *Journal of Medical Genetics* 42 (8): 630–32. doi:10.1136/jmg.2004.027953.

- Fang, F., S. Turcan, A. Rimner, A. Kaufman, D. Giri, L. G. T. Morris, R. Shen, et al. 2011. "Breast Cancer Methylomes Establish an Epigenomic Foundation for Metastasis." *Science Translational Medicine* 3 (75): 75ra25–75ra25. doi:10.1126/scitranslmed.3001875.

- Feinberg, A. P., C. W. Gehrke, K. C. Kuo, and M. Ehrlich. 1988. "Reduced Genomic 5-Methylcytosine Content in Human Colonic Neoplasia." *Cancer Research* 48 (5): 1159–61.

- Feng, Ying, Zheng Wang, Zhaoshi Bao, Wei Yan, Gan You, Yinyan Wang, Huimin Hu, Wei Zhang, Quangeng Zhang, and Tao Jiang. 2014. "SOCS3

Promoter Hypermethylation Is a Favorable Prognosticator and a Novel Indicator for G-CIMP-Positive GBM Patients." *PLoS ONE* 9 (3): e91829. doi:10.1371/journal.pone.0091829.

- Ferlay, Jacques, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. 2015. "Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012: Globocan 2012." *International Journal of Cancer* 136 (5): E359–86. doi:10.1002/ijc.29210.

- Figarella-Branger, Dominique, Corinne Bouvier, André Maues de Paula, Karima Mokhtari, Carole Colin, Anderson Loundou, Olivier Chinot, and Philippe Metellus. 2012. "Molecular Genetics of Adult Grade II Gliomas: Towards a Comprehensive Tumor Classification System." *Journal of Neuro-Oncology* 110 (2): 205–13. doi:10.1007/s11060-012-0953-x.

- Figueroa, Maria E., Omar Abdel-Wahab, Chao Lu, Patrick S. Ward, Jay Patel, Alan Shih, Yushan Li, et al. 2010. "Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation." *Cancer Cell* 18 (6): 553–67. doi:10.1016/j.ccr.2010.11.015.

- Fleisher, A. S., M. Esteller, S. Wang, G. Tamura, H. Suzuki, J. Yin, T. T. Zou, et al. 1999. "Hypermethylation of the hMLH1 Gene Promoter in Human Gastric Cancers with Microsatellite Instability." *Cancer Research* 59 (5): 1090–95.

- Forbes, S.A., G. Bhamra, S. Bamford, E. Dawson, C. Kok, J. Clements, A. Menzies, J.W. Teague, P.A. Futreal, and M.R. Stratton. 2008. "The Catalogue of Somatic Mutations in Cancer (COSMIC)." *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.]* CHAPTER (April): Unit – 10.11. doi:10.1002/0471142905.hg1011s57.

- Forbes, Simon A., David Beare, Prasad Gunasekaran, Kenric Leung, Nidhi Bindal, Harry Boutselakis, Minjie Ding, et al. 2015. "COSMIC: Exploring the World's Knowledge of Somatic Mutations in Human Cancer." *Nucleic Acids Research* 43 (D1): D805–11. doi:10.1093/nar/gku1075.

- Forbes, Simon A, Gurpreet Tang, Chai Kok, Mingming Jia, Sally Bamford, Jennifer Cole, Elisabeth Dawson, et al. 2008. "An Introduction to COSMIC, the Catalogue of Somatic Mutations in Cancer." *NCI Nature Pathway Interaction Database*, July. doi:10.1038/pid.2008.3.

- Gama-Sosa, M. A., V. A. Slagel, R. W. Trewyn, R. Oxenhandler, K. C. Kuo, C. W. Gehrke, and M. Ehrlich. 1983. "The 5-Methylcytosine Content of DNA from Human Tumors." *Nucleic Acids Research* 11 (19): 6883–94.

- Gartner, Jared J., Stephen C. J. Parker, Todd D. Prickett, Ken Dutton-Regester, Michael L. Stitzel, Jimmy C. Lin, Sean Davis, et al. 2013. "Whole-Genome Sequencing Identifies a Recurrent Functional Synonymous Mutation in Melanoma." *Proceedings of the National Academy of Sciences*, July, 201304227. doi:10.1073/pnas.1304227110.

- Gavrilovic, Igor T., and Jerome B. Posner. 2005a. "Brain Metastases: Epidemiology and Pathophysiology." *Journal of Neuro-Oncology* 75 (1): 5–14. doi:10.1007/s11060-004-8093-6.

- ———. 2005b. "Brain Metastases: Epidemiology and Pathophysiology." *Journal of Neuro-Oncology* 75 (1): 5–14. doi:10.1007/s11060-004-8093-6.

- Gerlinger, Marco, Andrew J. Rowan, Stuart Horswell, James Larkin, David Endesfelder, Eva Gronroos, Pierre Martinez, et al. 2012. "Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing." *New England Journal of Medicine* 366 (10): 883–92. doi:10.1056/NEJMoa1113205.

- Gonzalez-Gomez, Pilar, M Josefa Bello, M Eva Alonso, Jesus Lomas, Dolores Arjona, Jose M de Campos, Jesus Vaquero, et al. 2003. "CpG Island Methylation in Sporadic and Neurofibromatis Type 2-Associated Schwannomas." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 9 (15): 5601–6.

- Gonzalez-Zulueta, M., C. M. Bender, A. S. Yang, T. Nguyen, R. W. Beart, J. M. Van Tornout, and P. A. Jones. 1995. "Methylation of the 5' CpG Island of the p16/CDKN2 Tumor Suppressor Gene in Normal and Transformed Human Tissues Correlates with Gene Silencing." *Cancer Research* 55 (20): 4531–35.

- Götze, Silke, Valeska Feldhaus, Thilo Traska, Marietta Wolter, Guido Reifenberger, Andrea Tannapfel, Cornelius Kuhnen, Dirk Martin, Oliver Müller, and Sonja Sievers. 2009. "ECRG4 Is a Candidate Tumor Suppressor Gene

Frequently Hypermethylated in Colorectal Carcinoma and Glioma." *BMC Cancer* 9: 447. doi:10.1186/1471-2407-9-447.

- Götze, Silke, Marietta Wolter, Guido Reifenberger, Oliver Müller, and Sonja Sievers. 2010. "Frequent Promoter Hypermethylation of Wnt Pathway Inhibitor Genes in Malignant Astrocytic Gliomas." *International Journal of Cancer. Journal International Du Cancer* 126 (11): 2584–93. doi:10.1002/ijc.24981.

- Grada, Ayman, and Kate Weinbrecht. 2013. "Next-Generation Sequencing: Methodology and Application." *Journal of Investigative Dermatology* 133 (8): e11. doi:10.1038/jid.2013.248.

- Hadfield, K. D., M. J. Smith, J. E. Urquhart, A. J. Wallace, N. L. Bowers, A. T. King, S. A. Rutherford, D. Trump, W. G. Newman, and D. G. Evans. 2010. "Rates of Loss of Heterozygosity and Mitotic Recombination in NF2 Schwannomas, Sporadic Vestibular Schwannomas and Schwannomatosis Schwannomas." *Oncogene* 29 (47): 6216–21. doi:10.1038/onc.2010.363.

- Hamamoto, Ryuji, Yoichi Furukawa, Masashi Morita, Yuko Iimura, Fabio Pittella Silva, Meihua Li, Ryuichiro Yagyu, and Yusuke Nakamura. 2004. "SMYD3 Encodes a Histone Methyltransferase Involved in the Proliferation of Cancer Cells." *Nature Cell Biology* 6 (8): 731–40. doi:10.1038/ncb1151.

- Hartmann, Christian, Bettina Hentschel, Matthias Simon, Manfred Westphal, Gabriele Schackert, Jörg C. Tonn, Markus Loeffler, et al. 2013. "Long-Term Survival in Primary Glioblastoma With Versus Without Isocitrate Dehydrogenase

Mutations." *Clinical Cancer Research* 19 (18): 5146–57. doi:10.1158/1078-0432.CCR-13-0017.

- Hartmann, Christian, Bettina Hentschel, Marcos Tatagiba, Johannes Schramm, Oliver Schnell, Clemens Seidel, Robert Stein, et al. 2011. "Molecular Markers in Low-Grade Gliomas: Predictive or Prognostic?" *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 17 (13): 4588–99. doi:10.1158/1078-0432.CCR-10-3194.

- Hauer, Kristina, Julia Calzada-Wack, Katja Steiger, Thomas G. P. Grunewald, Daniel Baumhoer, Stephanie Plehm, Thorsten Buch, et al. 2013. "DKK2 Mediates Osteolysis, Invasiveness, and Metastatic Spread in Ewing Sarcoma." *Cancer Research* 73 (2): 967–77. doi:10.1158/0008-5472.CAN-12-1492.

- Hawkins, Shannon M., Holli A. Loomans, Ying-Wooi Wan, Triparna Ghosh-Choudhury, Donna Coffey, Weimin Xiao, Zhandong Liu, Haleh Sangi-Haghpeykar, and Matthew L. Anderson. 2013. "Expression and Functional Pathway Analysis of Nuclear Receptor NR2F2 in Ovarian Cancer." *The Journal of Clinical Endocrinology and Metabolism* 98 (7): E1152–62. doi:10.1210/jc.2013-1081.

- Hegi, Monika E., Annie-Claire Diserens, Thierry Gorlia, Marie-France Hamou, Nicolas de Tribolet, Michael Weller, Johan M. Kros, et al. 2005. "MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma." *The New England Journal of Medicine* 352 (10): 997–1003. doi:10.1056/NEJMoa043331.

- Herman, J. G., A. Umar, K. Polyak, J. R. Graff, N. Ahuja, J. P. Issa, S. Markowitz, et al. 1998. "Incidence and Functional Consequences of hMLH1 Promoter Hypermethylation in Colorectal Carcinoma." *Proceedings of the National Academy of Sciences of the United States of America* 95 (12): 6870–75.

- Herrlinger, Ulrich, Johannes Rieger, Dorothee Koch, Simon Loeser, Britta Blaschke, Rolf-Dieter Kortmann, Joachim P. Steinbach, et al. 2006. "Phase II Trial of Lomustine plus Temozolomide Chemotherapy in Addition to Radiotherapy in Newly Diagnosed Glioblastoma: UKT-03." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 24 (27): 4412–17. doi:10.1200/JCO.2006.06.9104.

- Hesson, Luke, Ivan Bièche, Dietmar Krex, Emmanuelle Criniere, Khê Hoang-Xuan, Eamonn R. Maher, and Farida Latif. 2004. "Frequent Epigenetic Inactivation of RASSF1A and BLU Genes Located within the Critical 3p21.3 Region in Gliomas." *Oncogene* 23 (13): 2408–19. doi:10.1038/sj.onc.1207407.

- Hill, Victoria K., Thoraia Shinawi, Christopher J. Ricketts, Dietmar Krex, Gabriele Schackert, Julien Bauer, Wenbin Wei, Garth Cruickshank, Eamonn R. Maher, and Farida Latif. 2014. "Stability of the CpG Island Methylator Phenotype during Glioma Progression and Identification of Methylated Loci in Secondary Glioblastomas." *BMC Cancer* 14: 506. doi:10.1186/1471-2407-14-506.

- Hoffmann, Michèle J., and Wolfgang A. Schulz. 2005. "Causes and Consequences of DNA Hypomethylation in Human Cancer." *Biochemistry and*

Cell Biology = Biochimie Et Biologie Cellulaire 83 (3): 296–321. doi:10.1139/o05-036.

- HORAK, CHRISTINE E., JONG HEUN LEE, JEAN-CLAUDE MARSHALL, S. MARTIN SHREEVE, and PATRICIA S. STEEG. 2008. "The Role of Metastasis Suppressor Genes in Metastatic Dormancy." *APMIS : Acta Pathologica, Microbiologica, et Immunologica Scandinavica* 116 (7-8): 586–601. doi:10.1111/j.1600-0463.2008.01213.x.

- Horbinski, Craig, Lindsey Kelly, Yuri E. Nikiforov, Mary Beth Durso, and Marina N. Nikiforova. 2010. "Detection of IDH1 and IDH2 Mutations by Fluorescence Melting Curve Analysis as a Diagnostic Tool for Brain Biopsies." *The Journal of Molecular Diagnostics: JMD* 12 (4): 487–92. doi:10.2353/jmoldx.2010.090228.

- Horiguchi, Keishi, Yoshio Tomizawa, Masahiko Tosaka, Shogo Ishiuchi, Hideyuki Kurihara, Masatomo Mori, and Nobuhito Saito. 2003. "Epigenetic Inactivation of RASSF1A Candidate Tumor Suppressor Gene at 3p21.3 in Brain Tumors." *Oncogene* 22 (49): 7862–65. doi:10.1038/sj.onc.1207082.

- Horn, Susanne, Adina Figl, P. Sivaramakrishna Rachakonda, Christine Fischer, Antje Sucker, Andreas Gast, Stephanie Kadel, et al. 2013. "TERT Promoter Mutations in Familial and Sporadic Melanoma." *Science* 339 (6122): 959–61. doi:10.1126/science.1230062.

- Huang, Franklin W., Eran Hodis, Mary Jue Xu, Gregory V. Kryukov, Lynda Chin, and Levi A. Garraway. 2013. "Highly Recurrent TERT Promoter Mutations

in Human Melanoma." *Science* 339 (6122): 957–59. doi:10.1126/science.1229259.

- Huang, Li, Junhua Luo, Qingqing Cai, Qiuhui Pan, Hong Zeng, Zhenghui Guo, Wen Dong, Jian Huang, and Tianxin Lin. 2011. "MicroRNA-125b Suppresses the Development of Bladder Cancer by Targeting E2F3." *International Journal of Cancer. Journal International Du Cancer* 128 (8): 1758–69. doi:10.1002/ijc.25509.

- Huse, Jason T., and Eric C. Holland. 2010. "Targeting Brain Cancer: Advances in the Molecular Pathology of Malignant Glioma and Medulloblastoma." *Nature Reviews Cancer* 10 (5): 319–31. doi:10.1038/nrc2818.

- "Integrating Common and Rare Genetic Variation in Diverse Human Populations." 2010. *Nature* 467 (7311): 52–58. doi:10.1038/nature09298.

- Iorio, Marilena V., Manuela Ferracin, Chang-Gong Liu, Angelo Veronese, Riccardo Spizzo, Silvia Sabbioni, Eros Magri, et al. 2005. "MicroRNA Gene Expression Deregulation in Human Breast Cancer." *Cancer Research* 65 (16): 7065–70. doi:10.1158/0008-5472.CAN-05-1783.

- Irizarry, Rafael A., Christine Ladd-Acosta, Bo Wen, Zhijin Wu, Carolina Montano, Patrick Onyango, Hengmi Cui, et al. 2009. "Genome-Wide Methylation Analysis of Human Colon Cancer Reveals Similar Hypo- and Hypermethylation at Conserved Tissue-Specific CpG Island Shores." *Nature Genetics* 41 (2): 178–86. doi:10.1038/ng.298.

- Issa, Charbel M., R. Semrau, R. Kath, and K. Höffken. 2002. "Isolated Brain Metastases as the Sole Manifestation of a Late Relapse in Breast Cancer." *Journal of Cancer Research and Clinical Oncology* 128 (1): 61–63. doi:10.1007/s004320100286.

- Jacoby, Lee B., Mia MacCollin, David N. Louls, Trina Mohney, Mari-Paz Rublo, Karen Pulaski, James A. Trofatter, et al. 1994. "Exon Scanning for Mutation of the NF2 Gene in Schwannomas." *Human Molecular Genetics* 3 (3): 413–19. doi:10.1093/hmg/3.3.413.

- Jansson, Martin D., and Anders H. Lund. 2012. "MicroRNA and Cancer." *Molecular Oncology*, Cancer epigenetics, 6 (6): 590–610. doi:10.1016/j.molonc.2012.09.006.

- Jia, Hong-Yan, Yu-Xuan Wang, Wen-Ting Yan, Hui-Yu Li, Yan-Zhang Tian, Shi-Ming Wang, and Hao-Liang Zhao. 2012. "MicroRNA-125b Functions as a Tumor Suppressor in Hepatocellular Carcinoma Cells." *International Journal of Molecular Sciences* 13 (7): 8762–74. doi:10.3390/ijms13078762.

- Jiang, G. L., and S. Huang. 2000. "The Yin-Yang of PR-Domain Family Genes in Tumorigenesis." *Histology and Histopathology* 15 (1): 109–17.

- Jin, Seung-Gi, Swati Kadam, and Gerd P. Pfeifer. 2010. "Examination of the Specificity of DNA Methylation Profiling Techniques towards 5-Methylcytosine and 5-Hydroxymethylcytosine." *Nucleic Acids Research* 38 (11): e125. doi:10.1093/nar/gkq223.

- Jones, Siân, Ralph H. Hruban, Mihoko Kamiyama, Michael Borges, Xiaosong Zhang, D. Williams Parsons, Jimmy Cheng-Ho Lin, et al. 2009. "Exomic Sequencing Identifies PALB2 as a Pancreatic Cancer Susceptibility Gene." *Science (New York, N.Y.)* 324 (5924): 217. doi:10.1126/science.1171202.

- Juratli, Tareq A., Matthias Kirsch, Katja Robel, Silke Soucek, Kathrin Geiger, Rüdiger von Kummer, Gabriele Schackert, and Dietmar Krex. 2012. "IDH Mutations as an Early and Consistent Marker in Low-Grade Astrocytomas WHO Grade II and Their Consecutive Secondary High-Grade Gliomas." *Journal of Neuro-Oncology* 108 (3): 403–10. doi:10.1007/s11060-012-0844-1.

- Kang, Yibin, Peter M. Siegel, Weiping Shu, Maria Drobnjak, Sanna M. Kakonen, Carlos Cordón-Cardo, Theresa A. Guise, and Joan Massagué. 2003. "A Multigenic Program Mediating Breast Cancer Metastasis to Bone." *Cancer Cell* 3 (6): 537–49.

- Kawakita, Akiko, Souichi Yanamoto, Shin-Ichi Yamada, Tomofumi Naruse, Hidenori Takahashi, Goro Kawasaki, and Masahiro Umeda. 2013. "MicroRNA-21 Promotes Oral Cancer Invasion via the Wnt/β-Catenin Pathway by Targeting DKK2." *Pathology Oncology Research: POR*, September. doi:10.1007/s12253-013-9689-y.

- Kikuyama, Mizuho, Hideyuki Takeshima, Takayuki Kinoshita, Eriko Okochi-Takada, Mika Wakabayashi, Sadako Akashi-Tanaka, Toshihisa Ogawa, Yasuyuki Seto, and Toshikazu Ushijima. 2012. "Development of a Novel Approach, the Epigenome-Based Outlier Approach, to Identify Tumor-Suppressor Genes

Silenced by Aberrant DNA Methylation." *Cancer Letters* 322 (2): 204–12. doi:10.1016/j.canlet.2012.03.016.

- Kim, Young-Ho, Sumihito Nobusawa, Michel Mittelbronn, Werner Paulus, Benjamin Brokinkel, Kathy Keyvani, Ulrich Sure, et al. 2010. "Molecular Classification of Low-Grade Diffuse Gliomas." *The American Journal of Pathology* 177 (6): 2708–14. doi:10.2353/ajpath.2010.100680.

- King, Mary-Claire, Joan H. Marks, and Jessica B. Mandell. 2003. "Breast and Ovarian Cancer Risks Due to Inherited Mutations in BRCA1 and BRCA2." *Science* 302 (5645): 643–46. doi:10.1126/science.1088759.

- Knobbe, Christiane B., Julia Reifenberger, Britta Blaschke, and Guido Reifenberger. 2004. "Hypermethylation and Transcriptional Downregulation of the Carboxyl-Terminal Modulator Protein Gene in Glioblastomas." *Journal of the National Cancer Institute* 96 (6): 483–86.

- Knudson, A. G. 1971. "Mutation and Cancer: Statistical Study of Retinoblastoma." *Proceedings of the National Academy of Sciences of the United States of America* 68 (4): 820–23.

- Kondziolka, Douglas, Seyed H Mousavi, Hideyuki Kano, John C Flickinger, and L Dade Lunsford. 2012. "The Newly Diagnosed Vestibular Schwannoma: Radiosurgery, Resection, or Observation?" *Neurosurgical Focus* 33 (3): E8. doi:10.3171/2012.6.FOCUS12192.

- Krell, Daniel, Mawuelikem Assoku, Malcolm Galloway, Paul Mulholland, Ian Tomlinson, and Chiara Bardella. 2011. "Screen for IDH1, IDH2, IDH3,

D2HGDH and L2HGDH Mutations in Glioblastoma." *PloS One* 6 (5): e19868. doi:10.1371/journal.pone.0019868.

- Krex, Dietmar, Barbara Klink, Christian Hartmann, Andreas von Deimling, Torsten Pietsch, Matthias Simon, Michael Sabel, et al. 2007. "Long-Term Survival with Glioblastoma Multiforme." *Brain: A Journal of Neurology* 130 (Pt 10): 2596–2606. doi:10.1093/brain/awm204.

- Kriaucionis, Skirmantas, and Nathaniel Heintz. 2009. "The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain." *Science* 324 (5929): 929–30. doi:10.1126/science.1169786.

- Kullar, P J, D M Pearson, D S Malley, V P Collins, and K Ichimura. 2010. "CpG Island Hypermethylation of the Neurofibromatosis Type 2 (NF2) Gene Is Rare in Sporadic Vestibular Schwannomas." *Neuropathology and Applied Neurobiology* 36 (6): 505–14. doi:10.1111/j.1365-2990.2010.01090.x.

- Kunitz, Annegret, Marietta Wolter, Jörg van den Boom, Jörg Felsberg, Björn Tews, Meinhard Hahn, Axel Benner, et al. 2007. "DNA Hypermethylation and Aberrant Expression of the EMP3 Gene at 19q13.3 in Human Gliomas." *Brain Pathology (Zurich, Switzerland)* 17 (4): 363–70. doi:10.1111/j.1750-3639.2007.00083.x.

- Lai, Hung-Cheng, Yu-Che Ou, Tze-Chien Chen, Huei-Jean Huang, Ya-Min Cheng, Chi-Hau Chen, Tang-Yuan Chu, et al. 2014. "PAX1/SOX1 DNA Methylation and Cervical Neoplasia Detection: A Taiwanese Gynecologic

Oncology Group (TGOG) Study." *Cancer Medicine* 3 (4): 1062–74. doi:10.1002/cam4.253.

- Lassaletta, Luis, M Josefa Bello, Laura Del Río, Carolina Alfonso, Jose Maria Roda, Juan A Rey, and Javier Gavilan. 2006. "DNA Methylation of Multiple Genes in Vestibular Schwannoma: Relationship with Clinical and Radiological Findings." *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology* 27 (8): 1180–85. doi:10.1097/01.mao.0000226291.42165.22.

- Lee, Jong Dae, Tae Jun Kwon, Un-Kyung Kim, and Won-Sang Lee. 2012. "Genetic and Epigenetic Alterations of the NF2 Gene in Sporadic Vestibular Schwannomas." *PloS One* 7 (1): e30418. doi:10.1371/journal.pone.0030418.

- Lee, W. H., R. Bookstein, F. Hong, L. J. Young, J. Y. Shew, and E. Y. Lee. 1987. "Human Retinoblastoma Susceptibility Gene: Cloning, Identification, and Sequence." *Science (New York, N.Y.)* 235 (4794): 1394–99.

- Leshchenko, Violetta V., Pei-Yu Kuo, Rita Shaknovich, David T. Yang, Tobias Gellen, Adam Petrich, Yiting Yu, et al. 2010. "Genomewide DNA Methylation Analysis Reveals Novel Targets for Drug Development in Mantle Cell Lymphoma." *Blood* 116 (7): 1025–34. doi:10.1182/blood-2009-12-257485.

- Li, D., and R. Roberts. 2001. "WD-Repeat Proteins: Structure Characteristics, Biological Function, and Their Involvement in Human Diseases." *Cellular and Molecular Life Sciences: CMLS* 58 (14): 2085–97.

- Li, Feng, Olga V. Glinskii, Jianjun Zhou, Landon S. Wilson, Stephen Barnes, Douglas C. Anthony, and Vladislav V. Glinsky. 2011. "Identification and Analysis of Signaling Networks Potentially Involved in Breast Carcinoma Metastasis to the Brain." *PLoS ONE* 6 (7): e21977. doi:10.1371/journal.pone.0021977.

- Lin, C. H., S. Y. Hsieh, I. S. Sheen, W. C. Lee, T. C. Chen, W. C. Shyu, and Y. F. Liaw. 2001. "Genome-Wide Hypomethylation in Hepatocellular Carcinogenesis." *Cancer Research* 61 (10): 4238–43.

- Lindemann, Carina, Oliver Hackmann, Sabit Delic, Natalie Schmidt, Guido Reifenberger, and Markus J. Riemenschneider. 2011. "SOCS3 Promoter Methylation Is Mutually Exclusive to EGFR Amplification in Gliomas and Promotes Glioma Cell Invasion through STAT3 and FAK Activation." *Acta Neuropathologica* 122 (2): 241–51. doi:10.1007/s00401-011-0832-0.

- Lin, Ya-Wen, Chun-Ming Tsao, Pei-Ning Yu, Yu-Lueng Shih, Chia-Hsin Lin, and Ming-De Yan. 2013. "SOX1 Suppresses Cell Growth and Invasion in Cervical Cancer." *Gynecologic Oncology* 131 (1): 174–81. doi:10.1016/j.ygyno.2013.07.111.

- Li, Qifeng, Ke Shen, Yang Zhao, Xiaoguang He, Chenkai Ma, Lin Wang, Baocheng Wang, Jianwen Liu, and Jie Ma. 2013. "MicroRNA-222 Promotes Tumorigenesis via Targeting DKK2 and Activating the Wnt/β-Catenin Signaling Pathway." *FEBS Letters* 587 (12): 1742–48. doi:10.1016/j.febslet.2013.04.002.

- Little, Annette S., Kathryn Balmanno, Matthew J. Sale, Scott Newman, Jonathan R. Dry, Mark Hampson, Paul A. W. Edwards, Paul D. Smith, and Simon J. Cook. 2011. "Amplification of the Driving Oncogene, KRAS or BRAF, Underpins Acquired Resistance to MEK1/2 Inhibitors in Colorectal Cancer Cells." *Science Signaling* 4 (166): ra17–ra17. doi:10.1126/scisignal.2001752.

- Liu, Bei, Laura H. Tang, Zhaojun Liu, Mei Mei, Run Yu, Deepti Dhall, Xin-Wei Qiao, et al. 2014. "A-Internexin: A Novel Biomarker for Pancreatic Neuroendocrine Tumor Aggressiveness." *The Journal of Clinical Endocrinology and Metabolism* 99 (5): E786–95. doi:10.1210/jc.2013-2874.

- Liu, Bo-lin, Jin-xiang Cheng, Xiang Zhang, Rui Wang, Wei Zhang, Hong Lin, Xian Xiao, Sang Cai, Xiao-yan Chen, and Hong Cheng. 2010. "Global Histone Modification Patterns as Prognostic Markers to Classify Glioma Patients." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 19 (11): 2888–96. doi:10.1158/1055-9965.EPI-10-0454.

- Liu, Yanhong, Sanjay Shete, Carol J. Etzel, Michael Scheurer, George Alexiou, Georgina Armstrong, Spyros Tsavachidis, et al. 2010. "Polymorphisms of LIG4, BTBD2, HMGA2, and RTEL1 Genes Involved in the Double-Strand Break Repair Pathway Predict Glioblastoma Survival." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 28 (14): 2467–74. doi:10.1200/JCO.2009.26.6213.

- Losman, Julie-Aurore, and William G. Kaelin. 2013. "What a Difference a Hydroxyl Makes: Mutant IDH, (R)-2-Hydroxyglutarate, and Cancer." *Genes & Development* 27 (8): 836–52. doi:10.1101/gad.217406.113.

- Louis, David N., Hiroko Ohgaki, Otmar D. Wiestler, Webster K. Cavenee, Peter C. Burger, Anne Jouvet, Bernd W. Scheithauer, and Paul Kleihues. 2007a. "The 2007 WHO Classification of Tumours of the Central Nervous System." *Acta Neuropathologica* 114 (2): 97–109. doi:10.1007/s00401-007-0243-4.

- Louis, David N, Hiroko Ohgaki, Otmar D Wiestler, Webster K Cavenee, Peter C Burger, Anne Jouvet, Bernd W Scheithauer, and Paul Kleihues. 2007b. "The 2007 WHO Classification of Tumours of the Central Nervous System." *Acta Neuropathologica* 114 (2): 97–109. doi:10.1007/s00401-007-0243-4.

- MacArthur, Daniel G, Suganthi Balasubramanian, Adam Frankish, Ni Huang, James Morris, Klaudia Walter, Luke Jostins, et al. 2012. "A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes." *Science (New York, N.Y.)* 335 (6070): 823–28. doi:10.1126/science.1215040.

- MacArthur, Daniel G, and Chris Tyler-Smith. 2010. "Loss-of-Function Variants in the Genomes of Healthy Humans." *Human Molecular Genetics* 19 (R2): R125–30. doi:10.1093/hmg/ddq365.

- MacCollin, Mia, Vijaya Ramesh, Lee B. Jacoby, David N. Louis, Mari-Paz Rubio, Karen Pulaski, James A. Trofatter, et al. 1994. "Mutational Analysis of Patients with Neurofibromatosis 2." *American Journal of Human Genetics* 55 (2): 314–20.

- Maksimovic, Jovana, Lavinia Gordon, and Alicia Oshlack. 2012. "SWAN: Subset-Quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips." *Genome Biology* 13 (6): R44. doi:10.1186/gb-2012-13-6-r44.

- Malzkorn, Bastian, Marietta Wolter, Markus J. Riemenschneider, and Guido Reifenberger. 2011. "Unraveling the Glioma Epigenome: From Molecular Mechanisms to Novel Biomarkers and Therapeutic Targets." *Brain Pathology (Zurich, Switzerland)* 21 (6): 619–32. doi:10.1111/j.1750-3639.2011.00536.x.

- Marsit, Carmen J., E. Andres Houseman, Brock C. Christensen, Karen Eddy, Raphael Bueno, David J. Sugarbaker, Heather H. Nelson, Margaret R. Karagas, and Karl T. Kelsey. 2006. "Examination of a CpG Island Methylator Phenotype and Implications of Methylation Profiles in Solid Tumors." *Cancer Research* 66 (21): 10621–29. doi:10.1158/0008-5472.CAN-06-1687.

- Martín, Berta, Ramón Aragüés, Rebeca Sanz, Baldo Oliva, Susana Boluda, Antonio Martínez, and Angels Sierra. 2008. "Biological Pathways Contributing to Organ-Specific Phenotype of Brain Metastatic Cells." *Journal of Proteome Research* 7 (3): 908–20. doi:10.1021/pr070426d.

- Martinez, Ramon, Jose I. Martin-Subero, Veit Rohde, Matthias Kirsch, Miguel Alaminos, Agustin F. Fernandez, Santiago Ropero, Gabriele Schackert, and Manel Esteller. 2009. "A Microarray-Based DNA Methylation Study of Glioblastoma Multiforme." *Epigenetics: Official Journal of the DNA Methylation Society* 4 (4): 255–64.

- McLendon, Roger, Allan Friedman, Darrell Bigner, Erwin G. Van Meir, Daniel J. Brat, Gena M. Mastrogianakis, Jeffrey J. Olson, et al. 2008. "Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways." *Nature* 455 (7216): 1061–68. doi:10.1038/nature07385.

- McRonald, Fiona E., Mark R. Morris, Dean Gentle, Laura Winchester, Dilair Baban, Jiannis Ragoussis, Noel W. Clarke, et al. 2009. "CpG Methylation Profiling in VHL Related and VHL Unrelated Renal Cell Carcinoma." *Molecular Cancer* 8 (1): 31. doi:10.1186/1476-4598-8-31.

- Melissa S Derycke, Shanaka R. Gunawardena. 2013. "Identification of Novel Variants in Colorectal Cancer Families by High-Throughput Exome Sequencing." *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*. doi:10.1158/1055-9965.EPI-12-1226.

- Mellai, Marta, Valentina Caldera, Laura Annovazzi, and Davide Schiffer. 2013. "The Distribution and Significance of IDH Mutations in Gliomas." In *Evolution of the Molecular Biology of Brain Tumors and the Therapeutic Implications*, edited by Terry Lichtor. InTech. http://www.intechopen.com/books/evolution-of-the-molecular-biology-of-brain-tumors-and-the-therapeutic-implications/the-distribution-and-significance-of-idh-mutations-in-gliomas#SEC33.

- Metellus, Philippe, Bema Coulibaly, Carole Colin, Andre Maues de Paula, Alexandre Vasiljevic, David Taieb, Anne Barlier, et al. 2010. "Absence of IDH Mutation Identifies a Novel Radiologic and Molecular Subtype of WHO Grade II

Gliomas with Dismal Prognosis." *Acta Neuropathologica* 120 (6): 719–29. doi:10.1007/s00401-010-0777-8.

- "Milestone 9 : Nature Milestones in Cancer." 2015. Accessed February 13. http://www.nature.com.ezproxyd.bham.ac.uk/milestones/milecancer/full/milecancer09.html.

- Miller, Gary J., Heidi L. Miller, Adrie van Bokhoven, James R. Lambert, Priya N. Werahera, Osvaldo Schirripa, M. Scott Lucia, and Steven K. Nordeen. 2003. "Aberrant HOXC Expression Accompanies the Malignant Phenotype in Human Prostate." *Cancer Research* 63 (18): 5879–88.

- Miller, Rupert, and David Siegmund. 1982. "Maximally Selected Chi Square Statistics." *Biometrics* 38 (4): 1011. doi:10.2307/2529881.

- Minn, Andy J., Gaorav P. Gupta, Peter M. Siegel, Paula D. Bos, Weiping Shu, Dilip D. Giri, Agnes Viale, Adam B. Olshen, William L. Gerald, and Joan Massagué. 2005. "Genes That Mediate Breast Cancer Metastasis to Lung." *Nature* 436 (7050): 518–24. doi:10.1038/nature03799.

- Minn, Andy J., Yibin Kang, Inna Serganova, Gaorav P. Gupta, Dilip D. Giri, Mikhail Doubrovin, Vladimir Ponomarev, William L. Gerald, Ronald Blasberg, and Joan Massagué. 2005. "Distinct Organ-Specific Metastatic Potential of Individual Breast Cancer Cells and Primary Tumors." *Journal of Clinical Investigation* 115 (1): 44–55. doi:10.1172/JCI22320.

- Mizuno, Hideaki, Kunio Kitada, Kenta Nakai, and Akinori Sarai. 2009. "PrognoScan: A New Database for Meta-Analysis of the Prognostic Value of Genes." *BMC Medical Genomics* 2 (1): 18. doi:10.1186/1755-8794-2-18.

- Mizuno, Kotaro, Hirotaka Osada, Hiroyuki Konishi, Yoshio Tatematsu, Yasushi Yatabe, Tetsuya Mitsudomi, Yoshitaka Fujii, and Takashi Takahashi. 2002. "Aberrant Hypermethylation of the CHFR Prophase Checkpoint Gene in Human Lung Cancers." *Oncogene* 21 (15): 2328–33. doi:10.1038/sj.onc.1205402.

- Molnar, Peter. 2011. "Classification of Primary Brain Tumors: Molecular Aspects." In *Management of CNS Tumors*, edited by Miklos Garami. InTech. http://www.intechopen.com/books/management-of-cns-tumors/classification-of-primary-brain-tumors-molecular-aspects.

- Mukasa, Akitake, Shunsaku Takayanagi, Kuniaki Saito, Junji Shibahara, Yusuke Tabei, Kazuhide Furuya, Takafumi Ide, et al. 2012. "Significance of IDH Mutations Varies with Tumor Histology, Grade, and Genetics in Japanese Glioma Patients." *Cancer Science* 103 (3): 587–92. doi:10.1111/j.1349-7006.2011.02175.x.

- Nakamura, Mitsutoshi, Eiwa Ishida, Keiji Shimada, Munehiro Kishi, Hiroyuki Nakase, Toshisuke Sakaki, and Noboru Konishi. 2005. "Frequent LOH on 22q12.3 and TIMP-3 Inactivation Occur in the Progression to Secondary Glioblastomas." *Laboratory Investigation; a Journal of Technical Methods and Pathology* 85 (2): 165–75. doi:10.1038/labinvest.3700223.

- Nakamura, Mitsutoshi, Takao Watanabe, Yasuhiro Yonekawa, Paul Kleihues, and Hiroko Ohgaki. 2001. "Promoter Methylation of the DNA Repair Gene MGMT in Astrocytomas Is Frequently Associated with G:C → A:T Mutations of the TP53 Tumor Suppressor Gene." *Carcinogenesis* 22 (10): 1715–19. doi:10.1093/carcin/22.10.1715.

- Nakamura, M., T. Watanabe, U. Klangby, C. Asker, K. Wiman, Y. Yonekawa, P. Kleihues, and H. Ohgaki. 2001. "p14ARF Deletion and Methylation in Genetic Pathways to Glioblastomas." *Brain Pathology (Zurich, Switzerland)* 11 (2): 159–68.

- Nakamura, M., Y. Yonekawa, P. Kleihues, and H. Ohgaki. 2001. "Promoter Hypermethylation of the RB1 Gene in Glioblastomas." *Laboratory Investigation; a Journal of Technical Methods and Pathology* 81 (1): 77–82.

- Network, The Cancer Genome Atlas. 2012. "Comprehensive Molecular Characterization of Human Colon and Rectal Cancer." *Nature* 487 (7407): 330–37. doi:10.1038/nature11252.

- Nichols, B. J., A. C. Perry, L. Hall, and R. M. Denton. 1995. "Molecular Cloning and Deduced Amino Acid Sequences of the Alpha- and Beta- Subunits of Mammalian NAD(+)-Isocitrate Dehydrogenase." *The Biochemical Journal* 310 (Pt 3) (September): 917–22.

- Nordling, C. O. 1953. "A New Theory on Cancer-Inducing Mechanism." *British Journal of Cancer* 7 (1): 68–72.

- Noushmehr, Houtan, Daniel J. Weisenberger, Kristin Diefes, Heidi S. Phillips, Kanan Pujara, Benjamin P. Berman, Fei Pan, et al. 2010. "Identification of a CpG Island Methylator Phenotype That Defines a Distinct Subgroup of Glioma." *Cancer Cell* 17 (5): 510–22. doi:10.1016/j.ccr.2010.03.017.

- Ogino, Shuji, Takako Kawasaki, Gregory J. Kirkner, Massimo Loda, and Charles S. Fuchs. 2006. "CpG Island Methylator Phenotype-Low (CIMP-Low) in Colorectal Cancer: Possible Associations with Male Sex and KRAS Mutations." *The Journal of Molecular Diagnostics: JMD* 8 (5): 582–88. doi:10.2353/jmoldx.2006.060082.

- Ongen, Halit, Claus L. Andersen, Jesper B. Bramsen, Bodil Oster, Mads H. Rasmussen, Pedro G. Ferreira, Juan Sandoval, et al. 2014. "Putative Cis-Regulatory Drivers in Colorectal Cancer." *Nature* 512 (7512): 87–90. doi:10.1038/nature13602.

- Paget, S. 1989. "Paget,stephen Paper Reproduced from the Lancet, 1889." *Cancer and Metastasis Reviews* 8 (2): 98–101.

- Palmieri, Diane, Quentin R. Smith, Paul R. Lockman, Julie Bronder, Brunilde Gril, Ann F. Chambers, Robert J. Weil, and Patricia S. Steeg. 2007. "Brain Metastases of Breast Cancer." *Breast Disease* 26 (1): 139–47.

- Park, Hongryeol, Hyei Yoon Jung, Hyun-Jung Choi, Dong Young Kim, Ji-Young Yoo, Chae-Ok Yun, Jeong-Ki Min, Young-Myoung Kim, and Young-Guen Kwon. 2014. "Distinct Roles of DKK1 and DKK2 in Tumor Angiogenesis." *Angiogenesis* 17 (1): 221–34. doi:10.1007/s10456-013-9390-5.

- Parsons, D Williams, Siân Jones, Xiaosong Zhang, Jimmy Cheng-Ho Lin, Rebecca J Leary, Philipp Angenendt, Parminder Mankoo, et al. 2008a. "An Integrated Genomic Analysis of Human Glioblastoma Multiforme." *Science (New York, N.Y.)* 321 (5897): 1807–12. doi:10.1126/science.1164382.

- Parsons, D. Williams, Siân Jones, Xiaosong Zhang, Jimmy Cheng-Ho Lin, Rebecca J. Leary, Philipp Angenendt, Parminder Mankoo, et al. 2008b. "An Integrated Genomic Analysis of Human Glioblastoma Multiforme." *Science* 321 (5897): 1807–12. doi:10.1126/science.1164382.

- Parsons, D. Williams, Meng Li, Xiaosong Zhang, Siân Jones, Rebecca J. Leary, Jimmy Cheng-Ho Lin, Simina M. Boca, et al. 2011. "The Genetic Landscape of the Childhood Cancer Medulloblastoma." *Science (New York, N.Y.)* 331 (6016): 435–39. doi:10.1126/science.1198056.

- Patchell, Roy A. 2003. "The Management of Brain Metastases." *Cancer Treatment Reviews* 29 (6): 533–40. doi:10.1016/S0305-7372(03)00105-1.

- Patel, Zubin H., Leah C. Kottyan, Sara Lazaro, Marc S. Williams, David H. Ledbetter, hbGerard Tromp, Andrew Rupert, et al. 2014. "The Struggle to Find Reliable Results in Exome Sequencing Data: Filtering out Mendelian Errors." *Frontiers in Genetics* 5 (February). doi:10.3389/fgene.2014.00016.

- Paugh, Barbara S., Chunxu Qu, Chris Jones, Zhaoli Liu, Martyna Adamowicz-Brice, Junyuan Zhang, Dorine A. Bax, et al. 2010. "Integrated Molecular Genetic Profiling of Pediatric High-Grade Gliomas Reveals Key Differences with the

Adult Disease." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 28 (18): 3061–68. doi:10.1200/JCO.2009.26.7252.

- Perri, Patrizia, Tiziana Bachetti, Luca Longo, Ivana Matera, Marco Seri, Gian Paolo Tonini, and Isabella Ceccherini. 2005. "PHOX2B Mutations and Genetic Predisposition to Neuroblastoma." *Oncogene* 24 (18): 3050–53. doi:10.1038/sj.onc.1208532.

- Pfeifer, Gerd P., Swati Kadam, and Seung-Gi Jin. 2013. "5-Hydroxymethylcytosine and Its Potential Roles in Development and Cancer." *Epigenetics & Chromatin* 6 (1): 10. doi:10.1186/1756-8935-6-10.

- Piotrowski, Arkadiusz, Jing Xie, Ying F. Liu, Andrzej B. Poplawski, Alicia R. Gomes, Piotr Madanecki, Chuanhua Fu, et al. 2014. "Germline Loss-of-Function Mutations in LZTR1 Predispose to an Inherited Disorder of Multiple Schwannomas." *Nature Genetics* 46 (2): 182–87. doi:10.1038/ng.2855.

- Plumb, J. A., G. Strathdee, J. Sludden, S. B. Kaye, and R. Brown. 2000. "Reversal of Drug Resistance in Human Tumor Xenografts by 2'-Deoxy-5-Azacytidine-Induced Demethylation of the hMLH1 Gene Promoter." *Cancer Research* 60 (21): 6039–44.

- Pollack, Ian F., Ronald L. Hamilton, Robert W. Sobol, Marina N. Nikiforova, Maureen A. Lyons-Weiler, William A. LaFramboise, Peter C. Burger, et al. 2011. "IDH1 Mutations Are Common in Malignant Gliomas Arising in Adolescents: A Report from the Children's Oncology Group." *Child's Nervous System: ChNS:*

Official Journal of the International Society for Pediatric Neurosurgery 27 (1): 87–94. doi:10.1007/s00381-010-1264-1.

- Puente, Xose S., Magda Pinyol, Víctor Quesada, Laura Conde, Gonzalo R. Ordóñez, Neus Villamor, Georgia Escaramis, et al. 2011. "Whole-Genome Sequencing Identifies Recurrent Mutations in Chronic Lymphocytic Leukaemia." *Nature* 475 (7354): 101–5. doi:10.1038/nature10113.

- Pusch, S., F. Sahm, J. Meyer, M. Mittelbronn, C. Hartmann, and A. von Deimling. 2011. "Glioma IDH1 Mutation Patterns off the Beaten Track." *Neuropathology and Applied Neurobiology* 37 (4): 428–30. doi:10.1111/j.1365-2990.2010.01127.x.

- Quail, Michael A., Miriam Smith, Paul Coupland, Thomas D. Otto, Simon R. Harris, Thomas R. Connor, Anna Bertoni, Harold P. Swerdlow, and Yong Gu. 2012. "A Tale of Three next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers." *BMC Genomics* 13 (1): 341. doi:10.1186/1471-2164-13-341.

- Rabbani, Bahareh, Mustafa Tekin, and Nejat Mahdieh. 2014. "The Promise of Whole-Exome Sequencing in Medical Genetics." *Journal of Human Genetics* 59 (1): 5–15. doi:10.1038/jhg.2013.114.

- Rahmathulla, Gazanfar, Steven A. Toms, and Robert J. Weil. 2012. "The Molecular Biology of Brain Metastasis." *Journal of Oncology* 2012 (March): e723541. doi:10.1155/2012/723541.

- Rajalingam, Krishnaraj, Ralf Schreck, Ulf R. Rapp, and Štefan Albert. 2007. "Ras Oncogenes and Their Downstream Targets." *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, Mitogen-Activated Protein Kinases: New Insights on Regulation, Function and Role in Human Disease, 1773 (8): 1177–95. doi:10.1016/j.bbamcr.2007.01.012.

- Rajendran, Ganeshkumar, Karthik Shanmuganandam, Ameya Bendre, Dattatraya Muzumdar, Dattatreya Mujumdar, Abhay Goel, and Anjali Shiras. 2011. "Epigenetic Regulation of DNA Methyltransferases: DNMT1 and DNMT3B in Gliomas." *Journal of Neuro-Oncology* 104 (2): 483–94. doi:10.1007/s11060-010-0520-2.

- Ramaswamy, Sridhar, Ken N. Ross, Eric S. Lander, and Todd R. Golub. 2003. "A Molecular Signature of Metastasis in Primary Solid Tumors." *Nature Genetics* 33 (1): 49–54. doi:10.1038/ng1060.

- Reyngold, Marsha, Sevin Turcan, Dilip Giri, Kasthuri Kannan, Logan A. Walsh, Agnes Viale, Marija Drobnjak, Linda T. Vahdat, William Lee, and Timothy A. Chan. 2014. "Remodeling of the Methylation Landscape in Breast Cancer Metastasis." *PLoS ONE* 9 (8): e103896. doi:10.1371/journal.pone.0103896.

- Riemenschneider, Markus J, Judith W M Jeuken, Pieter Wesseling, and Guido Reifenberger. 2010a. "Molecular Diagnostics of Gliomas: State of the Art." *Acta Neuropathologica* 120 (5): 567–84. doi:10.1007/s00401-010-0736-4.

- Riemenschneider, Markus J., Judith W. M. Jeuken, Pieter Wesseling, and Guido Reifenberger. 2010b. "Molecular Diagnostics of Gliomas: State of the Art." *Acta Neuropathologica* 120 (5): 567–84. doi:10.1007/s00401-010-0736-4.

- Rodriguez, Jairo, Laura Vives, Mireia Jordà, Cristina Morales, Mar Muñoz, Elisenda Vendrell, and Miguel A. Peinado. 2008. "Genome-Wide Tracking of Unmethylated DNA Alu Repeats in Normal and Cancer Cells." *Nucleic Acids Research* 36 (3): 770–84. doi:10.1093/nar/gkm1105.

- Roessler, Jessica, Ole Ammerpohl, Jana Gutwein, Britta Hasemeier, Sumadi Lukman Anwar, Hans Kreipe, and Ulrich Lehmann. 2012. "Quantitative Cross-Validation and Content Analysis of the 450k DNA Methylation Array from Illumina, Inc." *BMC Research Notes* 5: 210. doi:10.1186/1756-0500-5-210.

- Roman-Gomez, Jose, Antonio Jimenez-Velasco, Xabier Agirre, Juan A. Castillejo, German Navarro, Maria J. Calasanz, Leire Garate, et al. 2006. "CpG Island Methylator Phenotype Redefines the Prognostic Effect of t(12;21) in Childhood Acute Lymphoblastic Leukemia." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 12 (16): 4845–50. doi:10.1158/1078-0432.CCR-05-2592.

- Rousseau, Guillaume, Tetsuro Noguchi, Violaine Bourdon, Hagay Sobol, and Sylviane Olschwang. 2011. "SMARCB1/INI1 Germline Mutations Contribute to 10% of Sporadic Schwannomatosis." *BMC Neurology* 11: 9. doi:10.1186/1471-2377-11-9.

- Salhia, Bodour, Jeff Kiefer, Julianna T. D. Ross, Raghu Metapally, Rae Anne Martinez, Kyle N. Johnson, Danielle M. DiPerna, et al. 2014. "Integrated Genomic and Epigenomic Analysis of Breast Cancer Brain Metastasis." *PLoS ONE* 9 (1): e85448. doi:10.1371/journal.pone.0085448.

- Sanson, Marc, Yannick Marie, Sophie Paris, Ahmed Idbaih, Julien Laffaire, François Ducray, Soufiane El Hallani, et al. 2009. "Isocitrate Dehydrogenase 1 Codon 132 Mutation Is an Important Prognostic Biomarker in Gliomas." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 27 (25): 4150–54. doi:10.1200/JCO.2009.21.9832.

- Sarraf, Shireen A., and Irina Stancheva. 2004. "Methyl-CpG Binding Protein MBD1 Couples Histone H3 Methylation at Lysine 9 by SETDB1 to DNA Replication and Chromatin Assembly." *Molecular Cell* 15 (4): 595–605. doi:10.1016/j.molcel.2004.06.043.

- Sawada, Shun'ichi, Scott Florell, Smita M. Purandare, Mayumi Ota, Karen Stephens, and David Viskochil. 1996. "Identification of NF1 Mutations in Both Alleles of a Dermal Neurofibroma." *Nature Genetics* 14 (1): 110–12. doi:10.1038/ng0996-110.

- Schaefer, Annika, Monika Jung, Hans-Joachim Mollenkopf, Ina Wagner, Carsten Stephan, Florian Jentzmik, Kurt Miller, Michael Lein, Glen Kristiansen, and Klaus Jung. 2010. "Diagnostic and Prognostic Implications of microRNA Profiling in Prostate Carcinoma." *International Journal of Cancer. Journal International Du Cancer* 126 (5): 1166–76. doi:10.1002/ijc.24827.

- Scott, Gary K., Andrei Goga, Dipa Bhaumik, Crystal E. Berger, Christopher S. Sullivan, and Christopher C. Benz. 2007. "Coordinate Suppression of ERBB2 and ERBB3 by Enforced Expression of Micro-RNA miR-125a or miR-125b." *The Journal of Biological Chemistry* 282 (2): 1479–86. doi:10.1074/jbc.M609383200.

- Sekido, Yoshitaka. 2001. "Molecular Genetics of Malignant Mesothelioma." In *eLS*. John Wiley & Sons, Ltd. http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0022448/abstract.

- Shain, A. Hunter, and Jonathan R. Pollack. 2013. "The Spectrum of SWI/SNF Mutations, Ubiquitous in Human Cancers." *PLoS ONE* 8 (1): e55119. doi:10.1371/journal.pone.0055119.

- Shen, Lanlan, Minoru Toyota, Yutaka Kondo, E. Lin, Li Zhang, Yi Guo, Natalie Supunpong Hernandez, et al. 2007. "Integrated Genetic and Epigenetic Analysis Identifies Three Different Subclasses of Colon Cancer." *Proceedings of the National Academy of Sciences of the United States of America* 104 (47): 18654–59. doi:10.1073/pnas.0704652104.

- Shih, Yu-Lueng, Chung-Bao Hsieh, Ming-De Yan, Chun-Ming Tsao, Tsai-Yuan Hsieh, Chang-Hsin Liu, and Ya-Wen Lin. 2013. "Frequent Concomitant Epigenetic Silencing of SOX1 and Secreted Frizzled-Related Proteins (SFRPs) in Human Hepatocellular Carcinoma." *Journal of Gastroenterology and Hepatology* 28 (3): 551–59. doi:10.1111/jgh.12078.

- Shi, Xu-Bao, Lingru Xue, Joy Yang, Ai-Hong Ma, Jianjun Zhao, Ma Xu, Clifford G. Tepper, Christopher P. Evans, Hsing-Jien Kung, and Ralph W. deVere White.

2007. "An Androgen-Regulated miRNA Suppresses Bak1 Expression and Induces Androgen-Independent Growth of Prostate Cancer Cells." *Proceedings of the National Academy of Sciences* 104 (50): 19983–88. doi:10.1073/pnas.0706641104.

- Smith, Miriam J., Andrew J. Wallace, Naomi L. Bowers, Cecilie F. Rustad, C. Geoff Woods, Guy D. Leschziner, Rosalie E. Ferner, and D. Gareth R. Evans. 2012. "Frequency of SMARCB1 Mutations in Familial and Sporadic Schwannomatosis." *Neurogenetics* 13 (2): 141–45. doi:10.1007/s10048-012-0319-8.

- Smith, Temple F. 2008. "Diversity of WD-Repeat Proteins." In *The Coronin Family of Proteins*, edited by Christoph S. Clemen, Ludwig Eichinger, and Vasily Rybakin, 48:20–30. New York, NY: Springer New York. http://www.ncbi.nlm.nih.gov.ezproxyd.bham.ac.uk/books/NBK6426/.

- Stone, Annalisa R., William Bobo, Daniel J. Brat, Nara S. Devi, Erwin G. Van Meir, and Paula M. Vertino. 2004. "Aberrant Methylation and down-Regulation of TMS1/ASC in Human Glioblastoma." *The American Journal of Pathology* 165 (4): 1151–61. doi:10.1016/S0002-9440(10)63376-7.

- Strathdee, G., K. Appleton, M. Illand, D. W. Millan, J. Sargent, J. Paul, and R. Brown. 2001. "Primary Ovarian Carcinomas Display Multiple Methylator Phenotypes Involving Known Tumor Suppressor Genes." *The American Journal of Pathology* 158 (3): 1121–27. doi:10.1016/S0002-9440(10)64059-X.

- Su, Her-Young, Hung-Cheng Lai, Ya-Wen Lin, Yu-Ching Chou, Chin-Yu Liu, and Mu-Hsien Yu. 2009. "An Epigenetic Marker Panel for Screening and Prognostic Prediction of Ovarian Cancer." *International Journal of Cancer. Journal International Du Cancer* 124 (2): 387–93. doi:10.1002/ijc.23957.

- Sun, Yu-Meng, Kang-Yu Lin, and Yue-Qin Chen. 2013. "Diverse Functions of miR-125 Family in Different Cell Contexts." *Journal of Hematology & Oncology* 6 (1): 6. doi:10.1186/1756-8722-6-6.

- Supek, Fran, Belén Miñana, Juan Valcárcel, Toni Gabaldón, and Ben Lehner. 2014. "Synonymous Mutations Frequently Act as Driver Mutations in Human Cancers." *Cell* 156 (6): 1324–35. doi:10.1016/j.cell.2014.01.051.

- Svokos, Konstantina A., Bodour Salhia, and Steven A. Toms. 2014. "Molecular Biology of Brain Metastasis." *International Journal of Molecular Sciences* 15 (6): 9519–30. doi:10.3390/ijms15069519.

- Talpaz, Moshe, Neil P. Shah, Hagop Kantarjian, Nicholas Donato, John Nicoll, Ron Paquette, Jorge Cortes, et al. 2006. "Dasatinib in Imatinib-Resistant Philadelphia Chromosome-Positive Leukemias." *The New England Journal of Medicine* 354 (24): 2531–41. doi:10.1056/NEJMoa055229.

- Tang, Feng, Rui Zhang, Yunmian He, Meijuan Zou, Le Guo, and Tao Xi. 2012. "MicroRNA-125b Induces Metastasis by Targeting STARD13 in MCF-7 and MDA-MB-231 Breast Cancer Cells." *PLoS ONE* 7 (5): e35435. doi:10.1371/journal.pone.0035435.

- Tepel, Martin, Peter Roerig, Marietta Wolter, David H. Gutmann, Arie Perry, Guido Reifenberger, and Markus J. Riemenschneider. 2008. "Frequent Promoter Hypermethylation and Transcriptional Downregulation of the NDRG2 Gene at 14q11.2 in Primary Glioblastoma." *International Journal of Cancer. Journal International Du Cancer* 123 (9): 2080–86. doi:10.1002/ijc.23705.

- Tew, Kenneth D., and Paul B. Fisher. 2014. *Advances in Cancer Research*. Academic Press.

- "The Cancer Genome Atlas." 2015. *The Cancer Genome Atlas - National Cancer Institute*. Accessed February 10. http://cancergenome.nih.gov/.

- Torres-Martin, Miguel, Luis Lassaletta, Jose M. de Campos, Alberto Isla, Javier Gavilan, Giovanny R. Pinto, Rommel R. Burbano, et al. 2013. "Global Profiling in Vestibular Schwannomas Shows Critical Deregulation of MicroRNAs and Upregulation in Those Included in Chromosomal Region 14q32." *PLoS ONE* 8 (6): e65868. doi:10.1371/journal.pone.0065868.

- Torres-Martin, Miguel, Luis Lassaletta, Jesus San-Roman-Montero, Jose M De Campos, Alberto Isla, Javier Gavilan, Barbara Melendez, et al. 2013. "Microarray Analysis of Gene Expression in Vestibular Schwannomas Reveals SPP1/MET Signaling Pathway and Androgen Receptor Deregulation." *International Journal of Oncology* 42 (3): 848–62. doi:10.3892/ijo.2013.1798.

- Toyota, M., N. Ahuja, M. Ohe-Toyota, J. G. Herman, S. B. Baylin, and J. P. Issa. 1999. "CpG Island Methylator Phenotype in Colorectal Cancer." *Proceedings of*

the National Academy of Sciences of the United States of America 96 (15): 8681–86.

- Toyota, Minoru, Mutsumi Ohe-Toyota, Nita Ahuja, and Jean-Pierre J. Issa. 2000. "Distinct Genetic Profiles in Colorectal Tumors with or without the CpG Island Methylator Phenotype." *Proceedings of the National Academy of Sciences of the United States of America* 97 (2): 710–15.

- Tsao, Chun-Ming, Ming-De Yan, Yu-Lueng Shih, Pei-Ning Yu, Chih-Chi Kuo, Wen-Chi Lin, Hsin-Jung Li, and Ya-Wen Lin. 2012. "SOX1 Functions as a Tumor Suppressor by Antagonizing the WNT/β-Catenin Signaling Pathway in Hepatocellular Carcinoma." *Hepatology* 56 (6): 2277–87. doi:10.1002/hep.25933.

- Turcan, Sevin, Armida W. M. Fabius, Alexandra Borodovsky, Alicia Pedraza, Cameron Brennan, Jason Huse, Agnes Viale, Gregory J. Riggins, and Timothy A. Chan. 2013. "Efficient Induction of Differentiation and Growth Inhibition in IDH1 Mutant Glioma Cells by the DNMT Inhibitor Decitabine." *Oncotarget* 4 (10): 1729–36.

- Turcan, Sevin, Daniel Rohle, Anuj Goenka, Logan A. Walsh, Fang Fang, Emrullah Yilmaz, Carl Campos, et al. 2012a. "IDH1 Mutation Is Sufficient to Establish the Glioma Hypermethylator Phenotype." *Nature* 483 (7390): 479–83. doi:10.1038/nature10866.

- ———. 2012b. "IDH1 Mutation Is Sufficient to Establish the Glioma Hypermethylator Phenotype." *Nature* 483 (7390): 479–83. doi:10.1038/nature10866.

- Valastyan, Scott, and Robert A. Weinberg. 2011. "Tumor Metastasis: Molecular Insights and Evolving Paradigms." *Cell* 147 (2): 275–92. doi:10.1016/j.cell.2011.09.024.

- Van den Bent, Martin J., Lonneke A. Gravendeel, Thierry Gorlia, Johan M. Kros, Lariesa Lapre, Pieter Wesseling, Johannes L. Teepen, et al. 2011. "A Hypermethylated Phenotype Is a Better Predictor of Survival than MGMT Methylation in Anaplastic Oligodendroglial Brain Tumors: A Report from EORTC Study 26951." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 17 (22): 7148–55. doi:10.1158/1078-0432.CCR-11-1274.

- Várkondi, Edit, Ferenc Gyory, András Nagy, István Kiss, István Ember, and László Kozma. 2005. "Oncogene Amplification and Overexpression of Oncoproteins in Thyroid Papillary Cancer." *In Vivo (Athens, Greece)* 19 (2): 465–70.

- Ventura, Andrea, David G. Kirsch, Margaret E. McLaughlin, David A. Tuveson, Jan Grimm, Laura Lintault, Jamie Newman, Elizabeth E. Reczek, Ralph Weissleder, and Tyler Jacks. 2007. "Restoration of p53 Function Leads to Tumour Regression in Vivo." *Nature* 445 (7128): 661–65. doi:10.1038/nature05541.

- Vogan, Kyle. 2013. "Recurrent Mutations in Meningiomas." *Nature Genetics* 45 (3): 233–233. doi:10.1038/ng.2576.

- Wang, Dan, Li Yan, Qiang Hu, Lara E. Sucheston, Michael J. Higgins, Christine B. Ambrosone, Candace S. Johnson, Dominic J. Smiraglia, and Song Liu. 2012. "IMA: An R Package for High-Throughput Analysis of Illumina's 450K Infinium Methylation Data." *Bioinformatics* 28 (5): 729–30. doi:10.1093/bioinformatics/bts013.

- Wang, Hongjiang, Guang Tan, Lei Dong, Lei Cheng, Kejun Li, Zhongyu Wang, and Haifeng Luo. 2012. "Circulating MiR-125b as a Marker Predicting Chemoresistance in Breast Cancer." *PLoS ONE* 7 (4): e34210. doi:10.1371/journal.pone.0034210.

- Wang, Pu, Qiongzhu Dong, Chong Zhang, Pei-Fen Kuan, Yufeng Liu, William R. Jeck, Jesper B. Andersen, et al. 2013. "Mutations in Isocitrate Dehydrogenase 1 and 2 Occur Frequently in Intrahepatic Cholangiocarcinomas and Share Hypermethylation Targets with Glioblastomas." *Oncogene* 32 (25): 3091–3100. doi:10.1038/onc.2012.315.

- Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. "RNA-Seq: A Revolutionary Tool for Transcriptomics." *Nature Reviews. Genetics* 10 (1): 57–63. doi:10.1038/nrg2484.

- Wan, Liling, Klaus Pantel, and Yibin Kang. 2013. "Tumor Metastasis: Moving New Biological Insights into the Clinic." *Nature Medicine* 19 (11): 1450–64. doi:10.1038/nm.3391.

- Watt, Paul M., Rolee Kumar, and Ursula R. Kees. 2000. "Promoter Demethylation Accompanies Reactivation of the HOX11 Proto-Oncogene in

Leukemia." *Genes, Chromosomes and Cancer* 29 (4): 371–77. doi:10.1002/1098-2264(2000)9999:9999<::AID-GCC1050>3.0.CO;2-Y.

- Webb, Tracy. 2003. "Microarray Studies Challenge Theories of Metastasis." *Journal of the National Cancer Institute* 95 (5): 350–51. doi:10.1093/jnci/95.5.350.

- Wei, Peng, Roberto Pattarini, Yongqi Rong, Hong Guo, Parmil K. Bansal, Sheila V. Kusnoor, Ariel Y. Deutch, Jennifer Parris, and James I. Morgan. 2012. "The Cbln Family of Proteins Interact with Multiple Signaling Pathways." *Journal of Neurochemistry* 121 (5): 717–29. doi:10.1111/j.1471-4159.2012.07648.x.

- Weller, Michael, Jörg Felsberg, Christian Hartmann, Hilmar Berger, Joachim P. Steinbach, Johannes Schramm, Manfred Westphal, et al. 2009. "Molecular Predictors of Progression-Free and Overall Survival in Patients with Newly Diagnosed Glioblastoma: A Prospective Translational Study of the German Glioma Network." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 27 (34): 5743–50. doi:10.1200/JCO.2009.23.0805.

- Weller, Michael, Roger Stupp, Guido Reifenberger, Alba A. Brandes, Martin J. van den Bent, Wolfgang Wick, and Monika E. Hegi. 2010. "MGMT Promoter Methylation in Malignant Gliomas: Ready for Personalized Medicine?" *Nature Reviews. Neurology* 6 (1): 39–51. doi:10.1038/nrneurol.2009.197.

- Wen, Patrick Y, and Santosh Kesari. 2008. "Malignant Gliomas in Adults." *The New England Journal of Medicine* 359 (5): 492–507. doi:10.1056/NEJMra0708126.

- Wikman, Harriet, Laura Westphal, Felicitas Schmid, Sirkku Pollari, Jolanthe Kropidlowski, Bettina Sielaff-Frimpong, Markus Glatzel, et al. 2014. "Loss of CADM1 Expression Is Associated with Poor Prognosis and Brain Metastasis in Breast Cancer Patients." *Oncotarget* 5 (10): 3076–87.

- Wolff, Erika M., Yoshitomo Chihara, Fei Pan, Daniel J. Weisenberger, Kimberly D. Siegmund, Kokichi Sugano, Kiyotaka Kawashima, Peter W. Laird, Peter A. Jones, and Gangning Liang. 2010. "Unique DNA Methylation Patterns Distinguish Non-Invasive and Invasive Urothelial Cancers and Establish an Epigenetic Field Defect in Premalignant Tissue." *Cancer Research* 70 (20): 8169–78. doi:10.1158/0008-5472.CAN-10-1335.

- Wolff, R. K., K. A. Frazer, R. K. Jackler, M. J. Lanser, L. H. Pitts, and D. R. Cox. 1992. "Analysis of Chromosome 22 Deletions in Neurofibromatosis Type 2-Related Tumors." *American Journal of Human Genetics* 51 (3): 478–85.

- Wu, N., X. Lin, X. Zhao, L. Zheng, L. Xiao, J. Liu, L. Ge, and S. Cao. 2013. "MiR-125b Acts as an Oncogene in Glioblastoma Cells and Inhibits Cell Apoptosis through p53 and p38MAPK-Independent Pathways." *British Journal of Cancer* 109 (11): 2853–63. doi:10.1038/bjc.2013.672.

- Wu, Xiwei, Tibor A. Rauch, Xueyan Zhong, William P. Bennett, Farida Latif, Dietmar Krex, and Gerd P. Pfeifer. 2010. "CpG Island Hypermethylation in

Human Astrocytomas." *Cancer Research* 70 (7): 2718–27. doi:10.1158/0008-5472.CAN-09-3631.

- Xia, Hong-Fei, Tian-Zhu He, Chun-Mei Liu, Yi Cui, Pei-Pei Song, Xiao-Hua Jin, and Xu Ma. 2009. "MiR-125b Expression Affects the Proliferation and Apoptosis of Human Glioma Cells by Targeting Bmf." *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology* 23 (4-6): 347–58. doi:10.1159/000218181.

- Xiong, Zhenggang, and Peter W. Laird. 1997. "COBRA: A Sensitive and Quantitative DNA Methylation Assay." *Nucleic Acids Research* 25 (12): 2532–34. doi:10.1093/nar/25.12.2532.

- Xu, Wei, Hui Yang, Ying Liu, Ying Yang, Ping Wang, Se-Hee Kim, Shinsuke Ito, et al. 2011. "Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of A-Ketoglutarate-Dependent Dioxygenases." *Cancer Cell* 19 (1): 17–30. doi:10.1016/j.ccr.2010.12.014.

- Xu, Xiang, Jingyue Zhao, Zhen Xu, Baozhen Peng, Qiuhua Huang, Eddy Arnold, and Jianping Ding. 2004. "Structures of Human Cytosolic NADP-Dependent Isocitrate Dehydrogenase Reveal a Novel Self-Regulatory Mechanism of Activity." *The Journal of Biological Chemistry* 279 (32): 33946–57. doi:10.1074/jbc.M404298200.

- Yang, Yin-Xue. 2012. "Identification of Novel DNA Methylation Markers in Colorectal Cancer Using MIRA-Based Microarrays." *Oncology Reports*, April. doi:10.3892/or.2012.1779.

- Yan, Hai, D Williams Parsons, Genglin Jin, Roger McLendon, B Ahmed Rasheed, Weishi Yuan, Ivan Kos, et al. 2009a. "IDH1 and IDH2 Mutations in Gliomas." *The New England Journal of Medicine* 360 (8): 765–73. doi:10.1056/NEJMoa0808710.

- Yan, Hai, D. Williams Parsons, Genglin Jin, Roger McLendon, B. Ahmed Rasheed, Weishi Yuan, Ivan Kos, et al. 2009b. "IDH1 and IDH2 Mutations in Gliomas." *New England Journal of Medicine* 360 (8): 765–73. doi:10.1056/NEJMoa0808710.

- Yan, Xiao-Jing, Jie Xu, Zhao-Hui Gu, Chun-Ming Pan, Gang Lu, Yang Shen, Jing-Yi Shi, et al. 2011. "Exome Sequencing Identifies Somatic Mutations of DNA Methyltransferase Gene DNMT3A in Acute Monocytic Leukemia." *Nature Genetics* 43 (4): 309–15. doi:10.1038/ng.788.

- Zhang, Yan, Li-Xu Yan, Qi-Nian Wu, Zi-Ming Du, Jing Chen, Ding-Zhun Liao, Ma-Yan Huang, et al. 2011. "miR-125b Is Methylated and Functions as a Tumor Suppressor by Regulating the ETS1 Proto-Oncogene in Human Invasive Breast Cancer." *Cancer Research* 71 (10): 3552–62. doi:10.1158/0008-5472.CAN-10-2435.

- Zheng, Cui-Xia, Zhao-Hui Gu, Bing Han, Rong-Xin Zhang, Chun-Ming Pan, Yi Xiang, Xia-Jun Rong, Xia Chen, Qing-Yun Li, and Huan-Ying Wan. 2013. "Whole-Exome Sequencing to Identify Novel Somatic Mutations in Squamous Cell Lung Cancers." *International Journal of Oncology* 43 (3): 755–64. doi:10.3892/ijo.2013.1991.

- Zheng, Siyuan, Hoon Kim, and Roel G. W. Verhaak. 2014. "Silent Mutations Make Some Noise." *Cell* 156 (6): 1129–31. doi:10.1016/j.cell.2014.02.037.

- Zhi, Degui, and Rui Chen. 2012. "Statistical Guidance for Experimental Design and Data Analysis of Mutation Detection in Rare Monogenic Mendelian Diseases by Exome Sequencing." *PLoS ONE* 7 (2): e31358. doi:10.1371/journal.pone.0031358.

Chapter Nine: Peer reviewed publications

- Hill, Victoria K., **Thoraia Shinaw**i, Christopher J. Ricketts, Dietmar Krex, Gabriele Schackert, Julien Bauer, Wenbin Wei, Garth Cruickshank, Eamonn R. Maher, and Farida Latif. 2014. "Stability of the CpG Island Methylator Phenotype during Glioma Progression and Identification of Methylated Loci in Secondary Glioblastomas." *BMC Cancer* 14: 506. doi:10.1186/1471-2407-14-506.

- Mezzanotte, Jessica J., Victoria Hill, M. Lee Schmidt, **Thoraia Shinawi**, Stella Tommasi, Dietmar Krex, Gabriele Schackert, Gerd P. Pfeifer, Farida Latif, and Geoffrey J. Clark. 2014. "RASSF6 Exhibits Promoter Hypermethylation in Metastatic Melanoma and Inhibits Invasion in Melanoma Cells." *Epigenetics: Official Journal of the DNA Methylation Society* 9 (11): 1496–1503. doi:10.4161/15592294.2014.983361.

- Pangeni, Rajendra P., Prasanna Channathodiyil, David S. Huen, Lawrence W. Eagles, Balraj K. Johal, Dawar Pasha, Natasa Hadjistephanou, **Thoraia Shinawi**, et al. 2015. "The GALNT9, BNC1 and CCDC8 Genes Are Frequently Epigenetically Dysregulated in Breast Tumours That Metastasise to the Brain." *Clinical Epigenetics* 7 (1): 57. doi:10.1186/s13148-015-0089-x.

- **Thoraia Shinawi**, Victoria Hill, Antonis Dagklis, Panagiotis Baliakas, Kostas Stamatopoulos, Angleo Agathanggelou, Tanja Stankovic, Eamonn R. Maher, Paolo Ghia, and Farida Latif. 2012. "KIBRA Gene Methylation Is Associated with Unfavorable Biological Prognostic Parameters in Chronic Lymphocytic Leukemia." *Epigenetics: Official Journal of the DNA Methylation Society* 7 (3): 211–15. doi:10.4161/epi.7.3.19222.

- **Thoraia Shinawi**, Victoria K. Hill, Dietmar Krex, Gabriele Schackert, Dean Gentle, Mark R. Morris, Wenbin Wei, Garth Cruickshank, Eamonn R. Maher, and Farida Latif. 2013. "DNA Methylation Profiles of Long- and Short-Term Glioblastoma Survivors." *Epigenetics: Official Journal of the DNA Methylation Society* 8 (2): 149–56. doi:10.4161/epi.23398.