Genomic Landscape of Uveal Melanoma

Hima Anbunathan

CID:00797464
Imperial College London,
National Heart and Lung Institute

A thesis submitted for Doctor of Philosophy (PhD) degree
Faculty of Medicine, Imperial College London
2017

Abstract

Uveal melanoma (UM) is the most common cancer of the adult eye which can manifest as a highly aggressive form approximately half of the time. Here a comprehensive landscape of genetic alterations in UMs is described. It was identified by integrating copy number alterations (CNAs), and transcriptomic and whole exome sequencing data from 207 primary UMs. Focal copy number analysis with the GISTIC algorithm refined the boundaries of chromosomal segments with chromosomal gains or losses and candidate cancer genes within these segments were identified. Chromosome 8q24.3 was the region most frequently amplified in UMs, being detected in 72% of tumours. A comparison of focal copy gains and losses with that described by a pan-cancer study revealed Plectin 1 as a candidate gene within the 8q24.3 amplicon. Integration of copy number and transcriptomic data also revealed enrichment of genes within pathways leading to activation of NF-kappa B, WNT signaling and RNA splicing. Using a complementary bioinformatics approach, additional novel mutations in known dominant UM driver genes (*GNAQ*, *GNA11*, *BAP1, SF3B1*, *EIFIAX* and *CYSLTR2*) were identified and an accurate estimate of the frequencies of mutations in each gene were obtained. Finally, integration of data obtained from CNAs with mutational and transcriptome data reveled homozygous deletions, protein damaging mutations and gene fusions that targeted chromatin modifiers, and specifically genes encoding components of the human SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeling complex. Genes from the BAF complex (*ARID1A* and *ARID1B*) and the PBAF complex (*PHF10*) were subjected to functional loss through CNAs, gene fusions and mutations. Two of these chromatin modifiers (*ARID1B* and *PHF10*) map to chromosome 6q whose loss is associated with metastasis in a subset of UMs, and an *ARID1B* fusion is found in a tumour with a *BAP1* mutation that subsequently underwent metastasis. In conclusion, this study provides a comprehensive overview of the landscape of genomic alterations in UM, identifying candidate genes in regions of CNAs and providing further insights into the altered pathways of tumour development and progression.

Abbreviations

AJCC	American Joint Committee on Cancer
ALT	Alanine transaminase
ARID	ARID domain
ARID1A	AT-Rich Interaction Domain 1A
ARID1B	AT-Rich Interaction Domain 1B
ASCAT	Allelic specific copy number analysis of tumour
ASPCF	Allele-Specific Piecewise Constant Fitting
AST	Aspartate aminotransferase
AXIN1	Axis Inhibition Protein 1
BAF	B allele frequency
BAP1	BRCA1 associated protein 1
CHRPE	Congenital hypertrophy of the retinal pigment epithelium
CTBP1	C-Terminal Binding Protein 1
CYSLTR2	Cysteinyl leukotriene receptor 2
DUF3518	SWI/SNF-like complex subunit BAF250
EIF1AX	Eukaryotic translation initiation factor 1A
EIF1AY	Eukaryotic Translation Initiation Factor 1A, Y-Linked
ESE	Exonic splicing enhancer
GEP	Gene expression profiling
GNA11	G Protein Subunit Alpha 11
GNAQ	G Protein Subunit Alpha Q
GPCR	G protein-coupled receptors
HDAC	Histone deacetyase
HLA	human leukocyte antigen
IKBKB	Inhibitor of Nuclear Factor Kappa B Kinase Subunit Beta
KEGG	Kyoto Encyclopedia of Genes and Genomes
LFT	Liver function tests
LRR	Log R ratio
NCAPD3	Non-SMC Condensin II Complex Subunit D3
NSC	Lung non-small cell cancer
PA	Phosphatidic acid
PDT	Photodynamic laser photocoagulation therapy
PEHC	Hemorrhagic chorioretinopathy
PHF10	PHD Finger Protein 10
PKC	Protein kinase C
PLCB4	Phospholipase C 4
PLEC1	Plectin 1, Intermediate Filament Binding Protein
RAC3	Ras-Related C3 Botulinum Toxin Substrate 3
RPE	Retinal pigment epithelium
RPS6KA1	Ribosomal Protein S6 Kinase A1
RPS6KA2	Ribosomal Protein S6 Kinase A2
SASH1	SAM And SH3 Domain Containing 1
SCNA	Somatic copy number alterations
SF3B1	Splicing factor 3b subunit 1
SF3B5	Splicing Factor 3b Subunit 5
SIAH1	Siah E3 Ubiquitin Protein Ligase 1
SIGN	Scottish Intercollegiate Guidelines Network
TIRAP	TIR Domain Containing Adaptor Protein
TNM	Tumour node metastasis
TTT	transpupillary thermal therapy
VANGL1	VANGL Planar Cell Polarity Protein 1
WNT4	Wnt Family Member 4
WNT7B	Wnt Family Member 7B
+	Gain
-	Loss

Acknowledgements

Foremost, I want to express my most sincere gratitude and special appreciation to my supervisor Prof. Anne Bowcock who has truly been a wonderful mentor, always encouraging, sharing interesting ideas, being patient, allowing me to think independently and grow. I am extremely grateful to you for giving me this opportunity to conduct this research.

My sincere thanks go to Professor Cookson, Professor Lovett and Professor Moffatt for their insights, advice and encouragement though these years. I want to extend my gratitude to my friends and colleagues at work, Karl, Ruth, Leah, Adam, Anastasia, Spyros, Poonam, Anca, Verdiana, Giovanna, Mike, Stef, Hothri, Sharon, Sarah, Liz and everyone else in the office who have been really looking out for me through this process.

Last but not the least, I am ever grateful to my family, who always look out for me. I want to thank my parents Dr. Anitha and Dr. Anbunathan, grandma, Abi, Naveena, Mitra and rest of the family, who encouraged me to find and follow my passion and provided the support to pursue them. I want to specially thank my husband Sharath, for constantly supporting, helping and being patient with me during all the challenging times.

Declaration of Originality

I herewith certify that this dissertation is my original work and that all materials included which are not my own work have been acknowledged.

Copyright Declaration

Table of Contents

Figures

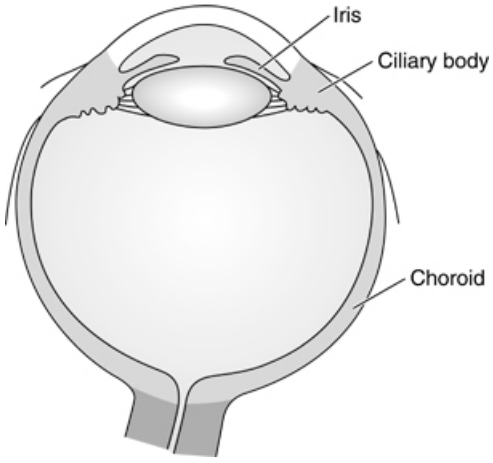
Tables

Chapter 1. Introduction

1.1 Clinical characteristics

The median age of diagnosis of UM is approximately between 55-65 years [1–3], with a peak in diagnosis in the 70-80 year age group [4]. Clinical presentation in UM depends on the size and location of the tumours. UM can develop in any area of the uveal track of the eye including the iris (anterior UM), choroid and the ciliary body (posterior UMs) **(Figure 1.1)**. Iris melanomas are clearly visible, whereas ciliary body and choroidal melanomas may escape notice without a careful fundoscopic examination. Thus, documentations of the various tumour's characteristics such as tumour thickness, and location by a colour fundus photography is important during follow-up to evaluating signs of malignant transformation. The incidence of choroidal melanoma is the highest (85%-90%) compared to ciliary body (5%-8%) and iris melanoma (3%-5%) with higher proportion of iris melanoma seen in younger patients (less than 20 years) [4,5].

Figure 1.1: The uveal tract of the eye [6].



The uveal tract of the eye comprising iris, ciliary body and choroid**.**

1.1.1 Clinical presentation

At the time of diagnosis, more than 80% of the patients present with symptoms, [2,7]. The most common include, blurred vision, defects in the visual field, photopsia, floaters, irritation, metamorphopsia, redness, change in the eye appearance and other rare clinical

feature such lipofuscin pigment [8–10] and retinal detachment [11]. Choroidal melanoma manifests as one of the 3 configurations including dome shaped in the majority of cases (75%), mushroom-shaped (20%) and diffused form (5%) [3]. When the choroidal tumour breaks through Bruch's membrane and hibernates into the sub-retinal space, it forms a bilobed structure giving rise to the mushroom-shaped choroidal melanoma. Diffused melanoma is flat in appearance and can be mistaken for a choroidal nevus. In addition, it shows a higher probability of metastasizing compared to the non-diffused tumours [12]. Ciliary body melanoma can be asymptomatic for a long time before it manifest. It can lead to displacement of the lens with consecutive refractive, accommodation balances (changes in the optical power to maintain clear image) and can cause localized cataract or increase intraocular pressure. These tumours may be associated with a dilated pupil and present as dome-shaped or sessile lesions. Iris melanoma manifests as growth of a previous iris lesion or a new pigmented spot and are usually asymptomatic. However, they are most likely to be discovered due to their visible location. Clinical features of iris melanoma include the involvement of the iridocorneal angle (the presence of tumour cells encroaching on the iris root with or without the trabecular meshwork or anterior ciliary body), prominent tumour vascularity, a distorted pupil, ectropion uveae or the presence of iris pigment epithelium on the anterior surface of the iris and rarely pigment dispersion and localized cataract [12].

1.1.2 Tumour classification and Diagnosis

Staging refers to the extent of the cancer with respect to tumour grow and spread and classification is based on clinical, pathological and genetic factors. Cancer staging follows the Tumour node metastasis (TNM) staging system of the American Joint Committee on Cancer (AJCC) for eye cancer [13]. The tumour size is evaluated and defined in the T category (from 1 to 4), lymph node involvement in the N category (includes NX, N0 and N1) and the presence of distant metastasis in the M category (including MX, M0, M1a, M1b, M1c). For posterior UM, T classification is based on tumour basal width and thickness and is then subdivided into groups reflecting ciliary body involvement and extrascleral extension of the tumour (a, b, c, d, e). This classification can predict prognosis, and the 5-year survival rate of iris melanoma patients is 56% of patients with T1 tumours, 34% for T2 tumour patients, 2% for T3 tumour patients and 1% for patients with tumours classified as T4 [14]. The rate of

metastasis in posterior UM at 5-years was found to be 8% for T1 tumours, 14% for T2 tumours, 31% for T3 tumours and 51% for T4 tumours [15].

The most standard and reliable way to make a diagnosis of choroidal melanoma is based on clinical examination by an experienced observer (Optometrist or Ophthalmologist). The ophthalmic examination can be done by slit lamp biomicroscopy, indirect ophthalmoscopy and other additional diagnostic testing such as ultrasonography, fluorescein angiography and optical coherence tomography. The accuracy of diagnosis established by clinical examination is quite high [16], however, Damato et al showed tumours were initially missed or misdiagnosed in 23% of patients in the United Kingdom, resulting in more advanced tumours and higher rate of primary enucleation was more likely in these patients who reported that their tumour was missed [2]. The differential diagnosis of UM includes around 54 different conditions that are able to mimic UM, the most frequent being choroidal nevus accounting more than 25% of psuedomelanomas [17,18]. Clinical features such as changes in drusen and retinal pigment epithelium (RPE) are more prevalent in choroidal nevi whereas retinal detachment, choroidal neovascularisation or haemorrhagic retinal detachment can occur in both, choroidal nevi and small melanomas. On B-scan Ocular Ultrasound, nevi have a high internal reflectivity; In addition, orange pigment and subretinal fluid, a feature of potential malignancy can be present in 10% and 18% of nevi respectively. Thus, differentiation between choroidal nevi and small melanomas remains a clinical challenge [19]. Other conditions for differential diagnosis include, congenital hypertrophy of the retinal pigment epithelium (CHRPE), optic disc melanocytoma, hyperplasia of the RPE, choroidal haemangioma, choroidal osteoma, peripheral exudative hemorrhagic chorioretinopathy (PEHC), choroidal haemorrhage, posterior nodular scleritis, intraocular leiomyoma and adenoma of RPE.

Histopathological Features in UM include 3 recognized categories based on H&E staining: spindle, epithelioid and mixed cell type differentiated [20]. Spindle cells exhibit elongated nuclei that may contain eosinophilic nucleoli; grow slowly and are associated with better prognosis. On the other hand, the epithelioid cells have a more polygonal cytoplasm and contain eccentric placed large pleomorphic nuclei and prominent eosinophilic nucleoli. They are fast growing with aggressive behaviour and are associated with poor prognosis. The mixed-cell type melanoma has a variable percentage of spindle and epithelioid cells with a minimum of 10% of either type. Immunohistochemistry can help with diagnosis where HMB45 antibody binds to gp100 expressed by melanocytes. This can be used to differentiate UM from non-melanocytic tumours [21].

1.1.3 Prognostic factors for metastasis of uveal melanoma

Most of the patients who present with metastatic disease have never been seen by an ophthalmologist. This is proposed to be due to the severity of other systematic medical problems so that visual related symptoms are not investigated. Fenton and Barak reported that the choroidal metastasis was found in 0% of breast cancer and only 2% of the asymptomatic choroidal metastases in patients with lung cancer [22,23]. Thus, the need for ophthalmic screening of all patients has been questioned to be able to justify the high cost of such low detection rate. At the times of diagnosis of choroidal melanoma, less than 4% of patients have detectable metastatic diseases [22]. However, despite the excellent rate of local control and treatment options, almost half of the patients eventually go on to develop metastatic disease. Early detection of metastasis in UM is critical for better management of the disease. This includes liver directed chemoembolization and recruiting patients into clinical trials of adjuvant therapy. UMs disseminate through the bloodstream (hematogenous) and the most common site of metastasis is the liver (93%), followed by the lung (24%) and bones (16%) [24]. UMs do not spread to regional lymph nodes due to lack of lymphatic drainage in the eye, except in rare cases of direct invasion of the conjunctiva leading to access to regional lymph nodes via conjunctival lymphatics [25]. Patients have better survival when they do not have liver metastasis or the primary site of metastasis is not liver [26]. The relevance of isolated or combined liver function tests (LFT) for aspartate aminotransferase (AST), alanine transaminase (ALT), yGT, LDH and phosphatidic acid (PA) was evaluated for early detection of liver metastasis in UM and found to not be helpful. However the high negative predictive value of these LFT can help clinicians to reassure patients when their LFT are negative [27]. Various predictors of metastatic outcomes and survival have been studied till date, including clinical staging, cytogenetic events, gene mutations and gene expression. These predictors have greatly increased the accuracy of predicting the outcome and eye biopsies are routinely performed following diagnosis.

Several histopathological and clinical features have been investigated to predict prognosis of UM. Histopathological features which are predictor of poor prognosis include epithelioid cell type, mitotic activity, tumour infiltration by M2-macrophages and lymphocytes, increased expression of human leukocyte antigen (HLA), extracellular matrix patterns and small blood vessel (microvascular) loops and networks [28–31]. However, these factors are not reliable enough to enable detection of high risk patients. Multivariate analysis of clinical factors that predict metastatic risk include increased tumour diameter, increasing

patient age, location of ciliary body, increasing tumour thickness, pigmented melanoma, intraocular haemorrhage and presence of sub-retinal fluid [3]. One of the most important clinical feature for predicting prognosis of UM is tumour size. The AJCC classification is based on tumour size, ciliary body involvement and extent of extraocular extension. Survival decreases rapidly with each increasing AJCC tumour stage (T1-T4). Compared to the T1 stage which has an estimated death of 4% at 5 years, 8% at 10 years and 11% at 20 years, the rate of metastasis and death for T2 is two times greater, T3 is four times greater and T4 is eight times greater [15]. Among the three sites, patients with iris melanoma have better prognosis: The frequency of metastasis at 5 and 10 years of follow is 4.1% and 6.9% respectively, compared to 15% and 25% respectively in choroidal melanoma, and 19% and 33% respectively for ciliary body melanoma [3].

Cytogenetic studies have significantly improved the prognosis in UMs. Aberrations on chromosomes 1, 3, 6 and 8 are common in UMs and are predictors of tumour behaviour [32–34]. Loss of heterozygosity on chromosome 3 (or Monosomy 3) is found in approximately in 50% of tumours and correlates with poor outcomes, where it is often present with abnormalities on chromosome 8 [32,35,36]. Monosomy 3 is found in approximately 70% of metastasizing and in approximately 20% of non-metastasizing UMs [32] and strongly correlates with metastases and decreased survival [33,35,36]. In addition, loss of the p arm of chromosome 1 (1p-) with concomitant monosomy 3 is a strong predictor of decreased disease free survival [37]. Gain of chromosome 6p (6p+) mainly occurs in the non-metastasizing tumours and tends to occur in a mutually exclusive pattern with monosomy 3 [38–40]. Cytogenetic based prognostic classifiers using status on specific chromosomal regions can also be used for identifying high-risk patients. Trolet et al [41] performed clustering analysis on microarray data from 86 primary and 66 liver metastases samples and found that disomy 3 tumours could be divided into two subgroups and that monosomy 3 tumours could be divided into three subgroups. The highest metastatic rates were found in the subgroup defined by 8p gain (8p+), monosomy 3, 8p loss (8p-) and 16q loss (16q-) and a prognostic classifier which included these chromosomal changes led to 85.9% classification accuracy.

Gene expression profiling (GEP) provides the best prediction of metastatic potential in UM. Primary UMs cluster into two distinct molecular classes based on their mRNA expression: Class 1 or low-grade tumours associated with low metastatic risk, and class 2 or high grade tumours strongly associated with high metastatic risk. The probability of survival is 95% for class 1 and 31% for class 2 tumours [42]. The molecular classes also show strong correlation with other risk factors such as patient age, cell type and chromosomal aberrations

[42]. The GEP based classification has been shown to be superior for predicting metastases compared to monosomy 3 and other clinical and histopathological prognostic factors [43,44]. Molecular classification of these tumours based on transcriptomic analysis from fine needle aspiration biopsy in patients is now established [45]. Currently, a 15-gene commercial gene-expression assay from Castle Biosciences is available which separates UM into two classes. This assay was initially validated by Onken et al, where metastatic rates of 1% and 25.9% were found for class 1 and class 2 tumours [43]. Recently, expression of a tumour antigen called *PRAME* has been reported to be an independent marker of metastasis and when combined with the GEP panel of 12-genes, it was possible to predict a 5-year metastatic rate of 0% in class 1/PRAME negative tumours, 38% in class 1/PRAME positive tumours and 71% in class 2 tumours[46].

1.1.4 Management of primary and metastatic UM

Once a clinician has obtained sufficient evidence to diagnose UM, management will depend on several subjective and objective features of the tumour. The management can be broadly classified as global-preserving therapy that includes radiation, surgery and laser therapy, and enucleation. The COMS study (Collaborative Ocular Melanoma Study) [20] in the United States is a multicentre investigation which evaluates the therapeutic interventions for patients with choroidal melanomas. It performed a multicentre randomized clinical trial comparing the effectiveness of brachytherapy to enucleation and found no difference in the mortality between the two groups for up to a 15 year follow up period. Based on the results of the COMS trial, majority of primary UM lesion in the US are treated with plaque brachytherapy. The United Kingdom (UK) UM guidelines [20] are convened under the UM Melanoma Study Group. The guidelines are available on the Melanoma Focus website and use an evidence based systematic approach, the Scottish Intercollegiate Guidelines Network (SIGN) checklist, as a guide. Similar procedures to those used in the U.S. are implemented in the management of primary tumours in the UK.

Radiation therapy includes brachytherapy, proton-based external beam radiation and charged particle radiation. Brachytherapy involves delivering high amounts of concentrated radiation dose (using radioisotopes ^{125}I or ^{106}Ru) to the tumour by securing a radiation plaque to the episclera. Although brachytherapy helps achieve good local control, it is associated with various complications. The most common are radiation induced retinopathy (45%-67%)

and cataracts (44%) followed by neovascular glaucoma in a third of the patients and macular edema in a quarter of the patients. These complications can lead to moderate vision loss and poor visual acuity [47,48]. Proton beam therapy has shown positive results in a recent retrospective cohort study of patients with T3-T4 choroidal melanomas. One study found a local control rate of 96.4% and overall eye retention rate of 95% in a cohort of 982 patients treated with proton beam therapy [49]. Charged-particle radiotherapy can be used for medium to large tumours where brachytherapy may not be an ideal option. Mishra et al reported long-term outcomes of a randomized controlled trial of helium-ion therapy versus brachytherapy for ciliochoroidal melanomas. They demonstrated improved local control of 100% versus 84% at 5 years and 98% versus 79% at 12 years in favour of charged-particle therapy, in addition to improved eye preservation and disease-free survival. However this study was based on a single center and extending the finding to the general population will require further studies [50].

Laser therapy such as photodynamic laser photocoagulation therapy (PDT) and transpupillary thermal therapy (TTT) aims to destroy tumour vascular supplies and reduce local recurrence. PDT involves injecting photosensitive dye followed by exposure of the area of interest to a laser (689 nm wavelength) which causes liberation of free radicals. TTT is used as an alternative to laser photocoagulation for small melanomas and uses an infrared diode laser to burn the tumour. However, TTT has been found to be ineffective in treatment of smaller or intermediate choroidal melanomas [51,52]. In addition, TTT does not shown significant improvement in visual outcome when compared to plaque radiotherapy [53].

However, undesirable, enucleation remains the most common surgical procedure performed in UM. It involves removing the entire eye while leaving the eye muscle and remaining orbital content intact. Enucleation is reserved for patients with the worst visual prognosis, large tumours, extraocular growth, circumferential tumour invasion, retinal detachment, severe haemorrhage or vitreous haemorrhage. Many studies have shown that there is no significant difference in mortality between eye-preserving and surgical treatment approaches. The COMS study reveal no significant difference in long term survival between patients who underwent treatment with brachytherapy and those who had an enucleation. Thus eye-preserving options have gained favour over enucleation [54,55].

Currently there is no standardized therapy that improves survival in metastasizing melanoma. Systematic treatment options for metastatic disease include intravenous chemotherapy, immunotherapy and other loco-regional techniques for providing local control for hepatic metastasis. These include immune-embolization, chemoembolization, isolated

liver perfusion and hepatic intra-arterial chemotherapy. Local resection involves excision of the tumour with scleral lamellar dissection. Disadvantages with this technique are the serious and immediate post-operative complications such as retinal detachment and vitreous haemorrhage. Adjuvant therapy has the potential to prevent metastases. Dacarbazine is an intravenous alkylating agent that has been evaluated by a randomized controlled study. Although initially shown to have significant impact in cutaneous melanoma, it did not demonstrate any survival advantage in UM [56]. Other adjuvant therapies include Fotemustine, which has received a good response although there is no statistical significance in survival benefit [57]. Other novel therapies include efforts to exploit the high levels of tyrosine-kinase receptors such as c-Met and c-Kit in UM [57], Crizotinib has been demonstrated to inhibit the establishment of metastasis and is currently being studied [57]. Histone deacetyase (HDAC) inhibitors such as valproic acid are currently being investigated, the rational being that HDACs promote cell differentiation and induce the transition from the aggressive class 2 to class 1 UM cells [58]. The future of treatments and management in UM is constantly evolving and novel therapeutic strategies are emerging with the new discoveries.

1.2 Epidemiology and risk factors in uveal melanoma

Uveal melanoma (UM) is the most common type of eye cancer and the second most common type of non-cutaneous melanoma [4]. The site of origin of UM is the uveal tract of the eye comprising the iris, choroid and the ciliary body. UM predominantly affect adults and is rare in children [59–62].

1.2.1 Incidence

Although UM is a common form of eye cancer, it is still a rare cancer with an overall incidence estimates of less than 10 cases reported for every million individuals across different countries. The incidence varies by gender and ethnicity. Stang et al [63] used data from internationally accredited population based cancer registries across the world to obtain estimates of the crude and age adjusted incidence trend covering 15 years between the period 1983-97 by country **(Table 1.1)**. In Europe, the incidence trend clearly reflects a decreasing gradient from north to south, with < 3 cases per million in Spain and Italy and > 6 per million in Nordic countries. This is consistent with specific incident rates reported previously by country [60,64–69] even though earlier studies had shown contradictory evidence for the north-

south variation [59,70]. The most recent epidemiology study from England reflects stable incident rates over the past decade (1999-2010) with approximately 1 case per 100,000 and is consistent with data from Stang et al [63]. Singh et al performed the largest study determining the UM incidence trends in the United States and reported a mean age-adjusted annual stable rate of 5.1 cases per million, that was consistent over a study period of 36 years from 1973 to 2008 [70]. McLaughlin et al [4] examined the difference in the incidence of ocular tumours in the United States by latitude and found the overall rates were lower in the south compared to the northern states, however the ocular tumours they compared also included conjunctival melanomas and tumours from other parts of the eye. A recent study by Keenan et al [71] found significant geographical variation in the incidence within England and the population-based incidence rates in their study, adjusted for age, may be attributed to difference in the ethnic groups in different areas. These geographical differences could also reflect some of the important risk factors of UM that include ancestry, light skin and iris pigmentation [63,72–74]. The incidence of UM among the white population is highest compared to all other non-white populations. Data from Stang et al's study [63] showed the highest incidence rates of UM in Northern Europe and Australia and the lowest rates in populations from Costa Rica, Osaka and Singapore. In addition, the crude incidence rate of UM in the United States was 20 times higher in the white population than in the Afro-Caribbean population. Keenan et al [71] showed a similar trend in England with a lower incidence rate in areas with a higher proportion of black and Indian populations including those born outside the UK. Margo et al [73] looked specifically at race related incidence of UM in the United States between the period 1981 to 1993 and found a low relative risk of UM for blacks compared to the non-Hispanic and Hispanic white population. The non-Hispanic men were found to have 72 times higher risk of developing UM compared to black men and non-Hispanic white woman experienced a 22-fold risk of developing UM compared to black women. The white Hispanic group was less likely to develop UM compared to their white non-Hispanic counterparts. Hu et al [74] reported age-adjusted incidence rates of UM in various ethnic groups from various cities in the United States including white Hispanics, non-Hispanics and blacks and found significant differences in comparisons between each group except for black versus Asian population. This study did not include any American Indians. Although the incidence of UM is rare in Asian population, a few studies have reported a low incidence (< 3 cases per million) in Chinese patients and they tend to be younger compared to the white population [75,76]. Another study [77] examining the ethnic variation among Jewish immigrants to Israel compared to Israeli born Jews and their parents found higher rates in individual born to European and American parents and

lower rates among individual with African or Asian ancestry. However, examining the incidence trends over time, the Israel-born Jews showed increased incidence rates compared to the European and American born Jews. Although the rates were not based on large numbers, the study concluded that within the Jewish population, the incidence rates were lowest in Jews from Asia and Africa and in non-Jews and highest in Jews born in Israel, Europe and the Americas.

Table 1.1: Age-adjusted incidence estimates of UM from 1983-1997

Countries	Men*	Women*
Europe		
Denmark	0.75	0.63
UK Scotland	0.69	0.53
Slovakia	0.63	0.51
Estonia	0.63	0.57
Czech	0.56	0.46
Eindhoven	0.56	0.31
France	0.55	0.44
England	0.47	0.42
Switzerland	0.42	0.4
Italy	0.29	0.26
Spain	0.26	0.17
North and Central America		
Canada w/o Quebec	0.59	0.48
SEER White	0.51	0.42
SEER Black	0.04	0.02
Costa Rica	0.04	0.09
Asia and Oceania		
Australia	0.62	0.52
Singapore	0.03	0.01
Osaka	0.02	0.02

Table adapted from Stang et al, (2005). * indicates adjusted to cases per 100,000

1.2.2 Ethnic and geographical risk factors

Other important risk factors relating to population ethnicity are eye colour and fair skin. A population based case control study in Australia on 290 cases and 916 controls found higher risk of UM in people with grey, hazel and blue eyes compared to those with brown eye colour [78]. A meta-analysis of 132 published reports on risk factors for UM provides evidence for increased risk with fair skin colour, light iris colour and ability to tan [79]. While many earlier studies have provided consistent evidence for eye colour [77,80–82], there have been some inconsistencies in skin and hair colour as important risk factors for UM [77,79,80,83]. Furthermore, the geographical variations reported in a few studies [63,69,71] could be attributed to ethnicity or environmental factors such as ultraviolet (UV) radiation or other occupational risk. Thus, there has been interest in understanding the etiological synergism between the light eye colour that correlates with population ancestry and the role of UV radiation to support the hypothesis of gene-environment interaction in the causation of UM. The increased risk of UM in people with lighter blue or grey iris colour compared to darker brown eyes suggests an increased risk due to less melanin in the choroidal melanocytes. This could lead to less protection of the DNA in the nuclei of these cells against solar radiation [80]. Wakamatsu et al [84] studied the relationship between the quality and type of melanin, brown eumelanin and yellow or red pheomelanin in relation to the iris colour in cultured uveal melanocytes from 61 patients. They found that the quantity and type of melanin, correlate with iris colour. The quantity of total melanin and eumelanin/pheomelanin ratio in dark eyes were higher compared to light eyes. In addition, the pheomelanin content was constant while eumelanin content increased with iris colour. They concluded that melanin in dark-coloured irides may be covered by eumelanin while the melanin in light-coloured irides behaves more as pheomelanin, a pro-oxidant, and this coupled with lower melanin could explain the high prevalence of UM in people with light-eyes. In addition, De Leeuw et al's [85] study demonstrated the protective effect of melanin in cultured skin melanocytes, where they showed high melanin content associates with resistance to UV-B cytotoxicity and found the best survival in the cell population with a high content of total melanin and low pheomelanin content.

Anatomically, much of the UV radiation impinging on the human eye is absorbed by the cornea and lens and less than 1 percent of the radiation below a wavelength of 340nm reaches the retina [86]. Thus, while exposure to the retina is low, the choroidal region can still be exposed to solar UV radiation. A study by Li et al [87] found that the distribution of UM

arising in the uveal tract was not uniform. The rate of occurrence was concentrated in the macular area and decreased progressively with increasing distance from the macula to the ciliary body consistent with the dose distribution of light on retinal sphere. This supports the possible role of solar exposure in the induction of these tumours.

The geographical variation in UM incidence has not been consistent. In the European data [69] incidences across latitudes, used as a proxy for sun exposure, where higher ground exposure of UV radiation occurs in lower latitudes, showed a decreasing north-south gradient. This could be due to the protective effect of ocular and skin pigmentation prevalent in the southern populations. On the contrary, Moan et al [88] showed no north-south gradient in UM incidences, discounting the role of UV radiation. Adding to this, the incidence estimates in UM have remained stable over decades [60,66,71,89] and unlike cutaneous melanoma, the role of ultraviolet radiations in the pathogenesis of these tumours is not clear. Further, the estimated risk of UM associated with personal exposure to UV-B radiation has been inconsistent both within and between case-control studies [80,83,90–92]. Giles et al [93] performed a case control study in Australia with 290 patients and included additional meta information on number of outdoor hours at different times in life, exposure in early life and sun sensitivity. They found a strong positive association between sun exposure and increased risk of uveal melanoma including choroid and ciliary body tumours. When taken together with previous evidence of an association with artificial source of UV radiation [80,91] , this supported the role for UV radiation as responsible agent of UM risk. These findings on the synergistic effect of light eye iris colour and UV radiation on the risk of UM requires further investigation.

1.2.3 Age and sex-related risk factors

The incidence rates show gender specific differences with higher rates in males compared to females [64,70,71]. The age adjusted incidence of UM is 5.8 per million males and 4.4 per million females in the United States [89]. The incidence estimates from an Australian study gives similar estimates in men and women in the younger age group but significantly higher incident rates in men compared to women above 65 years [74].

Age is a strong determinant factor for risk of UM with increasing incidence after the 7th decade of life [4,74]. Other potential risk factors include sex-related factors [94–96], smoking [97], viral infection [98,99] and chemicals [100], however no causative agent has been recognized so far. The role of social class is also controversial with some showing a positive association with

UM [97,101] whereas Lockington et al reported that socioeconomic status is not a significant factor for choroidal melanoma following a retrospective study of 536 from Scotland [101].

1.2.4 Congenital nevi and familial risk factors

Congenital ocular and oculodermal melanocytosis (ODM) also called nevus of Ota and nevus or lesions found in the uvea, including choroid and iris nevi, are predisposing factors for UM. Singh et al estimated the lifetime risk of developing UM from ODM is 1 in 400 individuals [3]. Melanocytic choroidal lesions (choroidal nevus) ranging between 0.75 to 6.0 mm are found at a prevalence of 4.6% -7.9% in white individuals and show a slow rate of malignant transformation (1 in 8845) to choroidal melanoma [3]. Larger choroidal nevi of 10mm or more in diameter to transform into melanoma in 18% of cases at 10 years [3]. In addition, other lesions related risk factors include atypical and common cutaneous nevi and cutaneous freckles [102].

A hereditary basis for UM predisposition has been reported by a number of studies [103–108], although familial incidence is rare. Singh et al [109] investigated 4500 UM patients and 17 kindreds where the first degree relative of the proband had also been affected with UM and found strong statistical evidence for non-coincidental occurrence of familial UM with the expected number of affected first degree relative being 0.81 (RR=20.99, 95% CI, 12.2-33.6). Recently, germline mutations in the BRCA1 associated protein 1 (*BAP1)* gene was found to be the cause of a novel autosomal dominant syndrome, characterized by high penetrance of melanocytic neoplasms and with an increased risk of UM and cutaneous melanoma [110]. Another independent study [111] also identified germline mutations in *BAP1* in two families with a high incidence of mesothelioma where some carriers also developed UM. These findings were obtained following our identification by the Bowcock laboratory of *BAP1* as a metastasis suppressor of UM (see below), highlighting the role of *BAP1* as a predisposition gene in UM.

1.3 Developmental lineage of uveal melanocytes

UMs develop from melanocytes of the uveal tract. Melanocytes are melanin pigment-producing cells that are found in many different parts of the body (skin, eyes, inner ear, central nervous system, heart and muscle) and which are derived from different

developmental pathways. During development, the pluripotent neural crest cells develop from the embryonic ectoderm cell layer. These then gives rise to non-pigmented melanoblasts which differentiate into melanocytes. While migrating during embryogenesis the neural crest cells can bypass the natural tissue barriers and basement membrane of the eye and migrate to the optic cup. The choroid develops from the mesenchyme that surrounds the inner external surface of the optic cup, whereas the iris and the ciliary body develop from the anterior part of the optic cup. Within the uvea, the non-pigmented melanoblast mature into melanocytes or give rise to the melanocytic stem cells. The location of these uveal melanocytic stems cells is unknown. In mammals *SOX10* expression specifies the differentiation of melanoblasts from the neural crest cells, which subsequently express *MIFT*, *DCT* and *KIT* and specialize to form differentiated melanocytes [112]. In humans subtle changes in the transcriptional activity of *MIFT* is regulated by *IRF1* which is partially responsible for conferring pale skin and blue eye [113]. In UM, gene expression profile (GEP) based classification (discussed below) separates the tumours into two classes that accurately predict metastatic death, namely class1 tumours with a low propensity to metastasize and class 2 tumours which have a high risk of metastasizing. The expression signature of the aggressive class 2 tumours include down regulation of neural crest and melanocytic-specific genes and upregulation of epithelial genes [114]. Loss of function mutations in the tumor suppressor gene *BAP1* are strongly associated with class 2 metastasizing tumours (discussed below)[115]. Interestingly, depletion of BAP1 in UM cell lines leads to downregulation of canonical differentiation genes involved in melanocyte differentiation including *MITF*, *TRPM1*, *TYR* and *DCT* [116].

1.4 Genetics of uveal melanoma

Despite advances in clinical management, UM continues to be an aggressive cancer with systemic metastasis seen in approximately half of the patients. For the last 2 decades, our group and others have focussed on trying to understand the genetics of UM and the molecular basis of tumour formation and systemic metastases. New technologies are helping to reveal genome-wide molecular alterations at high resolution.

1.4.1 Genome wide association studies in uveal melanoma

As discussed above, UM is associated with risk factors such as skin colour, eye colour and ethnicity, and can segregate in families as part of a highly penetrant cancer susceptibility syndrome. Thus, the role of genetic risk factors in the etiology of UM is strongly implicated. However, due to the rarity of this cancer and the absence of comprehensive genetic data from large cohorts of patients, genetic susceptibility to UM has been a challenge. Ferguson et al provided the first evidence for genetic factors associated with UM susceptibility. They selected 28 SNPs, previously identified as risk variants for cutaneous melanoma (CM) through a genome wide association study (GWAS) and tested them for association with UM [117]. They identified 5 variants that were significant after correction for multiple testing. The three most significant SNPs were rs12913832, rs1129038 and rs916977 mapping to chromosome 15q12 locus in the region of *HERC2/OCA2* gene. Variants in this gene are associated with different eye colours in the human population [117]. Although the study had limited power, the most significant association at the *HERC2/OCA2* locus had an odds ratio (OR) of risk effect of less than 0.6, with rs12913832 having the highest OR of 0.562 (*P-value* = $1.13e^{-05}$). Mobuchon et al performed an independent GWAS and found another candidate locus on chromosome 5p15.33 (rs421284: OR = 1.7, CI 1.43–2.05, $P = 6 \times 10^{-3}$) that they replicated in an independent cohort [118]. This locus harbours two candidate genes: *TERT* and *CLPTM1L* and many SNPs that were in LD with the original risk SNP [118]. The presence of the risk allele (base C) positively correlates with higher expression of *CLPTM1L* than the non-risk allele. Further work is required to determine the biological significance of these findings.

1.4.2 Copy number profiles in uveal melanoma

Genomic instability is an important characteristic of cancers and plays an important role in tumour development and progression [119]. Genomic instability leads to copy number alterations (CNA) which can involve gains and losses of small regions or whole chromosomes (aneuploidy). CNAs are observed in all major human cancer types [120]. Other chromosomal abnormality includes structural aberrations such as translocations. For example the Philadelphia chromosome involves fusion of the ABL proto-oncogene 1, non-receptor tyrosine kinase (*ABL1)* gene on chromosome 9 and the RhoGEF and GTPase activating

protein (*BCR)* gene on chromosome 22 in chronic myeloid leukaemia [121]. Structural aberrations and CNAs can affect gene expression either by altering gene dosage or disrupting regulatory regions as observed in the healthy population [122]. However, disease-specific CNAs are of greater interest since they can provide insights into disease pathogenesis. In cancer, CNAs can lead to activation of oncogenes or inactivation of tumour suppressor genes that can play roles in pathways related to cell growth, proliferation and metastasis. Common examples of oncogenic activation of genes due to chromosomal amplification include the v-myc avian myelocytomatosis viral oncogene homolog (*MYC)* observed in a lung cancer cell line [123] and erb-b2 receptor tyrosine kinase 2 (*ERBB2)* in breast cancers [124,125]. Inactivation of tumour suppressor genes due to homozygous deletions have been observed for RB transcriptional corepressor 1 (*RB1)* gene in retinoblastoma and osteosarcoma [126] and phosphatase and tensin homolog (*PTEN)* in brain, breast and prostate cancers [127]. Thus, in addition to providing new insights into tumour biology, identification of copy number alterations can enable discovery of oncogenes and tumour suppressor genes driving tumorigenesis. While most of the chromosomal alterations are likely to be non-specific by-products of overall genomic instability, molecular dissection of characteristic recurrent patterns can help in the identification of pathogenetically relevant alterations. UMs are notably uniform in their clinical and histopathologic features [128] and metastasize in up to 50% of patients. In addition, they exhibit characteristic recurrent copy number alterations but less genomic instability compared to most solid tumours such as cutaneous and breast cancers [129–131]. UMs thus present itself as an attractive solid tumour for studying aneuploidy.

Copy number profiles of uveal melanomas have been characterized by several groups and recurrent aneuploidy of a few specific chromosomes remains a key feature in these tumours. The first report on molecular abnormalities in UMs by Sisley et al were gross alterations of chromosome 1, 3, 6 and 8. This was observed in a karyotype from short-term cultures from six posterior UMs [132]. Since this landmark discovery several studies employing much higher resolution techniques have interrogated this tumour type to identify large-scale losses and gains events. Most studies on copy number characterization in UMs have identified discrete pattern of chromosomal alterations involving chromosome 1, 3, 6 and 8. These studies found loss of the short arm of chromosome 1(1p-), loss of both short and long arms of chromosome 3 (Monosomy 3), loss of the long arm of chromosome 6 (6q-), gain of the short arm of chromosome 6 (6p+), gain of short and long arms of chromosome 8 (8p+, 8q+) and loss of short arm of chromosome 8 (8p-). Less frequent abnormalities were reported on chromosomes 16q (loss) and 21 [40,133,134]. The significance of these recurrent alterations

was further evaluated by correlating them with clinical and histopathologic parameters. Monosomy 3 and large tumour diameter were found to be most significant predictor of survival in patients [40]. Co-existence of 6p+ and 6p- in same tumours likely represents isochromosome 6 i(6p) as it does for chromosome 8 i(8q) which are both commonly observed in UMs [135–137].

UMs exhibit chromosomal patterns characterized by chromosomal abnormalities that correlate with mutations and gene expression signatures illustrating distinct biological mechanisms that underlie the pathogenesis of these tumours. The most striking abnormality associated with clinical and histopathological prognostic factors and with increased metastatic risk involves complete and partial loss of chromosome 3. This is seen in nearly half of all UMs [36]. However, it is only monosomy 3 involving the whole chromosome, that is prognostically relevant, conferring a high risk of metastasis [138,139] A small number of UMs show intra-tumoral heterogeneity of monosomy 3, but this does not show any effect on survival. Mensink et al [140] evaluated genetic heterogeneity in UM with FISH and identified subclones with different percentage of monosomy 3 in ten out of sixteen tumours. Another study showed a heterogeneous distribution of monosomy 3 based on FISH analysis and concluded that fine needle biopsy may be subject sampling error [141]. When comparing tumours with a low percent of monosomy 3 with those with a high percent of monosomy 3 (present in most nuclei), no significant difference was found in metastasis-free survival [138]. Since the prognostic value of chromosome 3 has been confirmed by been several group [33,35,40,142], it is assumed that monosomy 3 is a primary event in UM that often occurs with other chromosomal alterations involving chromosome 1, 6 and 8 [33,132,135]. Co-occurrence of chromosome 3 and 8 have been well recorded in previous studies, other alterations that occur less frequently with monosomy 3 include gain of 6p, loss of 1p or 6q [102,135,143,144].

Gain of 8q is associated with poor prognosis mainly in the context of 8p loss, suggesting formation of isochromosome 8q. This occurs in nearly a quarter of all UMs and almost exclusively in tumours with monosomy 3 [135,145]. Considering the GEP status of these tumours, isochromosome 8q is associated with the highly metastatic class 2 signature. Deletion of 8p12-22 locus is associated with more rapid onset of metastasis and shown to be accompanied by silencing of the corresponding retained chromosomal arm [145]. Chromosome 8q loss is thought to be a secondary event in UM [146,147]. Van den Bosch et al [148] showed that in addition to the presence of monosomy 3 and 8q gain, high percent of these other aberrations strongly correlated with poor outcomes compared to cells with a lower percentage of aberrations. Increased copies of 8q gain can coincide with an absence or low percent of

monosomy 3, relate to larger tumour size and leads to worsening patient survival [148]. They suggested that tumours accumulate cytogenetic alterations in an increasing number of cells as they grow bigger and that this leads to additional copies of 8q and worsening survival.

Aberrations on chromosome 6 involve loss of 6q and gain of 6p and occur in up to a third of UMs [149]. Gain of 6p are frequently found in non-metastasizing disomy 3 tumours, suggesting a bifurcation in tumour progression [38–40]. The effect of gain of 6p on patient outcome is not conclusive as it has been shown to be related to longer survival [35] but is not associated with disease-free survival [37]. Harbour et al [38] applied statistical pattern recognition algorithm to study aneuploidy patterns in 388 primary UMs, 45 from their cohort and 336 from previously published cases in addition, they performed an integrative analysis of gene expression and aneuploidy. They classified tumours based on chromosome 3 and 6p alterations (normal copy; 3nl/6pnl, monosomy 3 and 6p gain; 3nl/6p+ and monosomy 3; 3− tumours). These changes correlated with their GEPs. Key findings from this study included the following: (a) the association between aneuploidy and poor clinical outcome is determined by early chromosomal events rather than overall aneuploidy, (b) monosomy 3 and 6p+ represent almost mutually exclusive alterations and there is a third group of tumours with a normal copy of chromosome 3 and 6p (3nl/6pnl). Gene expression profiling suggested this group was more genetically similar to 6p+ tumours potentially representing an earlier stage in 6p+ tumour development but these tumours exhibited better prognosis than 6p+ tumours, (c) the gene expression pattern associated with aneuploidy revealed enrichment of genes involved in genomic integrity and chromosomal regulation, (d) 8q+ did not achieve statistical significance as a predictor of clinical outcome since it was found in all groups, suggesting strong selective pressure for gain of genetic material on 8q in all the UM tumours, (e) monosomy 3 and 6p+ were thought to be early events and i(8q) and 8p- late events in monosomy 3 tumours. Thus, they confirmed that tumours with monosomy 3 and gain of 6p represent mutually exclusive alterations and support the idea of phenotypic bifurcation where tumours with either one of those changes progress independently.

Rearrangements on the short arm of chromosome 1 giving rise to loss of 1p are found in UM. However, these 1p aberrations are less common compared to other alterations such as monosomy 3, 6p gain and 8q gain, with a frequency of approximately 30% [39,135,142,147,150]. Loss of 1p provides prognostic information independent of monosomy 3 status, with its presence presaging decreased disease-free survival [37]. Other less frequent chromosomal aberrations that have been identified, include LOH of 9p, 13q and 17p [151], loss of

chromosome 11q23 [142], gain of chromosome 18q22 and 18q11.2 [152,153], gain of 21q [153] and loss of chromosome 16q [40,154].

Refined mapping of candidate genes on chromosomes with cytogenetic changes: These non-random chromosomal alterations have been further explored to identify candidate oncogenes and tumour suppressor involved in UM pathogenesis since they can narrow down the altered regions on the chromosomal arm **(Table 1.2)**. Hausler et al [155] performed microsatellite genotyping of 70 UMs and identified a smallest commonly deleted region on chromosome 1p spanning 55Mb at 1p31. This had been reported previously, by a study that screened for copy number alteration on chromosomes 1, 3, 6 and 8 across metastasizing and non-metastasizing UMs [32]. The genes mapping to this 1p31 locus (*PDE4B*, *IL12RB2* and *ITGB3BP*) showed reduced or absent expression in tumours with monosomy 3, however no gene mutation has been found to be in this locus to be associated with UM pathogenesis [37,155]. However, their role in UM has not been well established in subsequent literature. Partial or complete loss of 1p occurs in a quarter of UMs and more often in tumours with monosomy 3 [155]. It was found to be a better independent predictor of disease free survival other than chromosome 3 [37]. One study tried to provide evidence for oncogenic role for a spliced variant of *TP53* homologue *p73* gene on 1p however its significance is yet to be established [143].

In the case of chromosome 6 alterations it is unclear if 6q loss is pathogenetically more significant than 6p gain, although loss of 6q and/or gain of 6p occur in close to one-third of tumours often as isochromosome 6p [137] and associated with better prognosis compared to tumours with monosomy 3. This 'protective effect' against metastasis could be because this change is found in tumours with absence of monosomy 3 [149]. The common region on 6p and 6q identified with CGH and FISH includes gain of 6pter-6p21 and loss of 6q16.1-6q22 [156], however no pathological mutation or differential expression of genes in these regions have been identified.

The smallest region identified on chromosome 8q is at 8q23-24→ qter [135,157]. This contains several potential oncogenes such as *MYC*, *DDEF1* and *NBS1*. The 8q abnormality was further explored and the c-*MYC* oncogene was shown to occur in at least 30% of primary UMs and only tumours with monosomy 3 showed amplification of c-*MYC* [158]**,** however it is still unclear if c-*MYC* is the main driver on 8q. The pathogenic significance of any of these observation on 8q have not been established in UM. Harbour et al found that loss of 8p has higher prognostic value than gain of 8q and identified a region of minimal deletion at 8p12-22. This deleted region contained 11 well defined genes, six of which were downregulated in

GEP based class 2 tumors (*CNOT7*, *ASAH1*, *ATP6V1B2*, *LZTS1*, *DPYSL2* and *FZD3*). Of these six genes, *LZTS1* was most strongly linked to metastasis and examined as a candidate tumour suppressor gene however, further work is required to determine its significance [145].

Loss of 9p is found in nearly a quarter of UMs [149] and cytogenetic analyses have identified a small region of LOH at 9p21 in nearly a third of tumours [159,160]. This locus includes the *CDKN2A* tumour suppressor gene and methylation of the *CDKN2A* promoter occurs in 24-31% of cases. It has been suggested that inactivation of *CDKN2A* plays a role in UM progression [160,161].

Monosomy 3 occurs in nearly half of all UMs and is strongly associated with clinical, histopathologic prognostic factors. Metastasis is the most prognostically significant chromosomal marker in UMs with monosomy 3 [34,162]. Thus, chromosome 3 attracted considerable attention with the expectation that a tumor suppressor gene may be involved in UM progression [163]. Many groups had attempted to narrow this down to a more focally deleted region. These included 3p11-14 [164,165], 3p25-26 [165,166], 3p25.1-3p25.2 [167] and 3q [165,166]. And there were many unsuccessful attempts at identifying candidate genes on chromosome 3 that played a role in UM metastasis [168–170]. Harbour & Bowcock et al used next generation sequencing to sequence whole exomes from two tumours with monosomy 3 that had undergone metastasis and identified deleterious loss of function mutations in BRCA1-associated protein-1 (*BAP1*) which was located on chromosome 3p21.1 in both tumours [115]. They then screened additional tumours with Sanger sequencing and identified inactivating mutations in BAP1 in 47% UMs [115]. These mutations occurred in metastasizing UMs with monosomy 3 consistent with the Knudson's 'two-hit' model for cancer genes. Thus, information on copy number complemented with mutational profiling enabled the identification of a key tumor suppressor gene associated with metastasis in UM.

Harbour et al [42] also showed that UM tumors can be subdivided into prognostically significant groups based on a gene expression profile (GEP) of 15 discriminating genes where class 1 tumors (low-grade tumors) have longer metastasis-free survival compared to class 2 tumors with high propensity to metastasize [42]. Copy number alterations that correlated with these expression classes are gain of 6p in class 1 tumors and monosomy 3 and 8q gain in case of class 2 tumors [129,171]. A more recent study by Harbour et al identified high expression of a tumor antigen *PRAME* as an independent prognostic marker for metastasis where tumors (both class 1 and class 2) that expressed PRAME above a threshold value (PRAME+) were associated with an increased risk of metastasis in UM [46]. Both class 1 and class 2 tumors with PRAME+ status were associated with 6p gain, 6q loss, 8q gain and 16q loss whereas 1q gain

and 8p loss occurred exclusively in class 1 and class 2 respectively. Interestingly, PRAME expression was not associated with monosomy 3. These findings highlight greater significance of these CNAs when the tumor samples are stratified based on these genes expression classifiers and role of these CNAs mediated as cause or effect of PRAME upregulation is currently unknown.

Table 1.2: Summary of genomic finding from other studies in uveal melanoma

Chromosome arm	Alteration	Prognostic value	Focal refinement	Candidate genes	Known UM driver genes	References
1p	loss	Yes	1p31	PDE4B, IL12RB2 and ITGB3BP		37, 155
3p, 3q	loss	Yes	3p11-14, 3p25-26, 3p25.1-3p25.2, 3p21.1		BAP1	164-167, 115
6p	gain		6pter-6p21			156
6q	loss	Yes	6q16.1-6q22			156
8p	loss	Yes	8p12-22	CNOT7, ASAH1, ATP6V1B2, LZTS1, DPYSL2 and FZD3		145
8q	gain	Yes	8q23-24qter	MYC, DDEF1 and NBS1		135, 157
9p	loss		9p21	CDKN2A		160, 161
13q	loss					151
17p	loss					151
11q	loss		11q23			142
18q	gain		18q22, 18q11.2			152, 153
16q	loss					40, 154
21q	gain					153

1.4.3 Structural rearrangements in uveal melanoma

Genomic stability in a key attribute of cancer and can be in the form of gene mutations, copy number alterations and other structural aberrations involving rearrangement of the genome, such as translocations and inversions events. Structural chromosomal rearrangements events may result in convergence of coding or regulatory DNA sequences between genes located nearby or faraway from each other to form gene fusions. A few gene fusions are drivers of neoplasia and as such represent a novel disease mechanism leading tumourigenesis. Gene fusions can enable the sub-classification of tumours and be correlated with clinical phenotype. This makes them ideal for diagnostic purposes as seen in case of non-small cell lung cancers where *ALK* fusions are routinely screened for disease management [172]. At the transcriptomic level the balanced and unbalanced chromosomal rearrangements can result in the formation of fused RNAs.

Fusions can sometimes be seen as chromosomal rearrangements including translocations (transfer of segments from one chromosome to another), insertions and

inversions (rotation of a segment by 180 degrees within a chromosome) and unbalanced events resulting from deletion of an interstitial chromosomal segment (parts of the p and q arm of a chromosome which are neither close to the centromere nor telomeres). The first reported gene fusion event was a translocation observed on the Philadelphia chromosome involving fusion of the ABL proto-oncogene 1, non-receptor tyrosine kinase (*ABL1)* gene on chromosome 9 and the RhoGEF and GTPase activating protein (*BCR)* gene on chromosome 22 in chronic myeloid leukaemia (CML). Splicing of the hybrid transcript produced by the *BCR-ABL1* fusion gene, yields an mRNA in which the *BCR* sequence is fused with exon 2 of the *ABL1* gene. The presence of this fusion is associated with poor prognosis [121,173]. One of the most common gene fusions events can be seen in prostate cancers, and involves a recurrent fusion of the *TMPRSS2* gene to oncogenic ETS family members encoding transcription factors *ERG*, *ETV1* or *ETV4* [174,175].

Chromosomal abnormalities including copy number changes and structural alterations are key features of both solid and liquid (blood-borne) tumours. The number of studies characterizing cytogenetic abnormalities are increasing[176] and the advent of molecular techniques such as FISH along with high-throughput technologies such as whole genome sequencing and RNA sequencing have added a further level of sophistication to the analysis. Besides revealing novel chimeric fused genes associated with tumourigenesis these technologies can enable the identification of the specific chromosomal breakpoints associated with the fusion.

Despite solid tumours being more common and making a significantly greater contribution to morbidity and mortality than liquid cancers, less is known about the role of their chromosomal alterations and their importance in a clinical and biological context. For example, the reported number of chromosomal rearrangements contributing to gene fusions events are far greater in hematopoietic malignancies than in most solid tumours [177]. However this could due to publication bias as liquid tumours have been far more extensively characterized compared to solid tumours [178]. Nevertheless, this trend is changing and significant information is accumulating on structural chromosomal aberrations in solid tumours which exhibit widespread karyotypic changes. In the context of epithelial cancers, many gene fusion events have been characterized over the last few decades, including, *ALK*, *ROS1* and *RET* kinase fusions in lung cancer, *RAF* kinase fusions in brain tumours, melanoma, gastric and prostate cancers, R-spondin fusions in colorectal and prostate cancer, *CD44-SLC1A2* gene fusions in gastric cancers, MAST- and NOTCH-family gene fusions in breast cancers and *MITF* fusions in renal cancer [179].

Interestingly, to date only a single study characterizing chromosomal rearrangement events has been published in UM [180]; and none describe gene fusion events at a transcriptomic level, although several fusions targeting genes such as *PAK1* and *DGKB* have been reported recently in cutaneous and other non-cutaneous melanoma [181]. Bertrand et al recently performed deep whole genome sequencing of 24 primary UMs and 9 metastatic samples and found an average of 50 structural variants including a few recurrent inter-chromosomal rearrangements [180]. Their data suggested an absence of major drivers constituted by genetic rearrangement events. They reported translocations in 3 out of 33 samples involving chromosomes 6 and 8, disrupting the genes *UBE2W* and *MYO6* in two samples and an intergenic region in the third sample. Additionally, translocations between chromosome 13 and 17, and chromosomes 3 and 12 were reported. Although these recurrent translocations involved nearly same genomic regions, there was absence of a common gene affected by these events.

Yoshihara et al performed a comprehensive analysis of RNA sequencing data from 4366 primary tumour samples across 20 cancer types, including 80 UMs from The Cancer Genome Atlas (TCGA) dataset. They identified 7887 high confident fusion events and tabulated fusions per cancer type in the TCGA Fusion gene data portal (http://www.tumorfusions.org) [182]. They used a 4-tier system to classify the transcripts based on minimum discordant pairs, matched junction spanning reads and uniqueness of the gene partner within a sample. Tier 1: fusions harboured a minimum of, three discordant pairs, two junction spanning reads and gene partner uniqueness within a sample; Tier 2: fusions with a minimum of two discordant read pairs and one junction spanning reads and breakpoints within 100Kb from predicted junction point; Tier 3: fusions which showed high consistency of predicted junction, gene partner uniqueness within sample, minimum two discordant read pairs and one junction spanning reads; and Tier 4: all fusions not present in tiers 1 to 3. The validation rate (based on their analysis of the glioblastoma dataset) of tier 1 and tier 2 events was greater than the others (tier 3 & tier 4). Although, they did not include UMs in their downstream analysis of fusion transcripts, when querying (for the current study) the tier 1 results of the TCGA UM samples from the fusion gene data portal, revealed 22 reported fusions events across 11 samples. Recurrent fusion pairs were absent and seven in-frame events were reported. These were: *TFG-GPR128*, *OC1AD1-KIT*, *RFC1-C4orf34*, *FILIP1-KCNQ5*, *STX7-LMBRD1*, *9-SEP-BCAS3* and *KLK2-FGFR2*. The orthogonal validation of these reported fusion events with Sanger sequencing was not assessed probably due to the

unavailability of samples for experimental follow up. Thus, the relevance of these events requires further investigation.

Another groups of gene fusions or chimeric RNA are formed by intergenic splicing events, referred to as "non-canonical chimeras" [179]. These non-canonical chimeric RNAs can be formed by trans-splicing (i.e. splicing together exons from multiple separately encoded transcripts) or read-through transcripts which are formed when the transcription machinery goes beyond the normal termination and reads through intergenic regions of neighboring genes. This results in a final chimeric RNA composed of sequences from two different genes. The presence of these fusions can increase noise in the data. Thus, identifying gene fusion detection tools which filter these read-through transcripts are preferable. Since most studies have focused on canonical gene fusions in cancers the relevance of non-canonical chimeras, which were only discovered recently, is largely unknown. As described above, although, previously analysis of RNA-sequencing data to characterize fusion transcripts in UM has been attempted [183], a comprehensive and focused study in UM has not yet been done till date.

1.4.4 Gene expression based classification

Although cytogenetic alterations are valuable for use in clinical prognostication, they can be susceptible to sampling error due to the significant intra-tumoral heterogeneity present in UM [141,184]. Previously several groups explored the use of gene expression to gain prognostic insights in UM. Zuidervaart et al [185] compared the gene expression of tumour specimens and cell lines with nylon filter microarrays containing 1176 gene spots related to cancer development. They found 4 genes whose expression levels divided tumours into two distinct subgroups [185]. These genes included *ET2*, *VBP1*, *CUL2* and *LAMR1*. They also found a positive correlation for three of the genes (*ET2*, *LAMR1* and *VBP1*) with PAS-positive loops, which are linked to poor outcomes [186]. However, they could not find any association with survival due to the unavailability of data on all of the tumours.

Tschentscher et al [187] analysed the expression patterns of primary tumours with oligonucleotide arrays tiling 12500 probes sets. They identified genes that were differentially expressed between UMs with and without disomy 3. They found seven genes that were absent in monosomy 3 tumours, and they proposed two genes *CHL1* and *fls485* on 3p25 as candidate tumour suppressor genes. However, they could not identify any mutation or epigenetic alteration in either of these two genes. They also described several other genes that

were highly significantly differentially expressed between the tumours, including *SPP1*, *TIMP-3* and *HTR2B*. High expression of *HTR2B* was seen in the tumours with monosomy 3 and was proposed to be a marker of metastasizing tumours. In addition, expression data was used for class discovery by unsupervised clustering. This revealed two distinct tumour classes that correlated with cytogenetic alteration on chromosomes 3, 6 and 8. Monosomy 3 was confined to only one of the two classes while 6p and 8q gain were seen in both classes. They could not establish the prognostic significance of these two classes tumours since the samples in their study were from patients with recent diagnosis and data on metastasis were not available.

Onken et al [42] went on to show that GEP could classify tumours into two prognostically significant groups with unsupervised clustering techniques, namely class 1 and class 2 tumours. Class 1 or low grade tumours had a low risk and class 2 or high grade tumours had a higher risk of metastasis. They then used a supervised technique to identify the genes that separate the two classes and found 62 discriminating genes, including a significant cluster on chromosome 3 and 8q that correlated with metastatic death. However, further enrichment of these genes using a signal to noise threshold resulted in exclusion of most the genes on chromosomes 3 and 8q. The smallest set of genes that could accurately predict the class labels was three, (*PHLDA1*, *FZD6* and *ENPP2)*. Most notably, the prognostic accuracy of this GEP based classification outperformed other clinicopathologic prognostic factors [44], and was confirmed by several groups [188,189]. The GEP from high-density hybridization based microarray platform for predicting metastasis was then reduced to twelve discriminating and three control genes that could be assayed with a microfluidics platform. This is now used in the clinic following aspiration of small quantities of sample with a fine-needle biopsy [190]. The prognostic accuracy of this assay has been validated in a prospective multicentre study spanning 12 independent centers and found to be superior over monosomy 3 status for clinical prognostic testing [43]. One possible reasons suggested for the superiority of GEP over cytogenetic method for prognostication is that cytogenetic markers are distributed heterogeneously and subject to sampling error, while GEP represents a less variable functional 'snapshot' of the tumour microenvironment [149]. The most recent work by Harbour et al, has revealed four prognostically significant classes based on GEP, namely 1A, 1B, 2A and 2B [129]. The expression of two discriminating genes (*CDH1* and *RAB31*) from the GEP panel was proposed to identify the subset of class 1 tumours with increased metastatic risk. Tumours with low expression of *CDH1*/*RAB31* and lower predicted metastatic risk were called class 1A, whereas those with higher expression of *CDH1*/*RAB31* genes and higher

predicted metastatic risk were called class 1B [191]. Class 1A has the longest metastasis free survival and class 2B has the shortest survival. Class 1 tumours (1A and 1B) have a better prognosis compared to class 2 tumours, with risk of metastasis being 2% for class 1A tumours, 21% for 1B tumours and 72% for class 2 tumours [191].

In addition to providing clinical value, the expression profiles also provide valuable insight into UM pathogenesis. The class 1 tumours resemble normal uveal melanocytes and low grade spindle cell melanoma. On the other hand, class 2 tumours correspond to high-grade melanomas with more epithelioid cells and a reduced expression of melanocytic genes. Their transcriptomes resemble those of primitive neural/ectodermal cells [192]. BAP1 gene are frequently mutated in tumours with monosomy 3 and the morphology of cultured class 1 UM cells changed into a class 2-like epithelial phenotype when the *BAP1* gene is depleted. This is consistent with later work implicating BAP1 in maintenance of key aspects of melanocytic differentiation where its loss can lead to malignant progression [115].

Worley et al [193] investigated the expression patterns of micro-RNA expression using a microarray based approach and showed that a microRNA signature can also cluster tumours easily cluster into two groups corresponding to the GEP classification (class 1 and class 2). Overall, miRNA expression levels were significantly higher in class 1 compared to the class 2. However, there were only 6 miRNAs (let-7b, miR-199a, miR-199a*, miR-143, miR-193b and miR-625) that were identified as strong discriminators of class and these were all upregulated in the class 2 tumours. The most significant discriminating miRNAs were let-7b and miR-199a, and their differential expression was validated with q-PCR. The let -7b family is involved in lineage programming and let-7b plays a role in cell fate determination and neural crest development [193]. Predicted targets of let-7b are the melanocyte-related developmental regulator box 3, Rb1 and CCND1 genes. Not much is known about miR-199a, whose predicted targets include MITF and other melanocyte related genes including *CAV1*, *FZD6* and *PAX3*. Many additional studies have identified novel miRNAs in UMs [194–197], however, concordance is only seen for a small number of miRNAs across different studies. These discrepancies can be attributed to quality of samples, variation in profiling platforms and different protocols for sample processing, tumour heterogeneity and other obscure factors. A comprehensive understanding of miRNA dysregulation and abnormalities associated with UM is still to be determined.

1.4.5 Mutational signatures in uveal melanoma

UMs are characterized by relative genomic stability, with a low mutation and structural variation rate. The single nucleotide variant (SNV) mutation rate is less than 1 per Mb, compared to most solid tumours including adult cancers and other types of melanomas such as cutaneous and acral[198–201]. Interestingly the somatic mutation rate in UMs is similar to paediatric cancers such as rhabdoid tumors and medulloblastomas[180].

Epidemiological studies have implicated population specific factors such as light skin colour and blue eyes in UM. However, the role of environmental factors such as UV light exposure have not clearly been delineated [77,78,81,83,91]. Two independent groups recently examined the role of UV radiation at the molecular level and found an absence of a UV-induced mutational signatures in UM [180,199]. Both studies looked at whole genome sequencing data from primary UMs and analysed the mutational signatures based on single nucleotide variants (SNVs). They looked for a signature of UV-induced damage that is characterized by C>T transitions at the 3' position of pyrimidine dimers [180,199,202]. Although, the C>T transition were the most common type of substitution in UMs, they accounted for ~35% of the lesion and were not restricted to the 3' position of pyrimidine dinucleotide [199]. Thus, based on molecular data, UV-induced DNA damage does not appear to play a role in the aetiology of UM. Other major mutational signatures described in UM, based on the classification of Alexandrov et al [203] include, signatures 12/16, signature 1B, signature 3 and signature 6 [180,198]. Signature 12 is a feature of T>C mutations at N_pT_pN trinucleotides and signature 16 is characterized by T>C mutations at A_pT_pA, A_pT_pG and A_pT_pT. Changes at these trinucleotides show strong transcriptional strand bias and are thought to result from transcription coupled nucleotide excision repair [203]. Signature 3 exhibit more or less equal representation of all the 96 substitutions and are associated with a DNA double-strand break repair ("BRCA1/2 signature) [203]. Signature 6 is characterized predominantly by C>T transitions at the N_pC_pG trinucleotide and is strongly associated with inactivation of DNA mismatch repair genes [203]. Interestingly, there was no significant association of signature 6 and the presence of *BAP1* mutations in the samples analysed [198]. This is surprising since there are some reports that the BAP1 protein plays a role in DNA repair [204]. Signature 1B is characterized by C>T substitutions at the N_pC_pG trinucleotide and is not well separated from other cancer signatures. It arises from the spontaneous deamination of 5-methylcytosine and is correlated with the ageing process [203]. As a contrast, mutational signature in cutaneous melanoma are dominated by signatures 7, which are signatures of ultraviolet radiation

exposure [205]. Signature 7 is conspicuously absent in UM [198] consistent with lack of epidemiological evidence that UV is a risk factor in this cancer. Signature 7 shows a higher prevalence of C>T substitutions on the untranscribed compared to transcribed strand and may be formed through UV exposure of pyrimidine dimers and other lesions that are then repaired by transcription-coupled NER [203]. In addition, analysis of genes involved in cutaneous melanoma such as *BRAF*, *NRAS* and *NF1* revealed no somatic mutations in UM [180,206]. Thus, taken together these finding support the notion that uveal and cutaneous melanoma share different molecular aetiology.

For many years, causative genetic mutations in UM were a mystery, characterized by the absence of activating oncogenes such as *NRAS*, *BRAF* and *MYC* and inactivating mutations in the tumor suppressor genes such as *INK4A* and *PTEN* commonly found in melanomas [207,208]. Since then, the mutational profile in UMs have been investigated both at coding level by whole exome sequencing and more recently at genome wide scale using whole genome sequencing. Currently, the genetics of UM is dominated by frequent mutations involving five major genes, including *GNAQ*, *GNA11*, *BAP1*, *SF3B1* and *EIF1AX* (described below).

1.4.6 Driver genes in uveal melanoma

According to Volgelstein et al [209] , genomic landscapes of human cancers consists of genes altered in a high percent of tumours ("mountains") and much larger numbers of genes altered infrequently ("hills") that "drive" tumorigenesis (driver genes). Consistent with this, UMs display similar characteristic patterns of somatic mutations that have been identified from sequencing primary tumours samples. Several key cancer driver genes are mutated consistently across various studies. These are discussed below.

Alterations in G-protein coupled receptor pathways in UM

Mutations in *BRAF*, *NRAS*, *HRAS* and *KIT* are commonly found in the melanomas of the skin. However, UMs are characterized by the absence of mutations in these oncogenic drivers, even though they exhibit a similar constitutive activation of the MEK/ERK/ERK pathways [210,211]. One exception is iris melanoma where one study identified *BRAF* mutations (T1799A) in nearly half of all tumours suggesting genetic differences between anterior and posterior UMs [210]. The first described mutations in UM were activating mutations in genes

encoding the G protein-coupled receptors (GPCR), G Protein Subunit Alpha Q (*GNAQ*) and G Protein Subunit Alpha 11 (*GNA11*). These mutations occurred in a mutually exclusive manner, mostly targeting codon 209 and in some cases codon 183 [212,213]. Raamsdonk et al (2010) discovered in a screen of 30,000 mice that germline mutations in the orthologous genes, $G\alpha_q$ (V179M) and $G\alpha_{11}$ (I63V) lead to dermal hyperpigmentation [41]. The microscopic appearance of the skin in these mutant mice resembled blue nevi of humans [214,215]. A subsequent investigation of mutations in *GNAQ* and *GNA11* in blue nevi and other cutaneous melanocytic neoplasms revealed Q209 hotspot mutations in *GNAQ* and *GNA11* in blue nevi, and in primary and metastatic UMs [212,213]. Hotspot mutations affecting codon R183 compared to Q209 were less prevalent and were found in 2% in blue nevi and 6% UMs [212]. In total, 83% of all the UMs analysed have either a somatic mutation in *GNAQ* or *GNA11* genes occurring in a mutually exclusive manner [212]. Since their discovery, these mutations have been confirmed by several studies and are well established as UM specific oncogenic drivers [180,216–219]. Mutations in either *GNAQ* or *GNA11* are found in nearly 85% of all UMs, including benign uveal nevi, primary and metastatic tumours [212,216]. The *GNAQ* and *GNA11* mutations are not correlated with survival and are present in the majority of tumours irrespective of their cytogenetic status. These mutations are also not correlated with the gene expression profile based class status of UM, nor with *BAP1* mutation status, that is associated with metastatic risk [220,221]. Thus, *GNAQ/GNA11* mutations appear to be an early or initiating event in UM progression. Recently, rarer mutations in two additional oncogenes were identified in UM using whole genome and exome sequencing data. These were recurrent mutations in cysteinyl leukotriene receptor 2 (*CYSLTR2*) and phospholipase C β4 (*PLCB4*), targeting codon 129 (L129Q) in *CYSLTR2* and codon 630 (D630Y) in *PLCB4*. These mutations were identified in less than 5% of UMs [198,222] and occur in samples without mutations in *GNAQ* or *GNA11*[198,222].

GNAQ and *GNA11* encode the alpha subunits of the transmembrane guanidine nucleotide-binding protein (G-proteins), $G_\alpha q$ and $G_\alpha 11$ respectively and are important intermediates between membrane-bound GPCR and intracellular signalling cascades. In a normal cell, G-proteins are inactive when bound to guanosine diphosphase (GDP). When a ligand binds to the receptor, it changes its conformation and GDP is converted to guanosine triphosphase (GTP). This renders the G-protein active to downstream pathways. The GTPase activity of the G-protein is rendered inactive when GTP is hydrolysed to GDP [223]. However, the most common *GNAQ/GNA11* mutation affecting codon 209, is crucial for their GTPase

activities, and prevents the return of G-protein to an inactive state. This oncogenic mutation leads to constitutive activation of downstream MEK-ERK1/2 and Rho/Rac/YAP pathways, which promote cell proliferation [224,225]. Other oncogenic alterations at the level of the GPCR involve mutations in the *CYSLTR2* gene. This encodes the seven transmembrane GPCR that activates the Gq subunit [222]. The L129Q mutation in *CYSLTR2* affects the third transmembrane helix of the receptor, promoting ligand independent activation of the GPCR [226]. The *PLCB4* encodes a phospholipase enzyme that catalyzes the conversion of phosphatidylinositol 4,5-biphosphate (PIP2) in the plasma membrane to inositol 1,4,5-triphosphate (IP3) and diacylglycerol (DAG) [226]. Consequently, DAG and IP3 mediate signal transduction by downstream activation of protein kinase C (PKC). Mutations affecting codon 630 of *PLCB4* affect the Y-domain of the conserved catalytic core of *PLCB4* that controls the signal transduction [198]. Collectively, mutations affecting genes *GNAQ/GNA11*, *CYSLTR2* and potentially *PLCB4*, affect the GPCR signalling in UMs and raise the possibility of identifying therapeutics that target these mutations as well as downstream pathways affected by them.

BAP1 - metastasis associated driver gene

Monosomy 3 was the best predictor of predictor of metastasis in UM for a long time until the identification of a gene expression profile (GEP) based classification [36,227]. Based on the GEP, tumors can be classified as class 1, with a low propensity to metastasize and as class 2, which are aggressive and exhibit a high degree of metastasis. The presence of a class 2 GEP is usually associated with monosomy 3 [42]. Although, these markers offered superior value for prognostication, the identification of a metastasis suppressor gene on chromosome 3 had attracted enormous attention. With the help of whole exome sequencing data, Bowcock and Harbour et al (2010) identified BRCA1 associated protein 1 (*BAP1*) mapping to chromosome 3p21.1 as a metastasis suppressor of UM [228]. Inactivating somatic mutations in *BAP1* gene were found in 84% of metastasizing class 2 tumours and in only 1 class 1 tumour (2%). Loss of BAP1 function is achieved by loss of one chromosome 3 and a second hit on the remaining chromosome 3, consistent with Knudson's definition of a two-hit hypothesis of a tumour suppressor gene. Germline *BAP1* mutations have also been reported, in a small subset of patients (< 5%) and are associated with larger tumours and higher rate of ciliary body involvement which are known risk factors for metastases [229]. Germline *BAP1* mutations are estimated to account for 2-3% of all UM patients, although a higher estimate of 8% has been reported [230]. The discovery of BAP1 as a metastasis suppressor has been confirmed by

various other groups, and the types of cancers arising as a result of predisposing germline *BAP1* mutations has been expanded to include malignant mesothelioma, cutaneous melanoma, melanocytic BAP1-associated intradermal tumours (MBAITs) and renal cell carcinoma. This is now collectively recognized as the *BAP1* familial cancer syndrome [229–233].

The *BAP1* gene encodes a nuclear-localized ubiquitin carboxy-terminal hydrolase (UCH) enzyme [234] and removes ubiquitin molecules from target proteins. BAP1 is part of the polycomb repressor complex and many functions have been ascribed to it. The balance of transcriptional control is achieved by ubiquitination of polycomb repressive complex and de-ubiquitination by PR-DUB and loss of BAP1 significantly alters the expression of several polycomb targets [230]. *BAP1* has been implicated in regulation of the cell cycle, DNA damage repair, cellular differentiation and transcription. BAP1 can affect the cell cycle by binding and deubiquitinating lysine residues in the transcriptional regulator host cell factor 1 (HCF1) [235]. HCF1 is a cell cycle regulator and associates with both activator and repressor transcriptional regulators that control cell-cycle progression [236]. Transcription factors such as E2F1 recruits HCF1 to specific promoters, where it regulates transcription. HCF1 sustains the complex formed between the chromatin modifying enzymes, such as histone methyltransferase and the transcription factors [236]. Ubiquitylation of HCF1 blocks the E2F-responsive promoter activity and de-ubiquitylation of HCF1 by BAP1 would remove this inhibition and promote cell proliferation. BAP1 forms a ternary complex via its association with HCF1 and transcription factor Yin Yang 1 (YY1) and regulates gene expression **(Figure 1.2)** [235]. BAP1 is found to be associated with a multiprotein complex, called the BAP1 complex and these include HCF1, ASXL1/2, O-linked N-acetylglucosamine transferase (OGT) and forkhead transcription factors FoxK1/K2 [237,238]. The protein partners of this core vary under different cell types and conditions. The interaction between BAP1 and HCF1 and OCT could play a role in recruiting polycomb repressive deubiquitylase complex to the target genes [239]. BAP1 also interacts with ASXL1/2 to form a polycomb group repressive deubiquitinase complex (PR-DUB). This complex is a critical regulator of stem cell pluripotency, embryonic development, self-renewal and differentiation [230,240]. In addition, *BAP1* may play a role in regulating DNA damage and repair via its interaction with BRCA1/BARD1. The E3 ubiquitin ligase activity of the BRCA1/BARD1 complex which regulates DNA damage response is modulated by de-ubiquitination of BARD1 by BAP1 [241].

In UMs, GEP based class 1 tumors are composed of more differentiated cells with similar gene expression to normal melanocytes, whereas class 2 tumors contain cells that have lost morphological features of melanocytic differentiation and their gene expression

signatures are enriched for genes expressed in primitive ectodermal and neural stem cells [192]. Functional experiments by our group to elucidate the effect of *BAP1* loss in cultured class 1 UM cells revealed that depletion of *BAP1* leads to loss of differentiation and gain of stem cell like properties, including expression of stem cell markers (NANOG and OCT4), increased capacity for self-replication and augmented growth in stem cell conditions and acquisition of a class 2 gene expression profile. However, this depletion did not result in increased proliferation, migration, invasion and tumorigenicity [116]. Furthermore, treating short term BAP1-mutant UM cell lines with histone deacetylase (HDAC) inhibitors reverted the de-differentiated state of the class 2 UM cells to a differentiated class 1 phenotype and restored the normal level of expression of melanocytic genes that were downregulated by *BAP1* loss [58]. The idea behind treating the *BAP1* deficient cells with HDAC inhibitors is that *BAP1* loss leads to hyper ubiquitination of histone H2A, which is shown to be reversed by using HDAC inhibitors [58]. Thus, transient treatment of *BAP1* depleted UM cell lines with HDAC inhibitors leads to reversal of the phenotypic effects of *BAP1* loss induction of melanocytic differentiation and arrest of the cell cycle. HDAC inhibitors also inhibits the growth of UM tumours in an *in vivo* xenograft model [58]. Thus, *BAP1* appears to function as a regulator of differentiation, where cells with loss of *BAP1* exhibit stem cell like qualities [116]. The exact mechanism of how *BAP1* loss promotes metastasis in UM is yet to be elucidated.

Figure 1.2: Schematic diagram illustrating interaction of BAP1 with HCF1 complex.

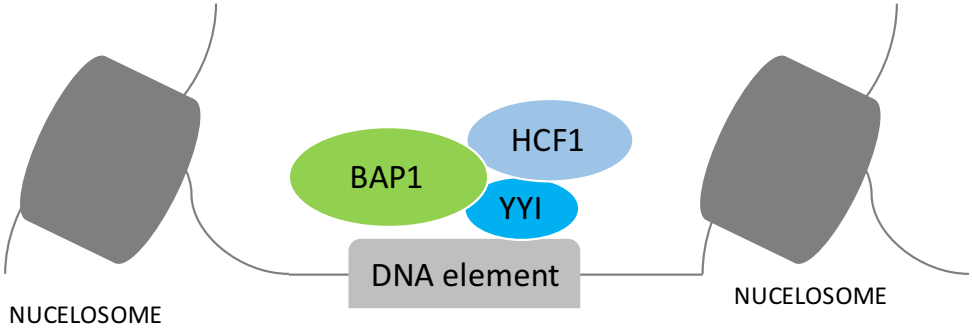


Figure adapted from Carbone et al.[241]. *BAP1* regulated gene transcription via its interaction with HCF1 and YY1.

SF3B1 – driver gene at transcriptional level in uveal melanoma

Our group has also shown that recurrent mutations affecting codon 625 of the splicing factor 3b subunit 1 (*SF3B1*) gene are found in nearly 24% of UM and are associated with favourable prognosis compared to *BAP1* [218,242]. *SF3B1* mutations are relatively specific to the GEP based class 1 tumours and are associated with younger patient age [218]. When adjusted for the impact of monosomy 3, patients with tumours with SF3B1 mutations and disomy 3 were associated with significantly worse prognosis and developed late metastasis [243]. Mutations in SF3B1 have subsequently been reported by number of studies and their frequency varies between 15% to 30% [199,218,219,242]. Interestingly, similar to GNAQ and GNA11, mutations in the SF3B1 gene are recurrent and predominantly affect codon 625 where the arginine residue (R) is changes to different amino acid depending upon the type of base change (R625C, R625H, R625G, R625L). This base is also altered in other cancers such as CLL and MDS [244,245]. *SF3B1* lies on chromosome 2q33.1 and evaluation of copy number of tumours with this mutation, shows no loss or gain of this locus [242]. Recently Yavuzyigitogle et al (2017) showed that tumours with *SF3B1* mutations are characterized by gain of chromosome 9q (24%), 6p (85%) and 8q (73%), and loss of 6q (52%), 11q (45%) and 1p (36%) [246]. Additionally, UMs with multiple chromosomal structural variants are more likely to be *SF3B1* mutants [246]. Martin et al reported that 29% of tumours with disomy 3 carry *SF3B1* mutations compared to 3% of tumours with monosomy 3 [219]. Mutations in *SF3B1* usually occur in a mutually exclusive manner with respect to the other UM driver genes *BAP1* and *EIF1AX* (discussed below) [218]. Resequencing of other exons in the SF3B1 gene for mutations in additional hotspots such as those in codon 700 which is found in other haematological and lymphoid malignancies [243,247,248] has revealed additional less common non-R625 mutations in UM [219].

SF3B1 is an essential part of the precursor-mRNA (pre-mRNA) splicing machinery and is a component of both the major (U2 small nuclear ribonucleoprotein or snRNP) and minor (U12 snRNP) spliceosome complexes [243]. *SF3B1* is involved in 3'splice site recognition and mutations affect splicing of target genes. *SF3B1* mutations may also affect DNA-damage repair in CLL [243]. Furney et al (2013) described a similar pattern of alternative splicing due to SF3B1 mutations in UM [199]. Differential alternative splicing of *UQCC* and *ABCC5*, and use of a cryptic splice site near exon 4 of the long coding RNA *CRNDE*, previously implicated in colorectal cancer, was identified [199]. Although Furney et al. (2013) showed diverse alternative splicing events in a large number of transcripts, including

alternative terminal exon usage, intron retention and cryptic splicing, a clear pattern of target genes affected by SF3B1 mutation was lacking [199]. Furney et al. (2013) showed that SF3B1 mutations are associated with alternative splicing patterns in UMs [199]. Differential alternative splicing of *UQCC* and *ABCC5*, and the use of a cryptic splice site near exon 4 of the long coding RNA (lincRNA) *CRNDE*, previously implicated in colorectal cancer , was identified [199]. Three independent studies by DeBoever et al. (2015) [249], Alsafadi et al. (2015) [250] and Darman et al. (2015) [251] elucidated the mechanistic role of *SF3B1* mutations in deregulation of canonical splicing. Some of the key points can be summarized as: (1) *SF3B1* mutant tumors exhibit global splicing defects by usage of an aberrant cryptic 3' splice site (3'SS) located 10-30 bases upstream of the canonical 3'SS (2) Significantly wildtype SF3B1 knockdowns and overexpression do not reproduce the effect of mutant SF3B1 and are proposed to lead to a "change of function" effect rather than gain or loss of function (3) The mutant SF3B1 preferentially recognizes an alternative branch point (BP) upstream of the canonical BP site (4) Aberrant splicing is observed in patients with hot-spot R625 mutations and the tumors with non-hotspot mutations cluster with wildtype tumours (5) Nearly 50% of the aberrantly spliced mRNAs are subjected to non-sense mediated decay and lead to downregulation of canonical mRNA or proteins. Taken together these studies provided a better understanding of the mechanism underlying the splicing alterations altered due to SF3B1 mutation, however, specific targets leading to tumour formation have not been identified.

EIF1AX – driver gene at translational level in UM

Mutations in the eukaryotic translation initiation factor 1A (*EIF1AX*) gene, located on chromosome Xp22, were recently discovered by whole exome sequencing of UMs [252]. EIF1AX is an essential in the initial phase of translation of mRNA to polypeptide in eukaryotic cells, the first step by which the mRNA is translated to proteins in eukaryotic cells by the transfer of methionyl initiator tRNA to the small ribosomal unit (40S) and stabilizes the formation of the ribosome to enable translation [253]. The reported frequencies of *EIF1AX* mutations in all UMs range between 14%-25% [198,216,218,219] and unlike the *SF3B1* mutations which frequently affects a single codon, *EIF1AX* mutations affect exon 1 and exon 2 of the gene. A few mutations affect splicing or lead to amino acid deletion in the N-terminal tail (NTT) of the protein, leaving the core protein with its RNA-binding domain intact [219]. In comparison to BAP1 mutations which are largely truncating and loss of function, *EIF1AX*

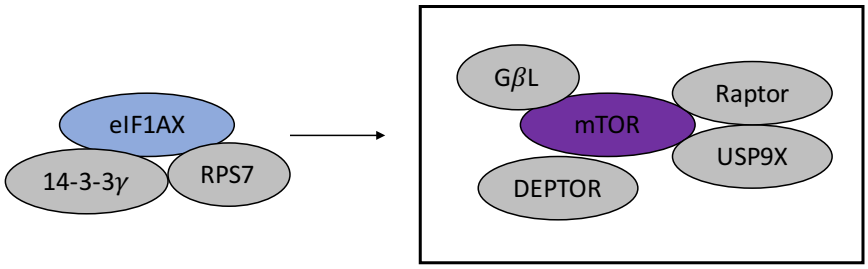
mutations are heterozygous missense variants in females and hemizygous in males. Evaluation of sex-specific effects of *EIF1AX* mutation revealed that all mutation target the active allele in females with UM, whereas the other allele is silenced at a result of X-chromosome inactivation [252]. Interestingly, mutations in *SF3B1* and *EIF1AX* were found to be more frequent in males compared to females with UM although the Y chromosome paralog of *EIF1AX* (*EIF1AY*) did not show any difference in expression compared to *EIF1AX* or harbour any mutation, indicating it does not contribute to the sex imbalance [252]. Immunostaining with an EIF1AX antibody revealed protein expression in the cytoplasm of all tumours irrespective of the mutation status[219].

Currently, the functional effects of the *EIF1AX* mutations are not clear, but are anticipated to have a global impact on production of protein. Chaudhuri et al have shown that mammalian eIF1a is essential for the transfer of the initiator Met-tRNAf (as part of the Met-tRNAf.eIF2.GTP ternary complex) to the 40 S ribosomal subunit in the absence of mRNA to form the 40 S preinitiation complex. This stabilizes the formation of the ribosome (joining the 40S and 60S to form the functional 80S initiation complex) around the AUG start codon [253]. Protein eIF1A encoded by *EIF1AX* may be linked to the mTOR pathway, via its interaction with RPS7 and 14-3-3γ **(Figure 1.3)** and may play a role in promoting protein synthesis and cell proliferation [254].

Mutations in the *EIF1AX* gene are predominantly found in non-metastatic class 1 UMs and are associated with good prognosis. *EIF1AX* mutations have a mutually exclusive relationship with *BAP1* and *SF3B1* mutations and as expected they are predominantly found in the disomy 3 tumours and rarely in tumours with monosomy 3 [218,219,255]. Combining information on *EIF1AX* mutation status with *BAP1* mutation status, enhances prognostication [218,255]. Mutations in the *EIF1AX* gene have been reported in other cancers such as thyroid and ovarian cancers [254,256]. The sequence of the *EIF1AX* gene harbouring mutations is highly conserved in eukaryotes and the regions with the substitutions correspond to portions of yeast eIF1A, OB-fold and ΔC domains. The assembly defects of the OB-fold mutation reduces 40S binding of eIF1A and impair ribosome function, reducing bulk translation [257]. This can be rescued by overexpressing eIF1, another eukaryotic translation initiation factor that communicates with eIF1A when bound to 40S during assembly of the preinitiation complex [257,258]. Another proposed effect of *EIF1AX* mutation is suppression of the recognition of near-cognate initiation sites or 5' proximal AUG codon to alter the relative use of different start

codons in mRNAs encoded by tumour promoting genes [219]. To understand the effects of *EIF1AX* mutation in UM warrants functional work.

Figure 1.3: Schematic diagram illustrating link between *EIF1AX* and mTOR complex.



EIF1AX forms a core component of the translation initiation complex in Eukaryotes and a potential promoter of mTOR activity part of the 14-3-3(gamma) and RPS7. Figure adapter from Hunter et al. [254].

Other low frequency driver gene mutations

Mutations in the promoter of telomerase reverse transcriptase gene (*TERT*) are commonly found in cutaneous melanoma and are consistent with a typical UV-damage signature [259]. In UMs, mutations in the *TERT* promoter are very rare [216,260]. However, elevated expression of *TERT* is reported in tumours with a wildtype *TERT* promoter suggesting that a different mechanism operates in UM that affects telomere maintenance [261,216].

As the number of high-throughput studies increases, a number of rare mutations of unclear significance will be identified. This imposes the challenge of distinguishing truly pathogenic events from "passenger" events that do not providing any selective advantage but which are swept along during tumour evolution. Revisiting Vogelstein's analogy of "hills" and "mountains" to describe driver genes, genes with few mutations ("hills") will continue to be identified and methods based on mutational frequency to identify these drivers will not be reliable. However, majority of the key drivers have been identified in UM and are slowly providing insights into the mechanistic role of tumour progression.

1.4.7 PRAME as an independent biomarker

Much research in UM has been focussed on the development of prognostic markers that predict metastasis. As previously described, cytogenetic and gene expression factors predict the risk of metastasis and currently a more accurate 15 gene assay is commercially available (DecisionDx-UM^{TM}). This helps identify the patients with high risk of metastases so that they

can be treated with adjuvant therapies and kept under surveillance [57]. Majority of the metastatic events are from the GEP class 2 subgroup of tumours and a very small subset of the class 1 tumours (1B). There is a 5-year metastasis rate of 21% for class 1B tumours compared to 2% for the class 1A tumours [191]. Recently, Field et al investigate mRNA expression of the metastasizing and non-metastasizing class 1 tumours at a genome-wide level and found overexpression of cancer-testis antigen *PRAME* in the metastasizing group [262]. Genes expressed by tumours with high levels of *PRAME* (the $PRAME^+$ group) were enriched for functions related to chromosome maintenance, meiotic recombination and telomere maintenance [262]. This study was further expanded to evaluate the prognostic value of upregulated PRAME in a much larger cohort of samples comprising class 1 and class 2 tumours. This established a $PRAME^+$ threshold that would allow the categorisation of samples into those with high versus low PRAME expression [46]. The key findings of this study were: (a) The only clinical feature associated with $PRAME^+$ status was large tumours size suggesting that its transcriptional activation is induced later during tumour progression (b) $PRAME^+$ status is more highly correlated with 1B compared to 1A tumours (c) $PRAME^+$ is associated with specific cytogenetic alterations including 6p gain, 6q loss, 8q gain and 16q loss. This provides evidence for the prognostic significance of 16q loss; a region that has not been well studied [40], additionally $PRAME^+$ was found in tumours with isochromosome of 6p and 8q (d) PRAME proteins are associated with OSGEP and LAGE3, human orthologues of an ancient EKC/KEOPS complex, implicated in transcription, telomere maintenance and chromosome segregation [263]. Hence, aberrant PRAME expression may predispose tumours cells to isochromosome formation and aneuploidy that in turn promotes tumorigenesis (e) $PRAME^+$ is associated with SF3B1mutation directly and EIF1AX mutations inversely (f) C_pG sites around PRAME promoters are aberrantly hypomethylated in $PRAME^+$ class 1 and class 2 tumours and loss of this CpG methylation is correlated with transcriptional activation of PRAME.

PRAME is specifically enriched at transcriptionally active promoters bound by NFY at a CCAAT binding motif, [264]. Field et al showed that many genes that upregulated in GEP class 1 tumours with PRAME overexpression, contain NFY sites in their promoter and, are located within regions of 1q gain and 6p gain [262]. Furthermore, these genes play a role in meiotic recombination, telomere and chromosome maintenance. Taken together these findings suggest that upregulation of PRAME may cooperate with 1q and 6q gain to facilitate chromosome instability associated with tumour progression [262]. However, it is also possible that upregulation of PRAME occurs prior to these chromosomal alterations.

Although, PRAME is a tumour antigen that is expressed in variety of cancers and provides prognostic value in UM due to its strong association with metastases, its exact role in tumour progression is unclear. A recent study identified PRAME as a dominant repressor of retinoic acid receptor (RAR) signalling. PRAME binds to RAR in the presence of retinoic acid, and prevents ligand-induced receptor activation, repressing target gene transcription through recruitment of Polycomb proteins *EZH2* [265]. It has therefore been proposed that overexpression of PRAME in cancer can confer growth and survival advantages by antagonizing RAR signaling. Since PRAME is a marker for poor outcome, it is likely to play a role in tumour progression rather than early oncogenic transformation. Further work is needed to elucidate the mechanistic role of PRAME in UM pathogenesis.

1.5 Perturbed pathways in uveal melanoma

Cancer development is a multistep process and can be defined by several features; Some of these "hallmarks of cancer" [266] that can be demonstrated in UM include: (1) resisting apoptosis (2) evading growth suppressors (3) self-sufficiency to grow (4) replicative immortality (5) sustained angiogenesis and (6) invasion of tissue and metastasis. Additionally, other emerging hallmarks include avoiding immune checkpoints and deregulating cellular energetics [266]. As indicated in the earlier section, UM arise from melanocytes in the uveal tract, that have differentiated from the non-pigmented neural-crest derived melanoblast precursors. The transition from melanocyte to metastasizing UM could be a combination of various genetic and epigenetic alterations. Several molecular pathways have been implicated in UM which help with tumour grow and progression. Inhibition of cell cycle control and apoptosis is mediated via retinoblastoma (Rb) and p53 pathways, that are functionally inhibited in UM. However, specific mutations targeting either *RB1* or *TP53* are lacking [267–270]. The Rb tumour suppressor pathway is disrupted by either hyperphosporylation of Rb, overexpression of cyclin D1 or inactivation via hypermethylation of the *INK4A*/CDKN2A/p16 promoter [267–272]. Hypermethylation of the *CDKN2A* promoter has been reported previously in UM, however with varying frequencies of between 4% and 33% [268,271–275]. Thus, its role as an additional mechanism for Rb inactivation is unclear. The p53 pathway is inhibited downstream of p53 activation and may be a consequence of *MDM2* overexpression which is seen in UM [267,275]. Defects in the BCL2 family of apoptosis regulators represent another way of evading apoptotic regulation and controlling cell survival

in UM. BCL2 is a direct transcriptional target of MITF and is highly expressed in UM [267]. Furthermore the PI3K/AKT pro-survival pathway is constitutively activated in UMs to avoid apoptosis [276,277]. Activation of AKT has been shown to be associated with higher risk of metastatic disease in enucleated eye samples [276] and this association with metastatic risk is only seen when phosphor-AKT (Thr308) is expressed [277]. Insulin-like growth factor 1 receptor (IGF1R) interacts with insulin like growth factory to regulate cell proliferation, differentiation and apoptosis, is often upregulated and can activate the PI3K-AKY pathway [278,279]. *PTEN* is a negative regulator of PI3K-AKT pathways and its role as a tumour suppressor gene has been investigated. Although mutations in *PTEN* are rare, loss of heterozygosity of the *PTEN* locus is reported in 76% of tumours [280,281]. Additionally, *PTEN* inactivation is associated with increased aneuploidy and decreased survival in UM, suggestive of *PTEN* loss being a later event in UM progression [38,280]. The role *PTEN* in UM progression requires further investigation.

MEK/MAPK signalling pathways are constitutively activated in UM [210]. Although UMs are characterized by the absence of oncogenic *BRAF*, *RAS* and *KIT* mutations, commonly seen in their cutaneous counterparts [206,282,283] the discovery of the oncogenic mutation in the *GNAQ/GNA11* genes shed light into the mechanism of MAPK pathway activation in UM [212,213] Mutant *GNAQ* and *GNA11* proteins activate the downstream MAPK cascade signaling via activation of phospholipase C (PLC). This signals through a critical secondary messenger protein kinase C (PKC) stimulated via IP3-DAG. The phosphorylation of PKC activates the MAPK cascade through sequential phosphorylation of RAF, MEK1/2 and ERK1/2, which in turn act on TFs in the nucleus and regulates proliferation and apoptosis [284].

Upregulation of growth factors such as vascular endothelial growth factor (VEGF) is observed in UM, although the mechanism for this overexpression is not known [285,286]. Evaluation of VEGF expression and its association with tumour formation and metastatic characteristics has revealed that it correlates with the presence of necrosis but does not correlate with marker of angiogenesis [287]. , However serum VEGF has been shown to be increased with metastasis [287]. One of the driving forces of angiogenic stimulators is intratumoral hypoxia. This is associated with increased metastases in many cancers. Hypoxia-inducible factor 1 (HIF-1) is a transcription factor that mediates an angiogenic phenotype and is seen in class 2 tumours [192]. Tissue invasion and metastasis in UM is characterized by activation of E-cadherin and Wnt/beta-catenin pathways [288,289]. Further evaluation of genes that contribute directly to epithelial alterations in class 2 tumours has revealed

downregulation of *ID2,* a gene which is highly expressed in normal uveal melanocytes [114]. Other features of tumour invasion in UM include increased expression of matrix metalloproteinases (MMP) and downregulation of the tissue inhibitors (TIMPs), ALCAM expression, activation of NOTCH pathway and bi-allelic methylation of embryonal Fyn-associated substrate (EFS) [290]. Other tumour hallmarks such as avoiding destruction by the immune system can be mediated by the expression of PD-L1 on the UM cells which suppresses the IL2 production and impairs T-cell function [291] and downregulation of HLA class 1 expression [291]. Varying densities of tumour infiltrating lymphocytes and macrophages are proposed to lead to tumour promoting inflammation [292–295]. The recently described role of PRAME overexpression in UM in regulating a CTL response[46] requires further investigation.

The interplay of all these different pathways, including PI3K-AKT, MAPK/MEK signaling, mTOR and mTOR-IGF-1R blockade, in mediating UM pathogenesis requires more functional investigation and may help in identifying strategies for targeted therapy.

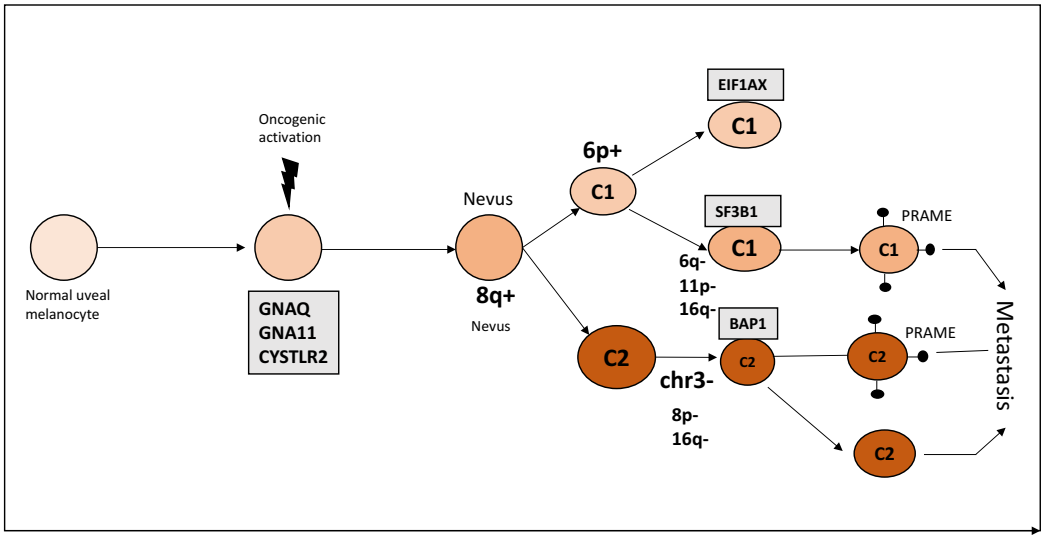
1.6 Current disease model

A provisional model can be suggested based on the data acquired on these tumours (**Figure 1.4**). The earlier initiating events include oncogenic activation of the Gα signalling pathway presumably in a normal uveal melanocyte by mutations affecting either a *GNAQ/GNA11* or *CYSLTR2* genes. This may trigger inappropriate cell cycle re-entry through activation of the MAPK pathway. Not all mutant cell clones will progress to form a melanoma, but rather undergo cellular senescence resulting in a nevus, or are eliminated by immune surveillance mechanisms [149]. One in 8000 nevi progress beyond this stage [296] to potentially form different GEP classes of melanoma. The class 1 tumours resemble the GEP of normal uveal melanocytes and nevi and class 2 exhibit melanocytic differentiation [192]. The GEP classes of tumours (Class 1 and Class 2) are associated with specific driver gene mutations that occur in a mutually exclusive manner.

Initiating mutations in GNAQ/GNA11/CYSLTR2 genes are followed by mutations in the BAP1/SF3B1/EIF1AX genes (referred in the current work as "secondary driver" genes since they appear to occur secondarily to the GNAQ/11 mutations). *SF3B1* and *BAP1* mutations occur mutually exclusively, in class 1 and class 2 respectively [218]. Within the class 1 tumours the *SF3B1* and *EIF1AX* mutations are mutually exclusive [218]. The class 1 tumours can be further differentiated based on the GEP, into the class 1A which rarely

metastasize and the class 1B tumours which represent slow growing tumours, with a higher propensity to metastasize than the 1As [129]. However, a more recently identified biomarker, *PRAME* overexpression can identify the slow growing class 1 tumours that are highly likely to metastasize [262]. Further, the GEP class 2 tumours are more likely to have overexpression of *PRAME* than class 1 tumours [46]. Cytogenetic alterations accompany these tumour groups based on gene expression and mutations. Gain of 8q is seen in almost all classes of tumours, irrespective of the GEP classification, although there is some evidence that it is more common in class 2 tumours [149]. Monosomy 3 is almost exclusively observed in the metastasizing class 2 tumours which have high *PRAME* and *BAP1* mutations [46]. Gain of 6p and is predominantly observed in the class 1 tumours or tumours with disomy 3 [38,39] and it is present in both *SF3B1* and *EIF1AX* mutant tumours [246]. Loss of 6p and 11p are seen in class 1 tumours with *SF3B1* mutations and almost never with *E1FIAX* mutants and is less commonly seen in class 2 tumours [246]. Loss of 8p occurs frequently in class 2 tumours with *PRAME* overexpression, often as isochromosome 8q. Loss of 16q occur in both class 1 and class 2 tumours that are associated with *PRAME* overexpression [46]. Loss of 1p is seen in close to a third of UM samples and are associated with *SF3B1* mutants [246]. Gain of 1q is associated with *PRAME* overexpression only in the class 1 tumours [46]. The metastatic tumours have a distinct gene expression profile similar to the GEP class 2 tumours [149].

Figure 1.4: Provisional disease model of formation of UM



Legends: C1 indicates class 1 tumours and C2 indicates class 2 tumours. Chromosome losses indicated by - and gains indicated by +.

1.7 Current study hypothesis and objectives

The provisional disease model provides broad insights into cytological and mutational events driving tumour processes. However, a comprehensive study with larger numbers of samples interrogated with high throughput technology is warranted to better understand UM tumorigenesis. The current knowledge indicates that while *GNAQ/11*, *CYSTLR2* and *PLCB4* mutant proteins are directly involved in the activation of signalling cascades, other key drivers including *BAP1*, *SF3B1* and *EIF1AX* play a role in transcriptional and post-transcriptional regulation of the cell machinery. The significance of these drivers is established at the level of recurrent mutation frequency, however, some less frequent, but important gene that are involved in tumorigenic processes may be missed due to small sample sizes and heterogeneity among different patients. Thus, much larger number of genes that are infrequently mutated but relevant to tumour biology ("hills") remain to be discovered. To date, most of the driver mutations in UM have been identified by non-systematic analysis of limited samples. However, this trend is changing with increasing number of studies employing whole genome or exome sequencing technology **(Table 1.3)**.

Table 1.3: Next gen sequencing based studies.

Year (month) of Publication	Study authors	Number of samples sequenced by Nex-gen Sequencing	Whole exome	Whole Genome	UM Driver Mutation identified	References
2010 (Dec)	Bowcock, Harbour et al.	2	2	0	GNAQ, GNA11, BAP1*	115
2013 (Feb)	Bowcock, Harbour et al.	18	18	0	GNAQ, GNA11, BAP1, SF3B1*	243
2013 (Aug)	Martin et al	22	22	0	GNAQ, GNA11, BAP1, SF3B1, EIF1AX*	220
2013 (Oct)	Furney et al	12	0	12	GNAQ, GNA11, BAP1, SF3B1	200
2016 (Apr)	Moore et al	136 (56 previously published + 80 TCGA samples)	136	0	GNAQ, GNA11, BAP1, SF3B1,EIF1AX, CYSLTR2*	223
2016 (Nov)	Royer-Bertrand et al.	33	0	33	GNAQ, GNA11, BAP1, SF3B1,EIF1AX	181
2016 (Dec)	Johansson et al.	28	15	13	GNAQ, GNA11, BAP1, SF3B1,EIF1AX, PLCB4*	199
2016 (Jul)	Field et al	104 (24 previously published + 80 TCGA samples)	104	0	GNAQ, GNA11, BAP1, SF3B1,EIF1AX	46

List of previously published studies utilizing next-generation sequencing technology and reported UM driver genes. * (star) indicates the study where the UM driver gene was first identified.

CNAs have extensively been characterized in UMs with various techniques such as Fluorescence In Situ hybridization (FISH), comparative genomic hybridization (CGH), multiplex ligation-dependent probe amplification (MLPA), SNP arrays and more recently with whole genome sequencing [180]. SNP arrays can interrogate both genotypes and copy number and can thus detect CNAs and copy neutral loss of heterozygosity events which cannot be detected by cytogenetics or CGH arrays. In addition, they have greater resolution compared to FISH [297].

The current understanding of cancer formation involves a stepwise acquisition of genetic alterations including point mutations, copy number alterations and fusion formation which affect critical genes regulating cellular growth and survival. Thus, the identification of genes targeted by these alterations will help accelerate progress in understanding the mechanistic basis of tumourigenesis and metastasis [266]. With the exception of BAP1, the target gene or genes underlying the recurrent cytogenetic events have not been identified. While all the major arm level chromosomal aberrations in UMs have been identified so far and key genes of prognostic significance have been established so far, a further systematic investigation of these CNAs is warranted to identify genes that play a role in mediating tumorigenesis and metastasis and to understand their role in the greater context of tumor biology. Although as described above, highly recurrent chromosomal gains and losses have been observed for some time in UM, most studies have been underpowered in sample size and genomic resolution to accurately identify the candidate genes involved in these events.

The underlying hypothesis proposed in the current study is that some copy number and structural alterations harbor genes that are highly selected for and are instrumental in causing defects in pathways that lead to UM pathogenesis. Thus, this study attempts to identify focal alterations, significantly altered in UM using a large cohort of primary tumors interrogated with high density SNP arrays and to identify potential candidate cancer genes relevant in UM. One of the limitations of the data analyzed in the current study is the heterogeneous sources of data generation. The SNP array platforms used here were from Affymetrix or Illumina and use different chemistries. Moreover, the error profiles of the array and methods to address the idiosyncrasies of the Illumina data are not well understood. Possible errors could come from the sample batch processing, array chemistry and sample heterogeneity. Thus, integrating results across multiple cohorts presented a challenge.

Thus, the main objective of the current study described in this thesis is to perform a comprehensive investigation of the genomic landscape of UM including copy number, mutations and the transcriptome. To achieve this, high throughput data has been analysed

with robust bioinformatics methods to gain further insights into the genetic events that occur in UM and to identify novel predisposing genes.

Major aims addressed in this study:

Chapter 3: Copy number profile in uveal melanoma

This chapter aims to perform a comprehensive copy number analysis of 182 UM primary tumours to call significant broad and focal copy number peaks that harbour candidate genes. Comparison of copy number peaks with pan cancer study to find CNAs common with other cancers and those specific to UMs. Further functional relevance of the copy number peaks will be accessed by integrating copy number data with gene expression from RNA-sequencing data followed up with pathway enrichment analysis of genes with copy number associated with change in expression to prioritise functionally relevant genes.

Chapter 4: Mutational landscape in uveal melanoma

Comprehensive investigation of primary UM will be performed by analysing the landscape of coding mutations in UM using a sample size of 120 samples (a previously published cohort) and sequencing data from additional samples (total N=131). In addition, this study describes an integration of copy number data from SNP arrays and variants identified from RNA-seq data with variants from exome sequencing to prioritise additional candidate genes that are likely to also play role in UM pathogenesis.

Chapter 5: Transcriptomic investigation of gene fusions in uveal melanoma

This chapter aims to investigate the global landscape of chimeric RNAs in primary UMs using RNA-sequencing data to identify canonical gene fusion events. Using newer algorithms, parsing and in-house developed filtering pipeline, this chapter attempts to confirm previously reported fusion transcripts and identify novel canonical gene fusion events that could play a role in UM pathogenesis.

Chapter 2. Data and Bioinformatics pipeline

2.1 Copy number analysis

2.1.1 Sample source and SNP array data

All samples used in the current study were enucleated specimens obtained from adult patients with informed consent. SNP-genotype data on 182 primary UM tumour samples and 87 normal samples were generated with different genotyping platforms from three different sources **(Table 2.1)**.

Table 2.1: Source of SNP array data

Cohort	Primary tumours Samples	Matched normal	SNP Platform
CC	57	0	Illumina - Human Quad66W
TCGA	80	80	Affymetrix SNParray 6
WASH-U	45	7	Illumina - Omniarray, CytoSNP12, OmniExpressExome_v1_3

In total of 182 primary tumors and normal UM samples

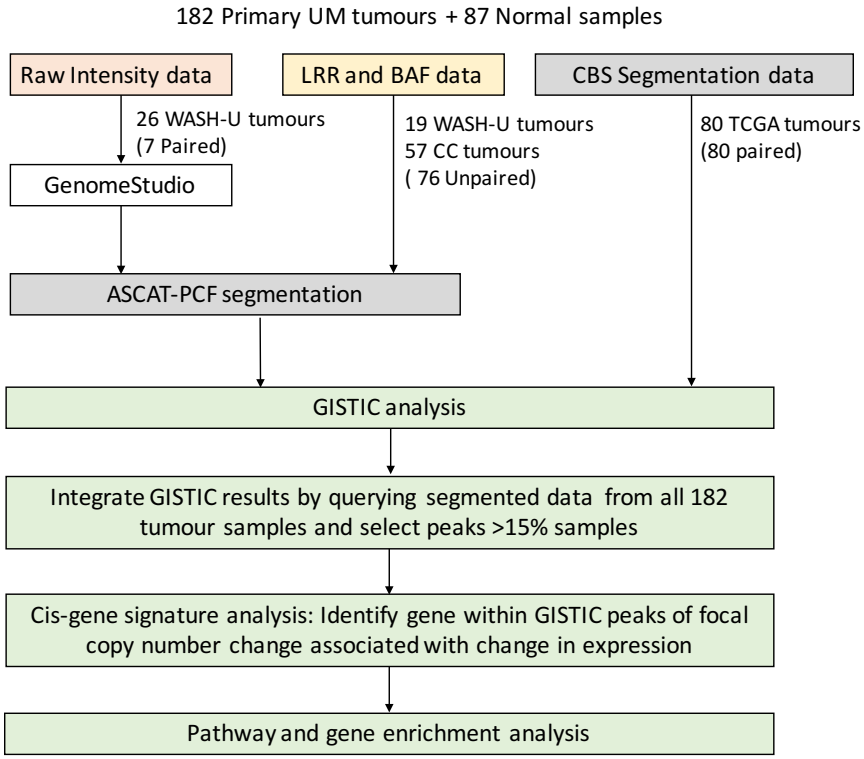
(1) Raw copy number data (LRR and BAF) for 45 tumour samples and 7 normal samples were obtained at Washington University in St Louis (WASH-U cohort). The collection of samples for all the tumour/germline DNA was approved by the Institution Review Board. Inclusion criteria for primary tumours included pathology-verified diagnosis and exclusion criteria include history of ocular brachytherapy or extensive tumour necrosis which can confound DNA profiling; The median age of patients was 60 years. Cytogenetic analysis of 45 primary UMs obtained at enucleation and matched blood on 7 samples was done using data obtained from three different Illumina SNP array platforms **(Table 2.1)**. Briefly, 200ng of DNA was used for the Infinium whole genome genotyping assay consisting of whole genome amplification, hybridization capture on BeadArray, array based primer extension SNP scoring and immunohistochemistry-based signal amplification was performed and data were generated on these samples comprising two intensity values (X, Y) for each SNP corresponding to each allele.

(2) Raw copy number data (LRR and BAF) was obtained for 57 primary tumour samples from Cleveland clinic, Ohio (CC cohort). The LRR and BAF data were downloaded using study accession number GSE44299. The median age of patients was 64 years (range 24-91 years)

(3) CBS segmented data from 80 tumour samples and 80 matched normal samples from the Cancer Genome Atlas Research Network (TCGA cohort) was downloaded from the dbGap portal. TCGA is supervised by the National Cancer Institute and the National Genome Research Institute and is funded by the US government. TCGA has several centers funded to generate and analyze data and the data are currently available for researchers in a three-level structure with tier 1 and 2 in a controlled access and the tier 3 without any restricted access. The segmented data for the 80 tumor and normal samples are available at tier 3 access. TCGA data on the UM cohort (TCGA code – UVM) was downloaded using the GDC data transfer tool (https://gdc.cancer.gov/access-data/gdc-data-transfer-tool). The median age of patients in the TCGA cohort was 61.5 years (range 22-86 years).

The data processing steps were performed at three different levels **(Figure 2.1)**: (1) Raw intensity data for WASH-U cohort; (2) LRR and BAF data for CC cohort; and (3) segmented data for the TCGA cohort. All the steps are described below.

Figure 2.1: Schematic representation of copy number analysis.



Schematic representation of the steps used for the analysis of copy number data.

2.1.2 Genotype data processing steps

The raw signal intensity data for data from WASH-C cohort genotyped on the Illumina Omniarray platform were processed to obtain the genotypes, Log R ratio and the B-allele frequency data. These were then used for downstream copy number calling. The allele specific image intensities were imported and normalized in Illumina's proprietary PC-Genome Studio software in order to call genotypes. The normalization algorithm was used to adjust for channel-dependent background and global intensities. It also removes outliers and scales the data. These two-color channels (X and Y) undergo affine transformations to make the data as canonical as possible with the homozygous SNPs positioned along the x and y intensity axis and followed by outlier removal, translational correction, rotational correction, shear correction and scaling correction[297]. After normalization, these intensities were transformed to polar coordinate plots of normalized intensities **R** ($R = X_{norm} + Y_{norm}$) and allelic intensity ratios **θ** ($\theta = 2/\pi * \arctan (Y_{norm} / X_{norm})$) where X_{norm} and Y_{norm} represent transformed signals from alleles A and B at a locus. Due to the unavailability of matched normal DNA for most of the samples, "single sample" analysis mode was used to compute log intensity ratios and allelic ratios by a comparison with canonical genotype clusters generated from large training sets of normal samples. Log R ratios (LRR) were computed from the normalized intensity R as the log (base2) ratio of normalized R value for the SNP divided by the expected normalized R values derived by interpolation of R value at the SNP's theta value for a sample relative to the canonical cluster generated by training the normal samples. The B allele frequency (BAF) was computed from the allelic intensity ratio **θ** for each SNP corrected for the canonical cluster position generated using normal samples. In other words, BAF represents the relative frequency of an allele at a locus in a reference population. These two transformed parameters, LRR and BAF are used for inferring copy number and LOH calls respectively. LRR measures copy number changes relative to the reference genome while BAF also referred to as allelic composition is the normalized measure of relative signal intensity between the two alleles (B/A). A LRR value of 0 can be interpreted as the presence of 2 alleles, LRR < 0 means loss of an allele or deletion and LRR > 0 means gain of an allele or amplification. A BAF of 0 represents allele 1 genotype (A/A or A/), whereas 0.5 represents a heterozygous genotype (A/B) and a BAF of 1 represents the allele 2 genotype (B/B or B/-).

2.1.3 Segmentation and copy number calling on SNP array data

Detection and defining copy number boundaries require a statistical framework that examines trends across multiple adjacent markers that share the same copy number and detects changes between two neighboring regions. Thus, these algorithms segment or delineate chromosomes into equal copy number difference regions. The array based data (LRR) for each marker represent the intensities from test and normal reference samples, assumed to not have any copy number alterations. This analysis was performed with a bioinformatics method called Allelic specific copy number analysis of tumour (ASCAT) [298] which takes into account the tumor ploidy and non-aberrant cell admixture to generates a genome-wide allele specific copy number profile for each tumor. ASCAT was applied to the LRR and BAF data from 102 primary tumours samples genotyped with the Illumina platform in the WASH-U (N=45) and CC (N=57) cohorts. ASCAT uses an Allele-Specific Piecewise Constant Fitting (ASPCF) algorithm to pre-process the LRR and BAF data. ASPCF algorithm is an extension of the univariate PCF algorithm [299] and takes as input the LRR and BAF data. It fits a piecewise constant regression functions to both the LRR and BAF data simultaneously to make the change points occur at the same position in both, iterating separately to each of 40 genomic regions corresponding to the chromosomal arms from 1 to X. Briefly, ASPCF tries to optimise partitioning of the genome into segments each containing consecutive probes to find the partition that minimizes the penalized optimization criterion comprising the number of segments, goodness of fit to the LRR data and the BAF data and a penalty variable for changes in the function due to discontinuities in the signal. After determining the change points and fitting the piecewise constant function the segment mean deviation from 0.5 is calculated. The ASPCF smoothed data are then used as input for the ASCAT algorithm to estimate the aberrant cell fraction and the tumour ploidy along with the allele specific copy number calls. ASCAT profiles can also be used to investigate copy number neutral loss of heterozygosity (CN-LOH) and LOH events by allowing the measurement of allele levels. A copy neutral event in the current analysis is defined as allelic bias for a SNP in the germline such that the total copy number does not change from tumour ploidy (differ more than 0.6) and copy number of A allele differs from copy number of B allele.

All steps were performed with ASCAT (version 2.4.2) package in the R computing environment (version 3.3). The following steps were involved **(Appendix 1 code 2.1)**: (1) The LRR and the BAF data for matched tumour and normal samples were loaded using the **ascat.loadData** function; (2) A GC correction step was performed to account for

the genomic 'wave artefacts' affecting the SNP arrays since this can affect the performance of copy number calling. The wave artefact are correlated with GC content of the surrounding genomic region and a wave adjustment procedure was applied to improve detection of copy number variation; (3) In the case of unmatched samples where only the tumour data were available, the germline prediction function was used to infer the germline genotypes from the tumour data and SNP array platform as input using the function **ascat.predictGermlineGenotypes**; (4) The ASPCF segmentation algorithm which calculates the allele specific copy number was applied to each sample with the **ascat.aspcf** function. The output file contained the rounded copy number values of A and B allele, the ploidy and aberrant cell fraction, segments with chromosomal start and stop positions and raw segmentation for each sample.

2.1.4 Broad copy number calling and unsupervised hierarchical clustering

The segmentation file for each sample was then used for calling broad and focal copy number aberrations. The segments can be classified as amplified or deleted based on a threshold of 0.15. The deleted or amplified segments affecting more than 50% of the chromosomal arm length were classified as broad arm level alterations. To identify the underlying pattern of copy number across all the tumours in an unbiased manner, unsupervised hierarchical clustering was applied. Here, an agglomerative method was used for clustering. First, the distance between two objects are calculated and followed by clustering using a bottom up (agglomerative method) approach where starting with each object in its own cluster, the best pair (cluster distance) is merged into a new cluster and repeated until all clusters are fused. A heatmap of the data with dendrogram (cluster tree) helps to visualize the inherent pattern in the data.

For hierarchical clustering, all the tumours were assessed for the presence (indicated by "1") or absence (indicated by "0") of each copy number alteration (samples, N=182) and a CNA status matrix was constructed. Samples with deletions were marked as "-1" and samples with amplification were "1" and the distance matrix was constructed using "Euclidean distance" method. This was passed to the R function 'hclust' for average hierarchical clustering and the number of clusters were chosen by visualizing the clustered heatmap. The 'cutree' function in R was used to extract the clusters. Association between the

metastatic outcomes and clusters were tested using chi-square statistic. Copy number segment heatmap was visualized with IGV [300].

2.1.5 Focal copy number analysis with GISTIC

The GISTIC algorithm [301] (version 2.0) was applied to the ASCAT segmented data to further narrow down the chromosomal regions of interest. GISTIC identifies somatic copy number alterations (SCNA) that are significantly altered at a higher frequency than regions which are selectively neutral or weakly deleterious 'passenger events'. It does this by evaluating the frequency and amplitude of the observed events. The algorithm performs 4 main steps:

1) Identification and separation of SCNA: The segmented profiles are deconstructed to identify the underlying SCNA by modelling the background rate of these alterations and separate out the arm-level and focal SCNAs based on length.
2) Assigning scores to each SCNA: The G scores are calculated as negative log of the probability of event occurring due to chance for a given background rate. To assign peaks that are statistically significantly, a *P*-value is computed for each marker by comparing the G score at each locus to the null distribution of random background score distribution generated by recalculating a G-score across all combinations of permutations of the marker within each sample.
3) Defining genomic regions to identify independent significant events: This step involves identification of most significant peak regions and then the use of an "arbitrary peel-off" method to prioritize independent potential targets under a significant peak. The segment scores of the most significant peaks are split among multiple potential peaks and iteratively subtracts segments covering the peaks until no region has an adjusted score that exceeds the significant threshold.
4) Accurate definition of the peak boundaries: This step defines the peak boundaries with a RegBounder method such that target genes are included at a pre-defined confidence interval regardless of the event frequency or the number of samples. RegBounder models the expected local variation in the G-score (due to neighboring passenger events or error determined in the segmentation analysis) to define boundaries predicted to contain true targets with user defined confidence intervals for the analyzed peaks.

Three independent GISTIC analyses were run on segmented data from 182 primary UM tumors spanning all three cohorts (7 tumour/normal and 38 unmatched tumour samples from

the WASH-U cohort, 80 tumour/normal samples from the TCGA cohort and 57 unmatched tumour samples from the CC cohort). Input files for GISTIC run were formatted as per GISTIC 2.0 (software version 6.10) (ftp://ftp.broadinstitute.org/pub/genepattern/modules_public_server_doc/GISTIC2.pdf). The marker file was generated as per the required GISTIC file format specification by taking the first 3 columns of the CN file used as input to the segmentation algorithm that produced the segmentation file. The threshold cut-off was determined by estimating the noise in each cohort (platform) and the parameter was determined heuristically by analyzing the histograms of the segment copy number and finding the first valleys on either side of the central peak. Each platform exhibited batch effects that could affect the results. To account for this variation, different amplification and deletion thresholds were assessed to determine if using platform specific characteristic threshold would be better than using a common threshold. There were few differences when comparing platform specific and generic cutoffs. Finally, a common threshold of 0.15 was used across all platforms for both amplifications and deletions. In addition, a gene GISTIC algorithm is available as an option which calculates the significance of deletion and amplification at gene level instead of marker level. An "all_thresholded.by_genes" file was included in the output which contained the discrete copy number state for each gene based on the threshold for amplification and deletion to help in the identification of low-level and high-level copy changes. The parameters used to run the analysis are as follows: Gene GISTIC analysis = YES, Amplification and deletion threshold = 0.2, Join segment size = 10 markers, Cap-value = 2.5, Confidence level= 0.95, q-value-0.25, Broad analysis= Yes /No and Arm peel-off= YES. A copy number variants (CNV) input file was used to exclude known structural variants in the germline of the healthy population obtained from the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/home) and HapMap normals. Frequency plots for the broad copy number alterations were plotted using copy number explorer [302].

2.1.6 Integration of GISTIC runs from different platforms

Although GISTIC was applied to samples from all 3 cohorts to identify focal peaks as described above, it cannot readily integrate samples from different platforms. Thus, a complementary downstream processing step was performed to find the most relevant CNAs across different segmented platforms. After GISTIC was applied to all the 182 samples to identify focal and broad peaks, a frequency based approach was used to prioritize the GISTIC

peaks identified across platforms using the segmented data. The focal copy number alterations (FCNA) were combined from independent runs on all three cohorts and the breakpoints predicted by GISTIC for a given minimal amplification and deletion peak were queried across segments from all the samples to determine whether a sample had a gain or loss at that region respectively. For the presence of amplification or deletion, a score of "1" was given and "0" was given in case of copy number neutral event. The frequency of GISTIC amplification and deletion peaks, was then calculated across all the 182 samples. To prioritize the more relevant peaks that would be less affected by the inherent noise in the data due variation between different platform and those most likely to have functional effects, peaks with frequencies > 15% were selected for further downstream analysis with gene expression data.

2.1.7 Integrated analysis of copy number and gene expression

Integrative analysis of focal copy number alterations and gene expression was performed for the selected list of focal peaks identified from the GISTIC analysis, to prune the list of candidate genes that lie within the region of interest and to assess if these alterations induced a significance change in their expression. RNA-sequencing count data from 80 TCGA samples, with matched copy number data used for GISTIC analysis, were downloaded from TCGA (https://portal.gdc.cancer.gov/) and data formatting was performed with R (version 3.2). The count data was used to generate count per million (CPM) to normalize for different sequencing depth for each sample using edgeR package in R followed by filtering of low expressed genes using a CPM threshold of 0.2. Additional quality control steps were performed by examining the library sizes and distribution plots and TMM normalization was applied to the remove composite biases between libraries. Finally, a voom transformation [303] was applied to the data matrix to make the data comparable and test for differences between tumors with and without alterations.

To assess the association between copy number level and gene expression for all genes that lie within the peak boundaries 70 differential analyses were performed. This included all the alterations identified for follow-up. To this end, for each gene, a two-group t-test was used to assess the difference in expression level between the group with and without the focal alterations. Multiple hypothesis testing (MHT) corrected q-values [304] were derived for a union of all the tested genes with the FDR procedure using the "q-value" package in R.

Here, a q-value was used rather than an adjusted p-value since it gives a better measure of each gene's significance and can be universally interpreted between studies. The q-value gives us a measure of the expected proportion of false positives incurred when calling the tested feature (gene) significant [304]. A false discovery adjusted p-value implies 5% of significance test will result in false positives, however, the q-value false discovery rate procedure is optimised by considering the distribution of p-values in an experiment. This is calculated by estimating the height at which the p-value distribution in an experiment flattens out. Thus, a q-value helps to identify how many of the significant results are false positives. In this study, 1010 genes across all the selected GISTIC peaks were tested and a q-value threshold of 0.05 resulted in 405 genes. Hence, 20 out of the 405 genes can be expected to be false positives.

The "cis-signature" genes were defined as a set of genes within each peak with an FDR q-value less than or equal to 0.05. Functional enrichment analysis based on the hypergeometric distribution was then performed for the significant genes using Broad's MsigDB website (http://software.broadinstitute.org/gsea/msigdb/index.jsp). The union of all the cis-signature genes were tested for enrichment against 860 genes in the C2 gene set collection (KEGG and Reactome pathways) from the MdSigDB repository. For network analysis, the genes from significantly enriched pathways were fed into GENEMANIA (http://genemania.org/). This is an interactive web service that helps predict the functional relationship between genes such as physical interaction (protein interaction), expression in the same tissue or cellular location (co-localization) or that have similar expression levels (co-expression). The networks were explored to generate a pathway diagram with Cytoscape (version 3.5) [305].

2.2 Mutational analysis

2.2.1 Samples and Exome/RNA-seq data

The exome sequenced patient samples came from two cohorts, WASH-U cohort and TCGA, consisting of 120 tumours and 104 normal samples in total (**Table 2.2)**. The distribution was as follows: (1) A collection of genomic DNAs from 40 tumours and 24 matched normal samples obtained from blood were obtained from the WASH-U cohort. Additionally, 22 tumour RNA samples that had been subjected to RNA-sequencing were included for variant calling (data described below under gene fusion analysis). These samples were sequenced

and downloaded as FASTQ files from GTAC (Washington University in St Louis). All the samples collected as part of the WASH-U cohort were approved by the Institution Review Board at Washington University in St Louis. Inclusion criteria for primary tumours included pathology-verified diagnosis and exclusion criteria include history of ocular brachytherapy or extensive tumour necrosis which can confound DNA profiling. Whole exome data for 80 tumors and matched normal TCGA-UVM samples were downloaded from controlled access tier one using the TCGA data portal. After obtaining a user certification for data access from the NIH dbGAP portal, bam files for 80 tumor/normal pair were downloaded with the GDC tool.

Table 2.2: Source of exome sequencing data

Cohort	Primary tumours Samples	Matched normal
TCGA	80	80
WASH-U	40	24

In total of 120 primary UM samples were used in the analysis.

2.2.2 Data quality control and coverage statistics

The quality of the sequencing data were checked by running FastQC [306] on the fastq files from each sample. The quality parameters included base quality, base distribution, sequence quality, base content, GC content, read duplication and adapter contamination. Picard's 'CollectHsMetrics' tool was used to generate mean target coverage, percent of bases reaching coverage level and percentage of bases excluded for each sample. The interval file with the location of all the probes were downloaded from the Roche's website and modified appropriately to be used for generating the metrics.

2.2.3 Alignment to human reference

Following QC of the Illumina reads (FASTQ files), alignment of reads to the human reference genome (hg19) was performed with the Burrows-Wheeler Transformation (BWA) alignment algorithm [307] (version 0.7.12). The BWA tool is an aligner based on a backward search with the Burror-Wheeler Transform (BWT) method to index the reference genome. This decreases the memory usage compared to other aligners such as MAQ which is based on hash table using a k-mer. Hwang et al [308] performed a systematic evaluated the performance

of various aligner and variant calling tools and determined that the BWA aligner with different variant callers performed better than other combination aligners such as Novoalign (http://novocraft.com/) and Bowtie [309]. Although they used the BWA-MEM [310], a newer version of the BWA algorithm, in the current study BWA Aln was used, since it was not feasible to re-analyze all the data with MEM due to project time constrains. The reference (hg19) fasta file was first indexed, followed by the BWA alignment step. The MAPQ for unmapped reads were set to zero using Picard's 'CleanSam' on the aligned SAM format file **(Appendix 1 Code 2.2)**.

2.2.4 Post-alignment processing

After the alignment step the bam files were run through a series of steps prior to variant calling to remove unwanted reads, remove misalignments and correct variation in the base quality reported by sequencers. These includee a series of steps using Picard [311] and Genome Analysis ToolKit (GATK) tools [312] **(Figure 2.2)**:

(a) Conversion of SAM to BAM: The aligned SAM format file was converted to BAM format which is the binary form of SAM file. The advantage of this conversion is that all the information in SAM is retained. This saves storage space files are can be manipulated in a faster fashion **(Appendix Code 2.3)**.

(b) Add-read group information: This step added read group information to each read in the aligned bam file including, Platform, barcode, sample name, library, sequencing center etc. The reads groups were added with Picard's 'AddOrReplaceReadGroups' command **(Appendix Code 2.3)**.

(c) Fix-mate Information: This step looked at all the mate-pair information between each read and its mate pair to ensure they were in sync with each other. Although this step is optional and not needed in the newer version of GATK tools, this step helps in handling some of the downstream errors resulting due to malformed mate information. This step was run with Picard 'FixMateInformation' command **(Appendix Code 2.3)**.

(d) Removal of duplicate reads: Duplicate reads of a DNA sequence can arise during the library preparation step (e.g. PCR amplification step) and represent non-independent measurement of that sequence. Such reads can inflate the counting support for the variant allele. Thus, removing or marking these duplicate reads is an important step to

avoid counting them. This step was run with Picard 'MarkDuplicate' command **(Appendix Code 2.3)**.

(e) Local realignment: The insertion and deletions (INDELs) in reads can cause mapping tools to mis-align reads with mismatches. These artifacts can affect the base quality recalibration and affect variant detection. This step realigns reads around INDELs and corrects mapping errors. This step is no longer necessary with the newer version of the Mutect2 and Haplotype variant callers, however it was required for the previous version of Mutect. Here, GATK's 'IndelRealigner' command was used.

(f) Base quality recalibration (BQSR): The quality scores assigned to the individual base calls affects the variant calling algorithms. This step applies a machine learning method to model the errors produced due to systematic technical artefacts and adjust the quality scores. After realignment, the quality of the reads is recalibrated to correct for variation in base quality emitted by the sequencing machines to provide accurate quality score for variant calling. This step was run with GATK's 'BaseRecalibrator' command **(Appendix Code 2.3)**.

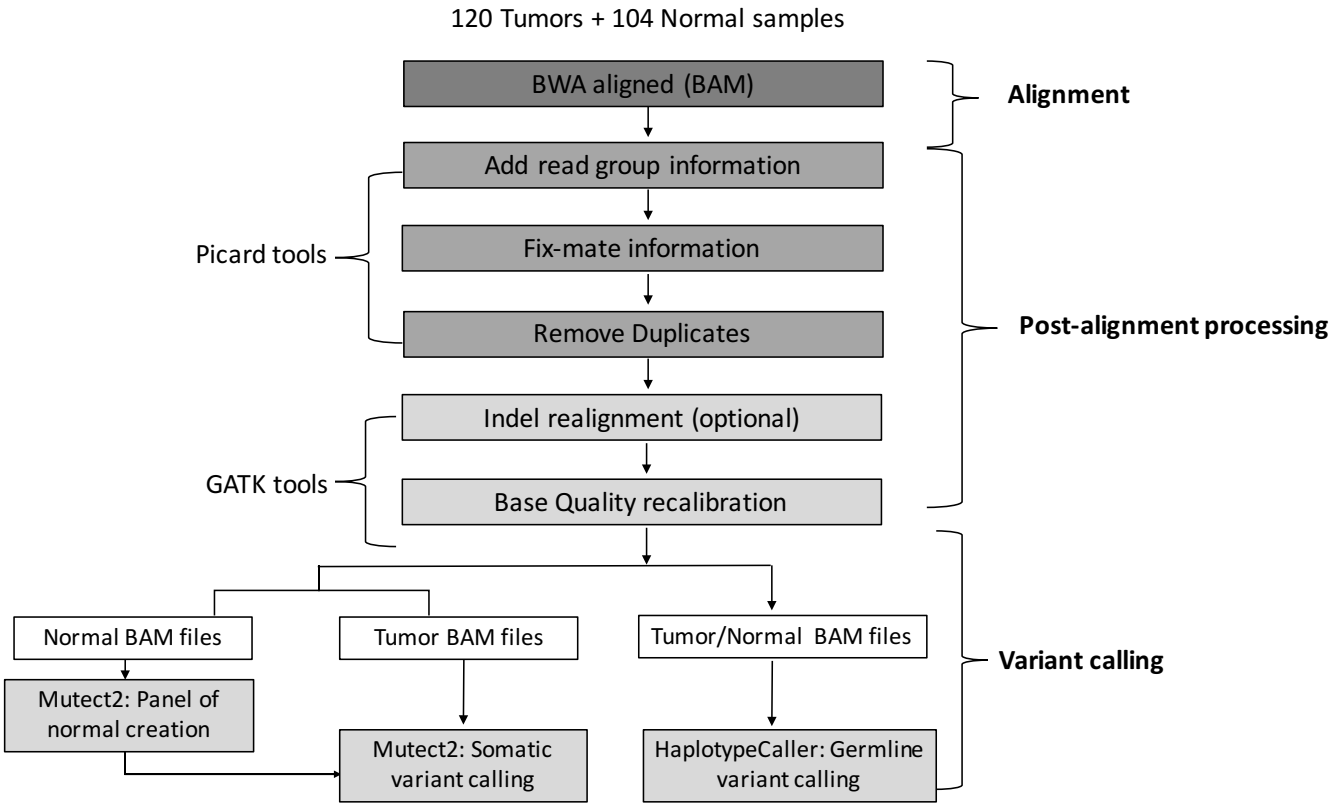
2.2.5 Variant calling from Exome sequencing data

GATK's recommended pipeline was used for somatic and germline variant calling **(Figure 2.2)**. Many variant calling tools are currently available, however, a comparison of variant callers for short read sequencing data have often indicated that the GATK tools show the best performance [313–315]. Recently, Hwang et al [308] compared the performance of thirteen variant calling pipeline based on gold standard reference variant calls from the Genome in a Bottle consortium. They found that the GATK HC method performed better than any other caller in calling INDELs, regardless of the alignment method used. However, for accurate variant calls, some caution needs to be exercised regarding homozygous SNP calls as GATK shows bias towards adding reference alleles [308].

For somatic single nucleotide variants (SNVs) and INDELs, MuTect2 [316] was used and for germline SNVs and INDELs, the HaplotypeCaller (HC) algorithm was used. MuTect is a highly sensitive method that applies Bayesian classifier to detect somatic mutations present at very low allele fractions. The underlying statistical framework of MuTect involves prediction of somatic mutation using two Bayesian classifiers: (1) For a given site, it is first determined whether the tumour is a non-reference; (2) For all the non-reference site, the second classifier ensures that it is not a germline variant allele. The

unavailability of matched normal samples for majority of tumour samples was circumvented with the Mutect2 pipeline which employs a "Panel of Normal" to identify the somatic variants across all the matched and unmatched tumour samples. The HC method is also capable of calling SNVs and INDELs simultaneously via local de-novo assembly of haplotypes in the regions of variation, making it more accurate in calling variants in regions which are difficult to call. For example, in regions where the different type of variants (SNV and INDELs) are close to each other. The main steps involved in the HC method include: (1) Determine the active region or regions of significant evidence for variation; (2) Build a De Bruijn like graph to re-assemble the active regions and identify the haplotypes present in the data. Each haplotype is then realigned against the reference haplotype using the SW algorithm to identify the variant site; (3) The likelihood of the allele for each potential variant site given the read data is calculated; (4) Finally, the program applies Bayes's rule to calculate the likelihood of each genotype per sample given the read data observed for that sample.

Figure 2.2: Schematic representation of exome analysis



Steps outlining exome analysis of 120 primary UM exome data from reference alignment to variant calling.

MuTect2 is an updated version of MuTect which combines the original method with the assembly-based machinery of HC. A key difference between Mutect2 and HC lies in the ploidy assumption for inferring the genotype likelihood and variant quality calculation, while HC relies on human samples being diploid. MuTect2 allows for a varying allelic fraction for each variant as in the case of tumour heterogeneity with less than 100% purity and the presence of multiple subclones. Since, HC calculates a variant likelihood that is not well suited for extreme allele frequencies (as in the case of variants with allelic ratios dramatically divergent from the expected diploid state), it is generally not suited for somatic variant discovery. Thus, a combination of HC and MuTect2 has been selected for the current analysis to identify germline *BAP1* mutations in UM, which have been reported in earlier studies [241]. The workflow for MuTect2 pipeline used here includes the following steps **(Figure 2.2)**: (1) Creation of a panel of normal variants **(Appendix 1 Code 2.4)**; (2) Tumour/Normal variant calling **(Appendix 1 Code 2.4)**; (3) Combine variants across all the VCF files **(Appendix 1 Code 2.4)**. The workflow for the HaplotypeCaller pipeline involves the steps: (1) Genomic VCFs (gVCF) generation per-sample with –ERC GVCF mode **(Appendix 1 Code 2.5)**; (2) Perform joint genotyping across all samples; (3) Perform variant filtering to remove variants that are likely to be artifacts (Variant Recalibration).

2.2.6 Variant calling from RNA-Sequencing data

Identification of genomic variants from existing RNA-sequencing data remains a challenge due to the intrinsic complexity of the transcriptome (e.g. splicing), which adds to the technical difficulty in calling variants and often leads to many false positive and negative calls. However, it can greatly increase the power to detect variants from known genes or help resolve the presence of variants with low coverage in samples with matched exome data. For the current analysis, GATK's recommended best practice workflow was used to call variants from RNA-sequencing data of 22 tumours **(Table 2.3)**. The steps involved were as follows: (1) Raw reads were mapped to the reference with STAR aligner [317], found to achieve highest sensitivity to both SNV and INDEL calls. This step uses a 2-pass approach where splice junction identified in the first alignment are then used to guide the final alignment **(Appendix 1 Code 2.6)**; (2) Read groups were added, duplicate reads were removed from the aligned bam files and sorted with Picard's 'SortSam'; (3) The reads were split into exon segments using GATK's 'SplitNCigarReads' tool. It also hard-clips any sequence overhanging into the intronic regions; (4) A Base Recalibration step was performed to correct where base qualities

have systematic errors due to sequencing artefacts; (5) The variant calling was performed with HC with 'dontUseSoftClippedBases' option to minimize the false positives and negative calls by taking into account the information about intron-exon split regions; (6) The variant filtering step involved applying hard filters to remove cluster of at least 3 SNPs within a window of 35 bases, Fishers strand value > 30 and Qual by Depth < 2. The filtered variants were then passed on to downstream processing involving variant annotation, filtering and variant prioritization.

2.2.7 Variant annotation (creation of MAF file)

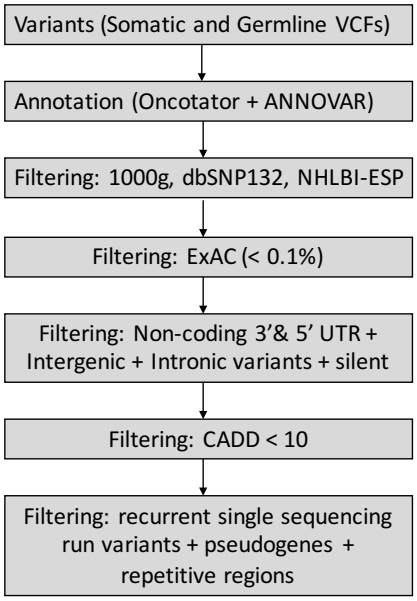
The identified variants (VCF file) were then run through annotation tools to get the genomic annotation, protein annotation, functional impact and variants were queried against various databases to get their allele frequency in normal and cancer specific populations. Currently, four major popular tools are available for variant annotation: Oncotator (http://portals.broadinstitute.org/oncotator/); ENSEMBL's variant effector predictor (VEP) [318]; ANNOVAR [319] and SnpEff [320]. McCarthy et al recently showed that the choice of annotation tool and the transcript set can have a significant impact on the classification of the variants [321]. They showed a concordance of about 65% between loss of function variants produced by ANNOVAR and VEP. Although some of the classification between tools disagree, filtering steps are performed on an aggregate set of variants and requires a common classification scheme. In the current study, Oncotator was used to create a MAF file with genomic annotations from GENCODE hg19 transcript reference, protein annotation and additional cancer annotations from COSMIC (http://cancer.sanger.ac.uk/cosmic), Cancer Cell Line Encyclopedia (https://portals.broadinstitute.org/ccle/home), Familial Cancer Database (http://www.familialcancerdatabase.nl/) and Clinvar (https://www.ncbi.nlm.nih.gov/clinvar/) and non-cancer annotations from dbSNP, 1000 Genomes and the NHLBI Go Exome Sequencing project. Further, all the variants were annotated with ANNOVAR, which provides additional information on the Exome Aggregation Consortium (ExAC) [322] frequency, Combined Annotation-Dependent Depletion (CADD) score [323] and region based annotations such as predicted TF binding sites, segmental duplication region and GWAS hits. Gene based annotation categories in ANNOVAR are more general and easier to build categories for making comparisons. The VCF files for both Exome and RNA-seq variants were used as input to run annotations with Oncotator (version 1.8) and ANNOVAR

(2016Feb01) **(Appendix 1 code 2.7)**. Additionally, genes were annotated as chromatin modifiers based on an in-house created list from the literature.

2.2.8 Downstream filtering and variant prioritization

After variant annotation, a series of filtering steps were performed to prioritize the variants important for cancer pathogenesis. The following filtering steps were applied **(Figure 2.3)**: (1) The polymorphic variants from 1000 genome, dbSNP132 and NHLBI-ESP databases were filtered; (2) The ExAC dataset, which is characterized by population allele frequencies from exome sequencing data on over 60,000 individuals provides a robust frequency estimates for rare variants. However, there exists ambiguity on the allele frequency to be considered for filtering these "too common" variants. In the current study, a pre-computed filtering allele frequency threshold of 0.1% described by Whiffin et al [324] was used to filter the variants; (3) All the non-coding variants in the 3' and 5' UTRs and intergenic regions were filtered; (4) The CADD scores are calculated by using different annotations such as SIFT, POLYPHEN etc and is quantitatively predictive of the deleteriousness and pathogenicity of variants. All the variants with CADD scaled C-scores (which ranks a variant relative to all possible substitution in the human genome) of less than 10 were filtered. This can be interpreted as variants that are not present in the 10% of most deleterious substitutions that can be present in the human genome; (5) Recurrent variants in multiple samples that are generated from same sequencing runs are likely to be artefacts and were filtered.

Figure 2.3: Workflow of post-variant calling steps.



Filtering was followed by variant prioritization, this involved the following steps: (1) Recurrent mutations across multiple samples; (2) Genes with multiple variants; (3) All the previously established known UM driver genes (*CYSLTR2*, *GNAQ*, *GNA11*, *BAP1*, *SF3B1* and *EIF1AX*); (4) The frequently altered regions with loss on chromosomes 1p, 3, 6q, 8p, 11q and 16q were examined for genes with secondary, hemizygous loss of function mutations indicative of tumour suppressors; (5) Variants from RNA-seq analysis were added to the list to identify additional frequent gene mutations; (6) Gene set enrichment analysis (GSEA) was performed using a list of chromatin modifiers gene set from the literature to test if the occurrence of mutations in the chromatin modifiers was statistically significant compared to observing the enrichment by chance. To perform GSEA, a fisher's exact test in R (version 3.3) was used to calculate the counts from the two input lists (exome data and gene set). Additionally, level 1 mutation calls using four different variant callers (Mutect, Varscan, Muse and somatic sniper) run on all the TCGA samples, generated by Broad Institute TCGA Genome Data Analysis Center (GDAC) Firehose analysis (doi:10.7908/C11G0KM9) were downloaded for independent verification of mutations identified in the current analysis. Published UM driver mutations (*GNAQ*, *GNA11*, *CYSLTR2*, *SF3B1*, *BAP1* and *EIF1AX*) in the WASH-U cohort and validated using sanger sequencing were used as complementary data to account for all mutations not identified in the current analysis.

2.2.9 Identification of significantly mutated genes (MutSigCV analysis)

As the sample size increases, the power to detect highly mutable cancer genes increases but this is at the cost of increasing the false positive rate. Therefore, we also performed an unbiased analysis for identifying genes that are mutated more often than one would expect by chance. To this end, Significant Mutational analysis in the MutSigCV algorithm (version1.2) [325] was applied to all the SNVs and INDELs identified from the exome analysis.. Briefly, MutSigCV scores every mutation against the background sample/patient specific background rate in which it is observed. The null distribution is calculated by combining all the patient specific null distributions. A p-value is calculated which compares the observed score to the null distribution. In addition, MutSigCV accounts for gene-specific differences in the background mutation rate. This background model incorporates patient specific factors (overall mutation rate and mutational spectrum such as transitions and transversions) and

genomic position specific factors (gene expression, DNA replication timing and HiC-based chromatin state estimation). Thus, mutations from all the chromosomes are aggregated and the total number of mutations per gene are then computed. This count is then converted to a score and compared with the background model to get the significance level. The input files to run the analysis include MAF file, coverage table file and covariate table file.

2.2.10 Mutational signature analysis

Mutational signature analysis was performed by decomposing all the single nucleotide substitution of nX96 dimensional matrix into r signatures, where n is the number of samples. The signatures are decomposed using non-negative matrix factorization (NMF) method and once decomposed compared against the known signature by Alexandrov et al [203]. The NMF algorithm is based on decomposition by parts and can reduce the dimension of the data to define the underlying pattern (similar to methods such as principal component analyses) [326]. Here the input to NMF comprises all the observed somatic mutations across all the tumours. It can detect signatures present in the somatic mutations across multiple tumours and determine the contribution of each signature to the somatic mutation in each tumour. Default parameters were tested where incremental number of mutational signatures are modelled, the signature with best possible value (maximum cophenetic coefficient) is selected. Cophenetic coefficient is a measure of how faithfully the clustered structures (pairwise distances between original unmodeled data points) are preserved thus, an indicator or stable reproducibility. All the steps were performed using R package Maftools [327] (version 1.2.30) in R computing environment (version 3.3.2).

2.2.11 Statistical analysis and plots

All the univariate statistical analyses were performed in R (version 3.3). Fishers exact test was used to evaluate the dichotomous categorical variables. Comparisons include mutational status versus gene expression profile, copy number status and metastasis status. A p-value of 0.05 was used to report statistically significant results. Plots were generated using R Package Maftools [327] (version 1.2.30) and GenvisR [328] (version 1.6.0) in R computing environment (version 3.3.2).

2.3 Gene Fusion analysis

2.3.1 Sample source and RNA-sequencing data

In total 22 primary UM tumour samples subjected to RNA sequencing by our group were interrogated for novel gene fusion events (**Table 2.3**). Of these, twelve samples were used to generate PolyA-selected RNA-sequencing data. Eight samples downloaded from a previously published study [242] and 4 additional samples (unpublished) sequenced at the same time. Total RNA from an additional 10 samples were sent to CNAG to generate RNA-seq data. The data used in the current study were initially generated as part of another project and used here to mine for potential novel gene fusion events. Therefore, two different RNA-seq libraries (polyA+ vs total RNA) were used. The adapter trimmed reads were downloaded and summary statistics on the raw data were generated with FastQC [306]. The samples were obtained from our collaborators and the study was approved by the Institution Review Board at Washington University in St Louis. To perform the RNA sequencing, total RNA was isolated from UM primary tumour biopsies with Qiagen AllPrep kits as per the manufacturer's instruction. The integrity and the quality were assessed on an Agilent 2100 Bioanalyzer (Agilent, CA) with an Agilent RNA 6000 Nano kit, as per the manufacturer's recommendation. Total RNA with high quality and integrity (RIN > 8) was sent to CNAG, Barcelona for library construction and to generate sequencing data. Briefly, 0.5 ug of quantified total RNA was subjected to ribosomal depletion with RiboZero Magnetic Gold Kit and fragmented by divalent cations at elevated temperature. This resulted in fragments of 80-450 nucleotides, with the majority peaking at 160 nucleotides. These were used to construct amplified libraries. This was followed by performing first strand cDNA synthesis using random hexamers as primeres. After second strand synthesis and end repair, the fragments then were ligated to Illumina barcoded adapters and the library was amplified with 15 PCR cycles. The size and quality of the library were assessed with the Agilent DNA 7500 Bioanalyzer assay (Agilent, CA) before being loaded into the flowcell for sequencing with the Illumina HiSeq2000 (Illumina Inc.) in paired end mode. The adapter trimmed sequencing data (FASTQ format) were downloaded and summary statistics on the read depth, read length and quality were generated with FastQC [306].

Table 2.3: UM primary tumour samples with RNA-seq data.

Sample	GEP class	RNAseq protocol	Number of reads	Read length
MM010	class 1	PolyA selected - unstranded	14296315	101
MM016	class 1	PolyA selected - unstranded	15069924	101
MM064	class 1	PolyA selected - unstranded	10569982	101
MM065	class 1	PolyA selected - unstranded	17608600	101
MM078	class 2	PolyA selected - unstranded	12191112	101
MM080	class 2	PolyA selected - unstranded	13261166	101
MM082	class 1	PolyA selected - unstranded	10711210	101
MM089	class 1	PolyA selected - unstranded	17243880	101
MM097	class 2	PolyA selected - unstranded	10639548	101
MM100	class 2	PolyA selected - unstranded	12754839	101
MM132	class 1	PolyA selected - unstranded	13230895	101
MM176	class 1	PolyA selected - unstranded	18107556	101
MM135	class 2	Total RNA - stranded	29426922	100
MM137	class 2	Total RNA - stranded	31172453	100
MM141	class 1	Total RNA - stranded	33318412	100
MM144	class 2	Total RNA - stranded	28325576	100
MM171	class 1	Total RNA - stranded	53498365	100
MM173	class 2	Total RNA - stranded	44025305	100
MM175	class 2	Total RNA - stranded	31076611	100
MM179	class 2	Total RNA - stranded	34318824	100
MM091	class 2	Total RNA - stranded	19600588	100
MM094	class 1	Total RNA - stranded	25302140	100

The tumour class, library type, read length and generated read counts are shown. PolyA selected RNA-seq data were from our previously published study [242].

2.3.2 JAFFA gene fusion detection pipeline

One of the major applications of RNA sequencing (RNA-seq) data has been identification of chimeric transcripts which can be generated by a gene fusion event at either DNA level (canonical gene fusions via chromosomal rearrangements) or RNA level (non-canonical gene fusions via trans-splicing or cis-splicing). Identification of these gene fusions events from RNA-seq data can be accomplished with the help of many software tools (~ 20 tools) freely available to the scientific community. Choosing the best optimum tool for fusion detection is a challenge since these tools are based on different algorithms, trained and test on different types of datasets and vary in the sensitivity and specificity of detected fusions. Additionally, each software requires different set of dependencies to be installed and have different computational requirements such as specific operating system, time consumption and memory usage.

Recently, Kumar S et al [329] reported an unbiased evaluation of twelve gene fusion detection tools (Bellerophontes, BreakFusion, Chimerascan, EricScript, FusionHunter, FusionMap, JAFFA, MapSplice, nFuse, SOAPfuse, and TopHat-Fusion) based on different datasets and compared the performance of these tools in terms of sensitivity of detecting fusions, positive predictive value and computational load (RAM and time consumption). In their comparisons based on the positive dataset containing 50 true fusions, JAFFA [330] was the most sensitive tool which detected 44/50 fusions (88% sensitivity) with relatively efficient memory usage and time consumption. Based on the positive predictive value, JAFFA performed relatively well (95.6%), although the computational performance was poor with respect to the negative and mixed datasets used for the evaluation. The performance of JAFFA drops with RNA-seq data of read length of less than or equal to 50 nucleotides (nt). However, for the current study this was not a problem since the reads lengths were greater or equal to 100nt **(Table 2.3)**. The primary goal of the current study was to detect as many true fusions events as possible. Thus, the JAFFA gene fusion discovery pipeline (JAFFA algorithm, version 1.09) was used to analyze the RNA-seq data [330].

JAFFA [330] is a recently developed fusion detection tool which compares a tumour transcriptome to the reference transcriptome, rather than the reference genome. JAFFA involves multiple steps starting from raw RNA-seq data and outputs a set of candidate fusions and their cDNA breakpoints. JAFFA is not a standalone tool, but is rather a pipeline relying on different tools implemented using a Bpipe platform and each step is stitched together with bash and R scripts. Depending on the read length of the RNA-seq data JAFFA can run on three different modes: (1) 'Assembly mode' (<70bp) which assembles short reads into contigs prior to fusion detection (2) 'Direct mode' (>100bp) which uses RNA-seq reads directly by selecting reads that do not map to known transcripts and (3) 'Hybrid mode' (between 70bp and 90bp) which is combination of assembly and direct modes and is the most sensitive of the three modes, however it is also the most computationally intensive. For the current analysis of long read lengths (>=100bp), the 'direct mode' option was used to perform the analysis.

The major steps performed by JAFFA pipeline can be described as follows **(Figure 2.4):**

(1) preliminary read filtering is performed to remove the intragenic and intronic sequences and this involves a two-step process. First the raw reads are aligned to the human reference transcriptome, GENCODE version hg19, and those reads that map to the transcriptome are

retained. In the second step, the unmapped reads are mapped to the human reference genome, hg19, where the reads that do not map concordantly to the genome are retained and merged with those from the initial step.

(2) Duplicate reads are removed with BBMap version 33.4.

(3) Reads are mapped to GENCODE version 19 reference with bowtie2 using '--un' option and those that do not map to known transcripts are selected.

(4) The transcript sequences are aligned to the transcriptome with BLAT.

(5) Reads that match multiple reference transcripts are selected, where the two reference transcripts are separated by more than 1 kilobase (kb) in the genome. Further, false chimeras are controlled by selecting only those fusion candidates where the chimeric transcripts have similar sequences of 13 bases or less in common between the reference genes.

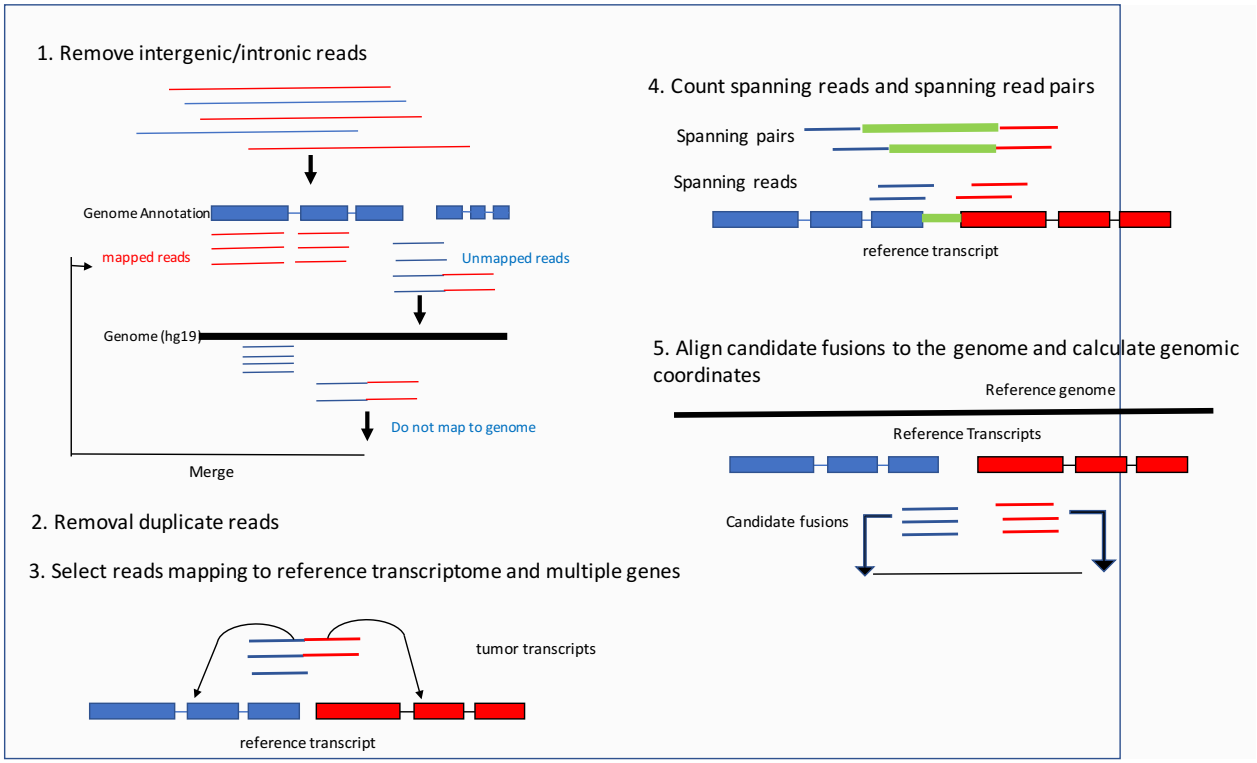
(6) The number of spanning reads and spanning pairs across the breakpoint are counted. Spanning reads are defined as the reads that lie across the breakpoint and the spanning pairs are defined as pairs in which the reads of each read-pair (paired-end sequencing reads) lie in their entirety, one on either side of the breakpoint.

(7) The candidate fusion transcript sequences are aligned to the human reference genome using BLAT.

(8) The genomic coordinates for each breakpoint are identified and the genomic gap size is calculated. Candidates with less than 10Kb gap size are discarded as they are likely to be false positives. In a few scenarios, genuine fusions with small gaps are distinguished from read-through transcription or unannotated splicing by looking for evidence of genomic rearrangement (such as inversion), based on direction of the fusion transcript with respect to the genome. Next, JAFFA checks if breakpoints lie on known exon-exon boundaries. This scenario would arise if the fusion occurred within intronic DNA and the exon structure was preserved and if this were true, fusions are checked to see if they are in-frame based on the most common gene isoform. Finally, based on spanning reads, spanning pairs, transcriptional

breakpoints aligned with exon boundaries and genomic gaps, JAFFA classifies the candidates as low, medium and high confident events and ranks them.

Figure 2.4: Schematic representation of the JAFFA pipeline



JAFFA pipeline was used for the analysis of RNA-seq data for gene fusion discovery.

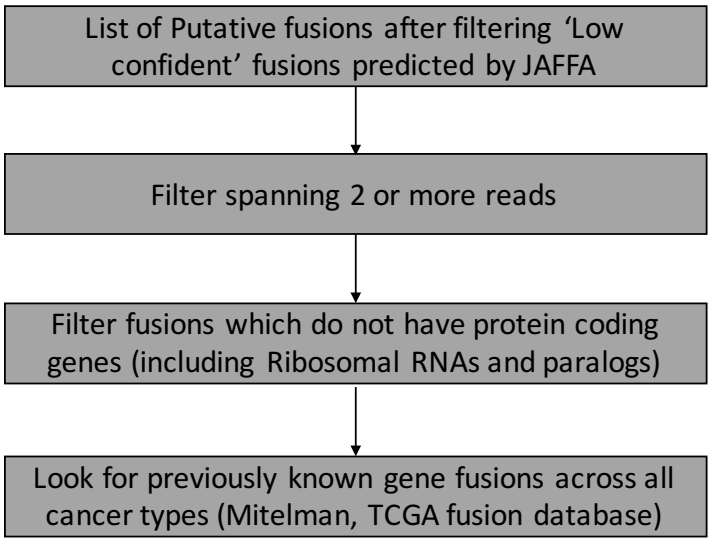
2.3.3 Downstream processing and gene fusion prioritization

2.3.3.1 Filtering non-significant gene fusions

Artifacts in gene fusion discovery are common despite the use of state of the art tools [331], thus post-processing filtering steps are necessary to mitigate the specific biases at the level of sensitivity and specificity. The high and the medium confident fusions categorized by JAFFA were selected for further downstream processing to refine the list and prioritize candidates of significance. This involved the following filtering steps **(Figure 2.5)**: (1) Filtering fusion candidates with less than 2 spanning reads aligned on the fused boundaries; (2) Filtering fusion candidates where fusion partner genes are known paralogs or same gene family as these are likely to be artifacts due to mapping errors; (3) Annotating genes with Ensemble human genes (GRCh37.p13) and their fusion candidates as un-annotated gene partners. For

example, pseudogenes and non-protein coding genes (ribosomal RNA, linc-RNAs) and highly polymorphic genes such as HLA are filtered since they produce reads that cannot be easily mapped to unique locations in the genome; (4) Removal of fusions that are present in the normal non-cancerous tissue based on Babiceanu et al's [332] list with the exception of genes that are previously implicated as germline cancer genes. All the fusions are examined in databases of known gene fusions across all cancer types to identify known or established gene fusions.

Figure 2.5: Steps for gene fusion prioritization



2.3.3.2 Classification of fusion events with Oncofuse

After filtering, hundreds of fusions events were identified, raising the question of how to differentiate the fusion events important for cancer development or 'driver event' and not just passenger events. A Bayesian classifier tool called Oncofuse [333] was used to extract various features from the candidate input list and returns a Bayesian probability value that such a fusion gene belongs to a class of passenger event or a driver event. The filtering steps were the following: First, all the candidate fusions were annotated and classified as driver or passenger events. This tool uses a set of features, including swapping of promoter and untranslated regions and a combination of specific protein domains and protein interaction interfaces (PII) to predict the oncogenic potential of the putative gene fusion. Oncofuse uses a Naïve Bayes network Classifier which is a model that assigns class labels to problem cases, represented as vectors of feature values. Here, 24 classification features were used for

classification of 'driver' versus' passenger events. These features were based on the following categories: (1) promoter features of the fusion protein gene (FPG), such as retention of the 5' FPG and loss of the 3' FPG; (2) six functional profile features (3) 3' UTR features; (4) PPI. To trained and test the model a supervised Discretization algorithm in Weka machine learning package was used on gene sets obtained from 10 different databases, including TICdb3.0, RTH, Mitelman, Oncogene (Sanger), RefSeq, NGS studies and ChimerDB2 [333].

To obtain the functional profile of a novel gene fusion the following steps were conducted: (1) A list of protein domains was extracted for a given fusion sequence and then the genes containing those domains were selected by querying a set of all human genes; (2) The gene ontology (GO) terms were extracted for these genes; (3) A functional family association score (FFAS) was calculated by counting the number of times a GO term overlaps with a set of six predefined functional families, including transcription factor, kinase activity, transcription cofactor, GTPase, helicase/histone modifiers and protein binding. These scores for each of the six functional profile features were then used for variant prioritization.

2.3.3.3 Identifying polygamous and private fusion events

After filtering and classifying fusions as driver events, the candidate fusions were further categorized into polygamous and private fusions based on the frequency of occurrence. If either the 5' or the 3' gene partner were detected in more than one sample, the fusions were categorized as polygamous fusions, whereas the those present in only single sample were categorized as private gene fusions. Additionally, the private fusion events were selected if gene fusions had a high driver probability (>60%) based on classification by Oncofuse. All the candidates were checked for previously reported fusions in other cancers, in known fusion database such the TCGA gene fusion portal and the 5' and 3' fusion gene partners are compared against the Mitelman Database of Chromosome Aberrations [176] to estimate the recurrence of known or established gene fusions.

2.3.3.4 Statistical tests and plots

Statistical tests were performed to test the association of gene fusions with tumour subgroups with respect to gene expression profile, known driver mutations and copy number alterations. The Mann-Whitney U test was performed in R (version 3.4) to compare the differences between the binary subgroups based on mutation, copy number and expression status. The circus plot of all curated fusion events was then plotted in R (version 3.4) with the 'chimeraviz' package (version 1.0.2).

Chapter 3. Profiling copy number abnormalities in uveal melanoma

3.1. Copy number analysis results

3.1.1 Broad copy number alterations in uveal melanomas

Previous genomic profiling studies in UM have been done at low resolution. In the current study, genomic DNA samples from 182 primary enucleated UMs (87 with matched normal DNA) were profiled by high-resolution arrays from three different platforms. These interrogated >700K probes with a median interprobe spacing ranging from ~24 to ~18Kb. Data was compiled from three different sources: high quality Affymetrix SNP6.0 mapping array data from The Cancer Genome Atlas (TCGA, 80 cases), Illumina660WQuad array data from Cleveland Clinic (CC, 57 cases) and Illumina Omniarray and CytoSNP850K from samples recruited at Washington University in St Louis (WASH-U, 45 cases). Tumor specific alterations were identified by normalizing the signals with normal reference samples to exclude potential germline copy number variation. Copy number alteration (CNA) detection was performed on the ASCAT PCF segmented CC and WASH-U cohorts and circular binary segmented TCGA cohort. CNAs affecting more than 50% of a chromosomal arm were classified as broad events and were inferred from segmented data. A genome-wide comparison of CNAs indicated the presence of a high-degree of non-random CNAs across samples from all the cohorts.

Overall, 1094 broad arm level aberrations were identified (6 on average per sample; range 0-39). Twelve broad events affected >10% of all samples including five broad arm level amplifications that included 1q gain (1q+), 6p gain (6p+), 8p gain (8p+), gain (8q+), 21q gain (21q+) and seven broad arm level deletions that include 1p loss (1p-), 3p loss (3p-), 3q loss (3q-), 6q loss (6q-), 8p loss (8p-), 9p loss (9p-) and 16q loss (16q-) (frequencies are provided in **Table 3.1, Figure 3.1**). The most frequent CNAs were amplification of chromosome 8q and monosomy 3 which were detected in more than 50% of all samples consistent with previous studies. Gains of 6p and loss of 6q occurred in 23% (41/182) of tumours indicative of isochromosome 6p formation. Similarly gains of 8q and loss of 8p were observed in 19% (35/182) of tumours forming an isochromosome 8q.

Unsupervised hierarchical clustering of all UM samples (N=182) was performed and the number of clusters were chosen by visualizing the clustered heatmap with four distinct clusters. The four main groups include **(Figure 3.2**): Group 1: 1p-/1q+, monosomy 3, 6p+/6q-, 8p+/8q- and 16q-; Group2: 6p+/6q-, 8q+; Group 3: 6p+ and Group 4: monosomy 3 and 8p-/8q+ (similar to group 1 with less 1p- and 6p+/6q-). In addition, clusters were tested to find correlations with metastatic outcome. Group 1 was not significantly associated with metastatic outcome (p-value 0.0728), and Groups 2 (p-value 0.0137) and 3 (p-value 0.0021) were significantly associated with non-metastatic outcomes. Group 4 was associated with associated with metastatic outcome (p-value 0.0085).

When examined independently, chromosomal alterations that were associated with metastatic disease were loss of 3p (P value = 2.29e-08), loss of 3q (P Value = 2.95e-07), loss of 6q (P Value = 0.0246), gain of 8q (P Value = 0.0001) and loss of 8p (P Value = 6.29e-05). Other alterations which were not directly associated with metastasis but were previously shown to be associated with PRAME positive tumours [46] which were not detected included, 1q gain (P Value = 0.3589), 6p gain (P Value = 0.0538) and 16q loss (P Value = 0.065). Whole chromosome arm-level CN-LOH affecting more than 50% of arm length was observed on chromosome 3 (9/76 samples, 11%), chromosome 6 (5/103, ~5%), 9p (2/103, ~2%), 15q (2/103, ~2%) and 1 sample on chromosome 8, chromosome 10 and 9p.

3.1.2 Ploidy and aberrant tumour cells in uveal melanoma

The ploidy estimates predicted by ASCAT were examined for 103 tumours samples (from WASH-U and CC cohorts) genotyped on the Illumina platform to investigate the relevance of aneuploidy in UM. The tumours showed an average ploidy of 2.4n. The ploidy estimates were then correlated with gene expression which predicts UM subtypes. Evaluating the ploidy distribution by stratifying the samples based on the GEP classifier information available for 45 tumours samples from the WASH-U cohort showed a common state of diploidy for majority of both, class 1 and class 2 tumours. However, a minority of tumours had undergone polyploidization by endoreduplication resulting in tetraploid state. This was more so in class 2 tumours compared to class 1 although this difference was not statistically significant (P Value=0.60) **(Figure 3.3 A)**. In addition to ploidy estimates, the percentage of aberrant cells estimated by ASCAT analysis showed a mean of 86% tumour content per samples which was similar to the pathological review. Stratifying the aberrant cell fraction

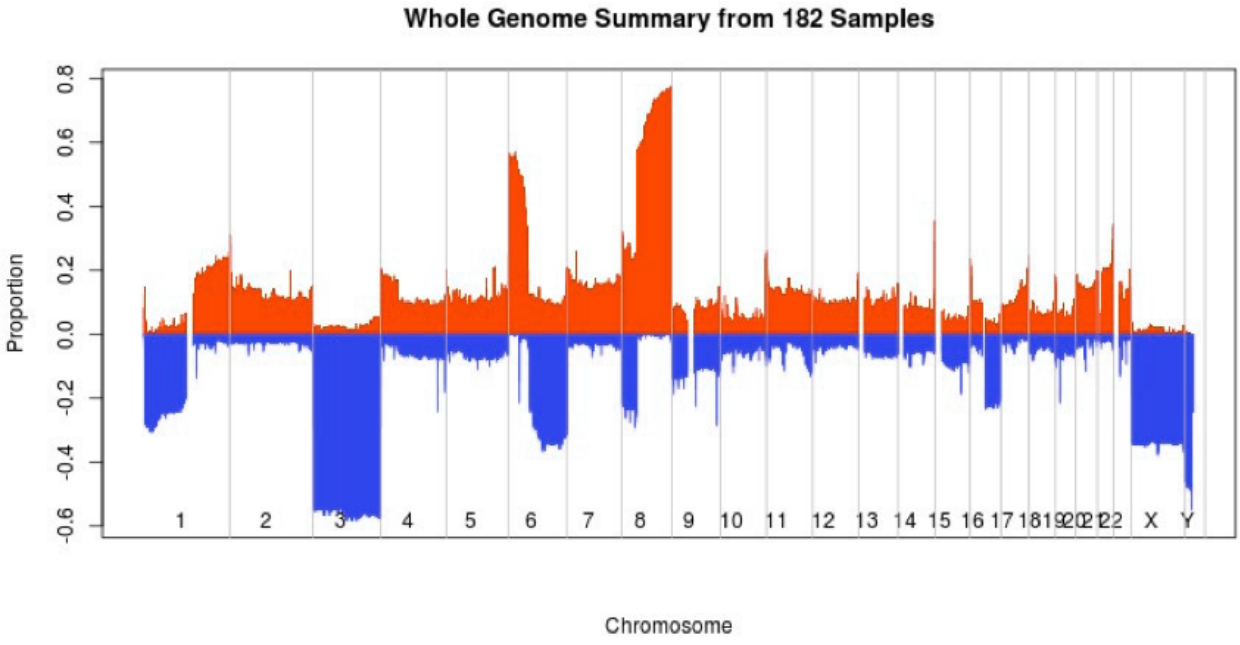
estimates by class 1 versus 2 tumour subgroups revealed considerable differences in the class 1 tumour subgroup compared to class 2 **(Figure 3.3 B)**.

Table 3.1: Recurrent broad chromosomal arm level amplifications and deletions.

CN Event	Chromosomal arm affected	Number of samples with alterations (N=182)	
		# Event	%
Loss	3q	98	54%
Loss	3p	94	52%
Loss	6q	57	31%
Loss	1p	38	21%
Loss	8p	36	20%
Loss	16q	34	19%
Loss	9p	21	12%
Gain	8q	125	69%
Gain	6p	83	46%
Gain	8p	37	20%
Gain	1q	26	14%
Gain	21q	26	14%

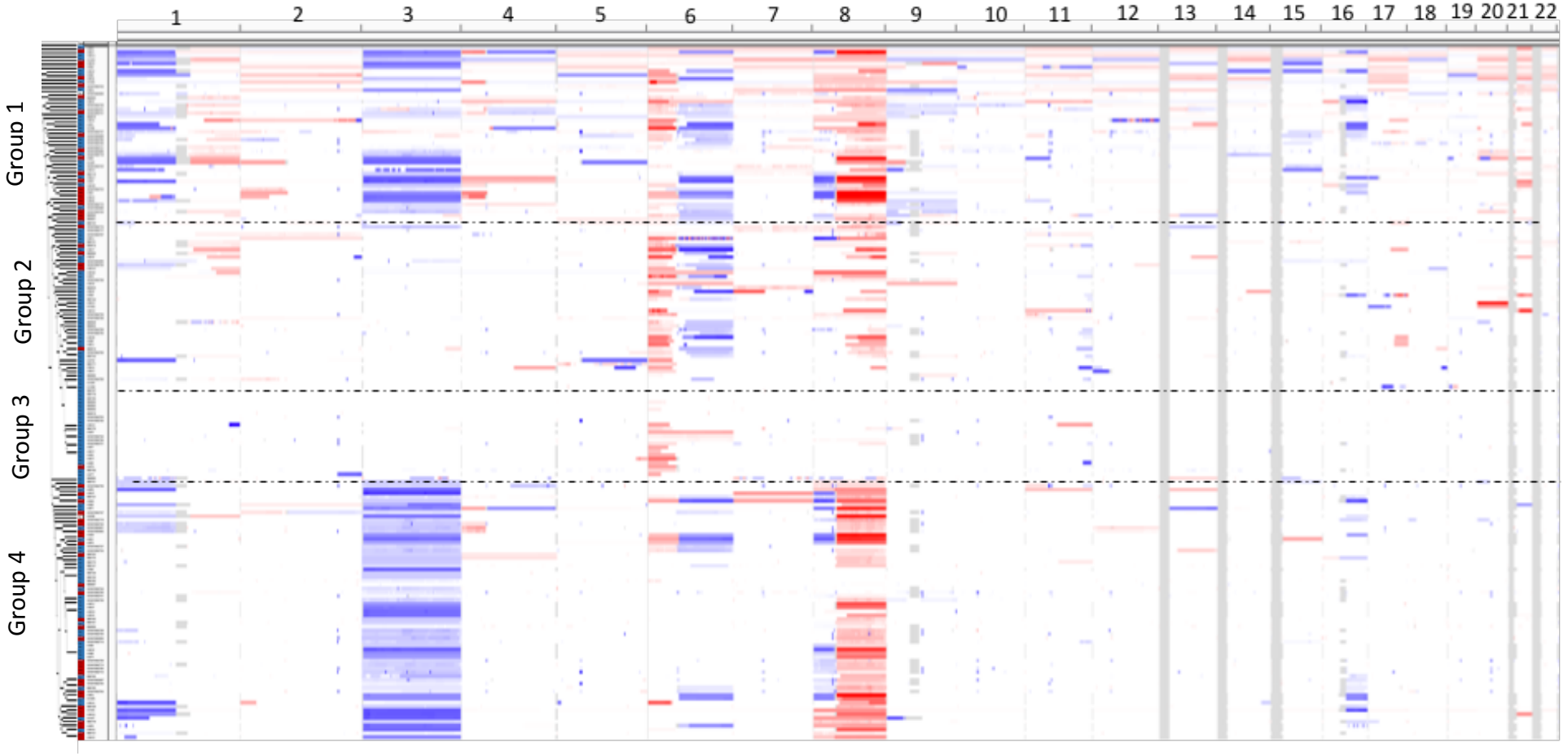
The alterations are present in >10% of UM primary tumour samples.

Figure 3.1: Genomic profile of the most common gains and losses in UMs.



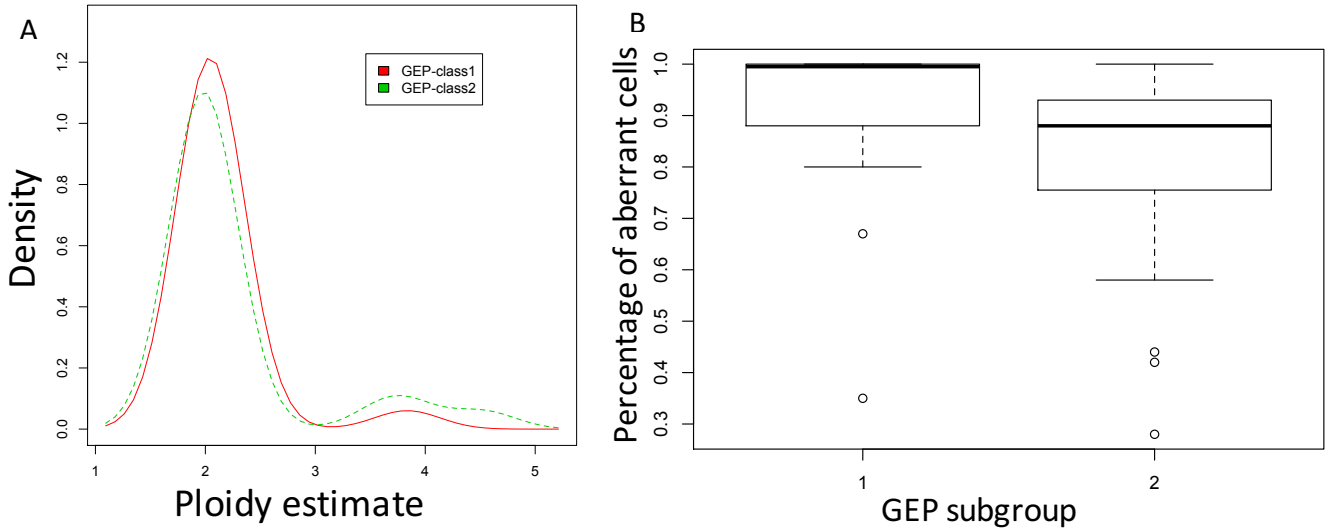
Gains are denoted by red and losses denoted by blue for all the samples (N=182). X-axis represents the chromosomes and Y-axis represents the proportion of gains and losses in the combined cohort.

Figure 3.2: Heatmap of copy number profiles from 182 UM tumours.



Unsupervised hierarchical clustering of all tumours revealed that gains and losses clustered together within 4 groups. Metastasis status represented by color labels on the left side; red indicating the metastasis group.

Figure 3.3: (A) Distribution of ploidy estimates from ASCAT. (B) Aberrant cell fraction estimates from ASCAT



Box and density plots were plotted for 45 uveal melanomas across class 1 (N=22) and class 2 (N=23) GEP based subgroups. Most of the tumours show a ploidy close to 2n and a smaller fraction showed a ploidy close to 4n. The class 2 subgroup displays the higher level of ploidy and a broader range (class 2: mean=2.24, range 1.78-4.75; class 1: mean=2.13, range=1.96-3.83). (B) Box plot showing the percentage of aberrant cell fraction across the class 1 and class 2 subgroups. The class 1 group has the highest level of aberrant tumour cell fraction compared to class 2 (P value 0.0027, 2-tailed Mann-Whitney U test). Most the data in the class 1 group have a median close to 1 (0.995) as seen from the plot where the median is close to the upper quartile with no whiskers, indicating tumors with 100% tumor content.

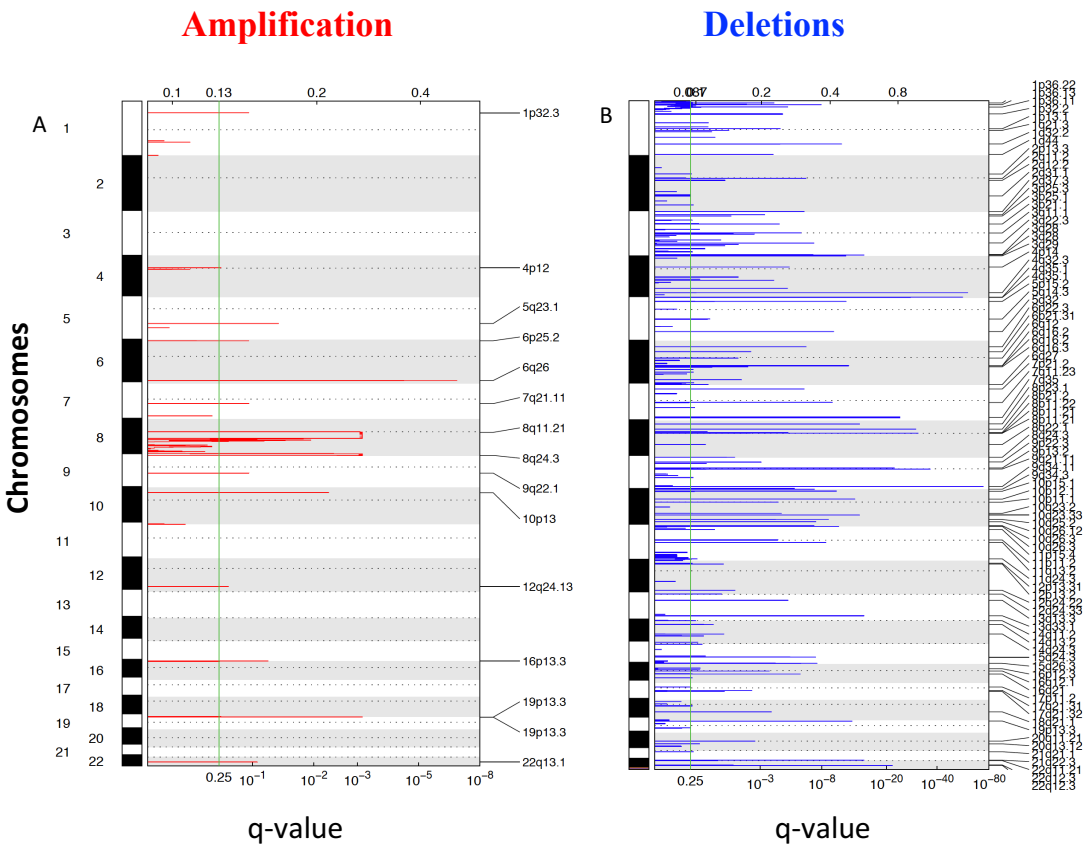
3.1.3 Focal copy number alterations identified with GISTIC

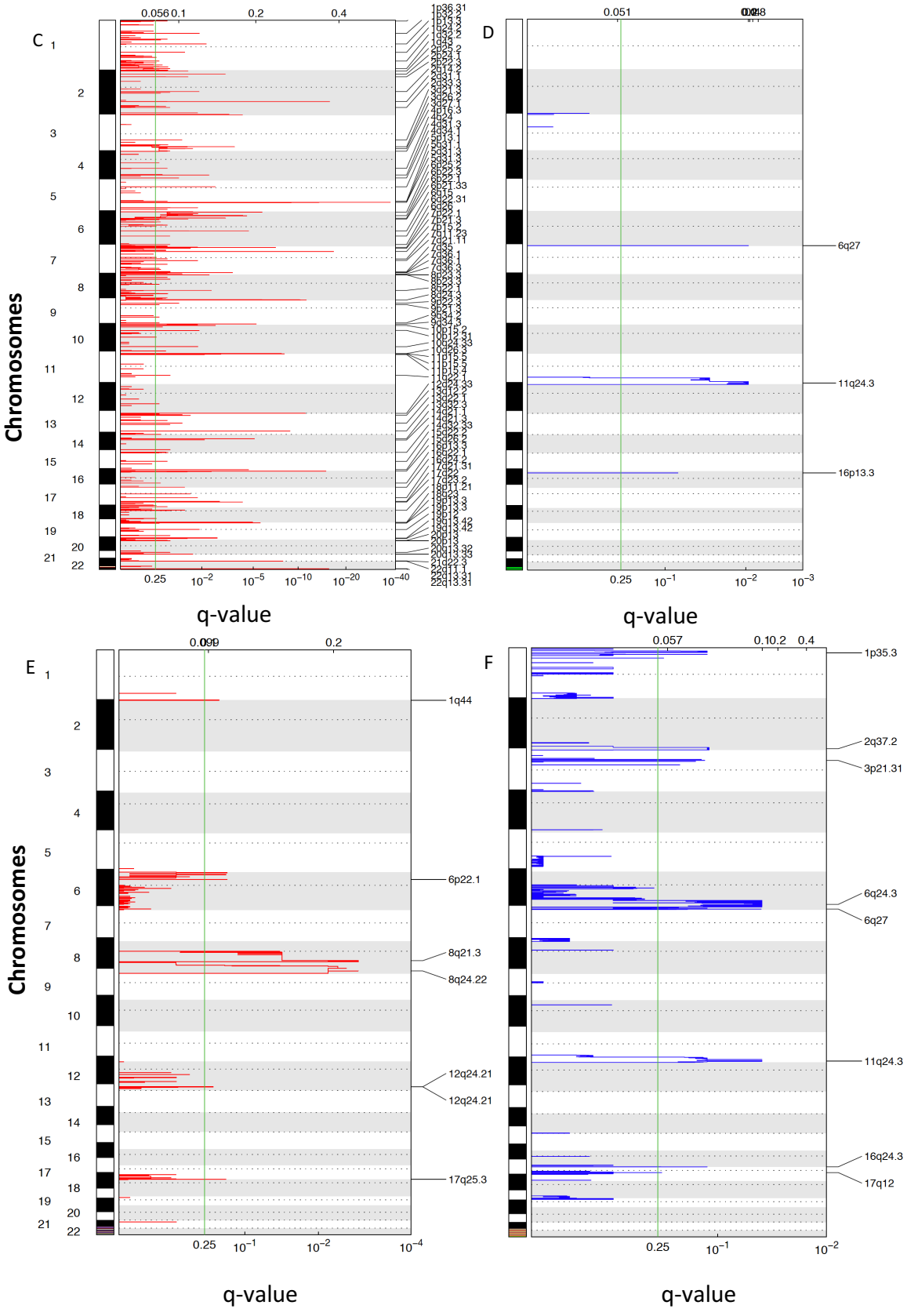
The large size of chromosomal arm-level alterations makes it difficult to identify specific target gene or genes of interest. However, the identification of focal CNAs (FCNA) can target genes relevant to tumour biology [334–337]. Copy number profiles of the 182 UM primary cancer specimens were analysed to identified FCNAs that occur at a significantly higher frequency compared to the average background rate in the dataset with GISTIC (genomic identification of significant target in cancer) algorithm [301]. For each sample, the probe intensity (log R ratio) and allelic intensity (B allele frequency) data were processed and segmented with the ASCAT-PCF segmentation method to obtain a copy number profile. To distinguish biologically significant FCNAs from random events in UMs, GISTIC was run

through the segmented data and processed downstream to integrate results from each cohort. Although GISTIC was applied to all the 182 samples to identify focal and broad peaks as described in the methods, it cannot readily integrate samples from different platforms. Thus, the frequency of alterations for each cytoband (GISTIC peak from each independent cohort analysis) was determined across all samples using platform independent segmented copy number profiles. The term focal CNAs (or FCNA) used here refers to a copy number segment smaller than half the length of the chromosomal arm.

In total 203 independent regions of significantly recurrent somatic FCNAs were identified. These included 103 amplifications and 100 deletions across 182 samples from pooled analysis on all cohorts without any filtering. The significant GISTIC peaks were identified across all the cohorts using a q value threshold of 0.25 **(Figure 3.4)**.

Figure 3.4: Significant GISTIC amplifications and deletion peaks identified.

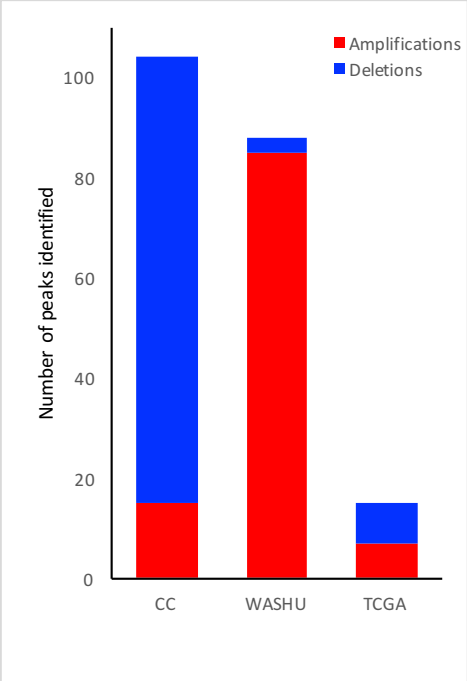




Here, CN gain peaks represented in red and CN-loss peaks represented in blue. Green lines denote the threshold for significance FDR value (q-value) < 0.25. (A-B) CC cohort. (C-D) WASHU cohort (E-F) TCGA cohort.

The peaks identified from the TCGA data (N=80) were then compared to those identified in WASH-U (N=45) and CC (N=57) cohorts **(Figure 3.5)**. The number of peaks in the TCGA cohort were substantially higher compared to the WASHU and CC cohort. Although, the significance q-value threshold of 0.25 was initially selected based on the literature, the analysis was re-run using a lower q-value threshold of 0.05 to account for potential false positives however this difference persisted **(Appendix 2)**. This difference could be attributed to the variation in the platforms used to generate data across all three cohorts in addition to the private germline CNV that could not be filtered out due to the unavailability of matched normal samples for some samples in the WASHU and CC cohorts. The TCGA data had matched normal for all the tumour samples and thus germline CNVs were filtered out for all the samples. Thus, in the case of the TCGA data only truly somatic segmented profiles were used for GISTIC peaks calling.

Figure 3.5: Number of GISTIC amplifications and deletions peaks identified across all three cohorts



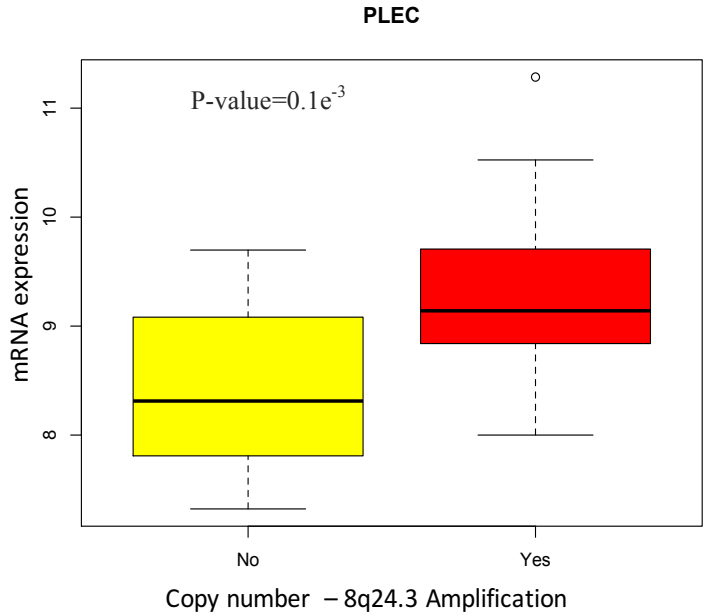
The germline CNAs were filtered by adopting a compiled list of genomic variants from the Database of Genomic Variants (DGV) and HapMap normals. On average, there were 34 focal alterations per sample after germline variants were removed. The most frequent of these significant focal alterations were amplification of *PLEC1* on chromosome 8q24.3 (74% of samples) and deletion of *PCCB* on 3q22.3 (55% of samples). The mRNA expression

of *PLEC1* and *PCCB* (normalized RNA-seq read counts) were analyzed between the samples with and without copy number change. Samples with focal amplification of 8q24.3 showed significant change in *PLEC1* expression (p-value=$0.1e^{-3}$) compared to samples with no amplification (FC=1.84) **(Figure 3.5)**. However, no significant difference was observed for *PCCB* expression between the samples with deletion of 3q22.3 and wildtype samples (FC 0.95, p-value =0.9). The low or partial correlation of gene expression with *PCCB* deletion is not clear but could be due to the CNV identification method where a lower threshold was used for deletions. Genes that showed significant change in expression that correlated with copy number alterations were selected as putative candidates for follow up, since they were more likely to correspond to biologically meaningful events.

The least frequent events were observed in 2% of samples. Among these 203 regions, the CNA boundaries for amplification events revealed a median size of 33.3 Kb (1.5-1610 Kb) for amplifications and a median size of 90.2Kb (9.5-21104 Kb) for deletions. To prioritize the most relevant alterations with respect to uveal melanoma, seventy FCNAs (from a total of 207) which occurred in more than 15% of all samples were selected. This also eliminated lower frequency spurious alterations and ensured that the alterations were robust to changes in the number of samples. These FCNAs include 29 recurrently amplified and 41 recurrently deleted regions **(Table 3.2)**. For each of the 70 significant FCNAs, a "peak" region was identified by GISTIC that is most likely to contain genes or loci involved in tumourigenesis. These peaks FCNA peaks lie within a 95% confidence window. The 29 FCNA peaks contained a median of 3 genes each (range 0-81, including microRNAs and other non-coding RNAs); 2 regions contained more than 25 genes each and the remaining 27 regions encompassed in aggregate 123 potential target genes. The 41 focal deletion peaks contained a median of 4 genes each (range 1-505, including microRNAs and other non-coding RNAs). Six regions contain at least 25 genes and the remaining 35 regions contain in aggregate 145 potential target genes.

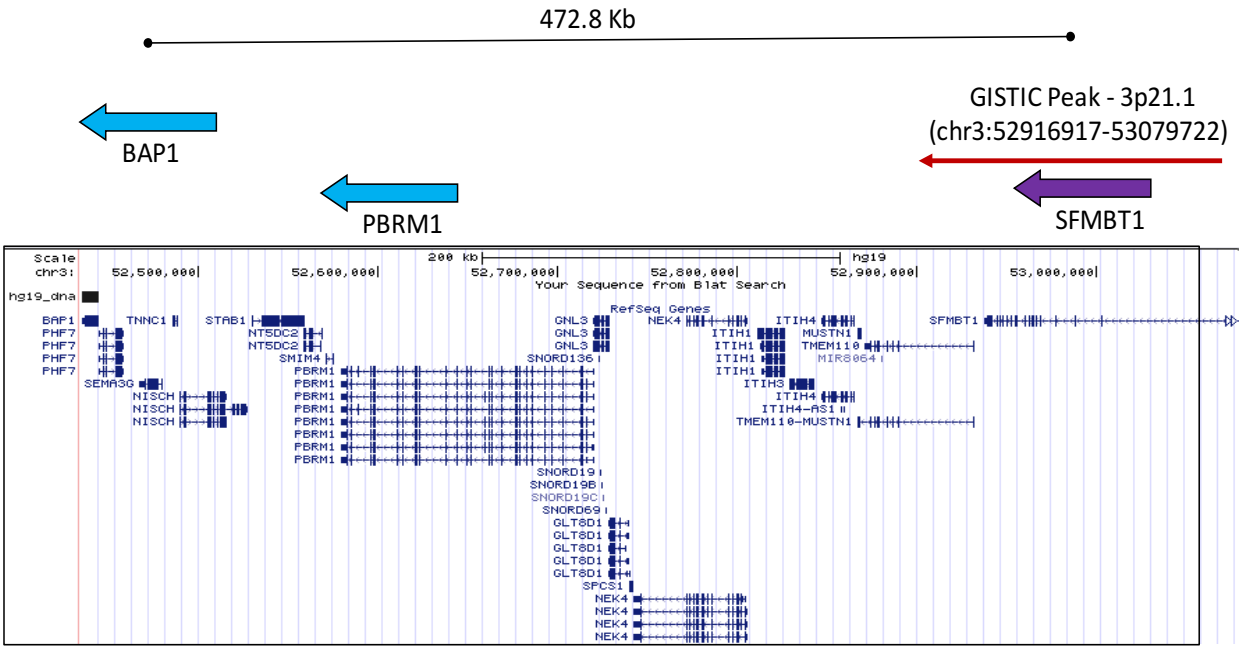
GISTIC tended to identify very small regions with the high-resolution platform data used in this study and as a result some of the expected genes were missed. For example, on 3p21.1 it identified a significant peak spanning 162.8 Kb and containing the genes *TMEM110*, *TMEM110-MUSTN1* and *SFMBT1*. This peak did not intersect with the *BAP1* gene, an established tumor suppressor gene in UM [115] which is located 472.8 Kb from the reported peak in the same chromosomal band (3p21.1) **(Figure 3.6)**.

Figure 3.6: Box and whisker plot of PLEC expression.



Comparison of PLEC expression between in tumors with and without focal amplification of 8q24.4 locus compared using samples in TCGA cohort (N=80). X-axis: Copy number gain and wildtype; Y-axis: Normalized read count from RNA-seq data.

Figure 3.7: GISTIC deletion peak on 3p21.1.



The peak is close, but does not overlap with the locus containing BAP1 gene.

Another event that was examined was amplification of *MYC* gene that lies on 8q24.21. *MYC* is amplified in many cancers, and its role in uveal melanoma is of interest, given the common occurrence of 8q amplification in this tumour type. However, the closest

focal peak identified by GISTIC at 8q in the current analysis was at 8q24.22, which lies more than 5 megabases from the *MYC* locus. This peak contained *PLEC1* (indicated above), which showed significant correlation between gene expression and copy number change.

To increase confidence in identifying candidate genes, all the GISTIC peaks were combined into a single dataset and those peaks that had overlapping boundaries across any two of the three cohorts were selected as high confident peaks. These revealed only four overlapping deletion peaks mapping to chromosomes 1p35, 2q37.2, 6q27 and 11q24.3. (Table 3). Candidate genes within or near these deletion peaks are also shown and correspond to *PDLIM1* at chromosome 1p35.2, LOC200772 at chromosome 2q37.2, *WDR27*, *PHF10* and c6orf120 at chromosome 6127 and *FLIT1* on chromosome 11q24.3. Although there were no overlapping amplification peaks, there were six peaks where the boundaries between dataset-pairs were within 1Mb of each other and these included chromosomes 6p25.2, 6p22.1, 6q26, 8q24.3, 16p13.3 and 19p13.3 **(Table 3.3).**

Table 3.2.: GISTIC amplification and deletion peaks.

Copy Number alteration	Band	GISTIC peak limits	Frequency across all samples (N=182)	TCGA (N=80)	WASH-U (N=45)	CC (N=57)
Amplification-01	1q32.2	chr1:210068118-210122321	18%	13	9	10
Amplification-02	1q43	chr1:240845575-240893966	16%	13	7	10
Amplification-03	1q44	chr1:247640512-249250621	16%	14	5	10
Amplification-04	2p25.2	chr2:5828703-5847674	15%	13	7	7
Amplification-05	2q31.1	chr2:176974105-177008913	16%	9	19	1
Amplification-06	4p16.3	chr4:1190519-1535003	15%	11	8	8
Amplification-07	6p21.33	chr6:30416832-30468162	48%	44	23	20
Amplification-08	6p22.1-peak1	chr6:28967039-29009567	48%	44	22	21
Amplification-09	6p22.1-peak2	chr6:28229882-28258984	45%	44	17	21
Amplification-10	6p22.3	chr6:19827920-19855932	51%	46	24	23
Amplification-11	6p25.2-peak1	chr6:3144029-3152404	52%	46	21	27
Amplification-12	6p25.2-peak2	chr6:2886796-2890282	50%	46	23	22
Amplification-13	7p15.2	chr7:27205283-27281156	19%	10	21	4
Amplification-14	7p22.1	chr7:6656831-6694717	15%	10	14	4
Amplification-15	7q21.11	chr7:85486010-85511551	16%	9	2	18
Amplification-16	8p23.3-peak1	chr8:1784860-1811396	25%	20	13	13
Amplification-17	8p23.3-peak2	chr8:1921161-1954463	25%	20	17	9
Amplification-18	8q11.21	chr8:52020710-52031773	56%	47	20	35
Amplification-19	8q21.3	chr8:89811299-91200675	68%	58	26	39
Amplification-20	8q22.1	chr8:94700402-94749564	70%	60	29	38
Amplification-21	8q24.22	chr8:133991650-134011691	72%	61	27	43
Amplification-22	8q24.3-peak1	chr8:144229104-144482855	74%	61	30	44
Amplification-23	8q24.3-peak2	chr8:144956465-145044748	74%	61	31	42
Amplification-24	11p15.5-peak1	chr11:549120-567390	16%	9	16	4
Amplification-25	11p15.5-peak2	chr11:824043-831882	15%	9	13	6
Amplification-26	16p13.3	chr16:1-889655	19%	4	15	16
Amplification-27	17q25.3	chr17:79976479-80215585	15%	15	10	2
Amplification-28	21q22.3	chr21:47416325-47465204	20%	16	14	6
Amplification-29	22q13.31	chr22:46363740-46480601	15%	6	18	4
Deletion-01	1p13.1	chr1:116174525-116243379	20%	19	4	13
Deletion-02	1p32.2	chr1:57450404-58743937	26%	23	3	21
Deletion-03	1p35.3 (overlap 1	chr1:19660871-36108704	34%	32	6	23
Deletion-41	1p36.11	chr1:26412170-26460781	25%	32	6	23
Deletion-04	1p36.13	chr1:17629005-17694614	27%	23	6	20
Deletion-05	1p36.22	chr1:10442868-10512055	25%	23	4	18
Deletion-06	2q37.2	chr2:231727807-242218174	22%	16	2	22
Deletion-07	3p21.1	chr3:52916917-53079722	52%	44	19	31
Deletion-08	3p21.31	chr3:46959694-47026134	51%	44	19	30
Deletion-09	3p25.1	chr3:15693242-15845263	51%	42	19	31
Deletion-10	3p25.3	chr3:11583255-11767552	51%	42	18	33
Deletion-11	3q11.1	chr3:90476372-93759078	54%	44	19	35
Deletion-12	3q22.3	chr3:135956306-136055458	55%	45	20	36
Deletion-13	3q28-peak1	chr3:191040175-191128440	55%	46	19	35
Deletion-14	3q28-peak2	chr3:189670949-189844960	53%	46	19	32
Deletion-15	3q29	chr3:194400231-194418125	52%	46	18	31
Deletion-16	4q32.3	chr4:165991449-166038190	27%	5	0	44
Deletion-17	4q35.1	chr4:186386588-186459025	23%	5	0	36
Deletion-18	6q12	chr6:64353648-66424344	30%	18	9	28
Deletion-19	6q16.2-peak1	chr6:99982340-100034579	37%	22	16	29
Deletion-20	6q16.2-peak2	chr6:100046934-100089847	35%	21	16	27
Deletion-21	6q16.3	chr6:105154522-105319124	34%	22	16	23
Deletion-22	6q24.3	chr6:137517511-158622124	39%	29	15	27
Deletion-23	6q27-peak1	chr6:170073603-170152990	30%	23	13	19
Deletion-24	6q27-peak2	chr6:166821252-167278270	30%	23	13	19
Deletion-25	7q35	chr7:145807448-148139556	19%	1	0	34
Deletion-26	8p11.21-peak1	chr8:42100553-42191342	30%	20	5	29
Deletion-27	8p11.21-peak2	chr8:42193876-42259298	28%	20	5	26
Deletion-28	8p11.22	chr8:39192482-39405336	27%	20	7	23
Deletion-29	8p21.2	chr8:25037493-25281328	35%	20	5	38
Deletion-30	8p23.1	chr8:11223023-11307117	25%	20	5	21
Deletion-31	9p13.2	chr9:36314567-36401351	18%	8	2	23
Deletion-32	9q21.11	chr9:71719619-71882817	22%	6	0	34
Deletion-33	9q34.11	chr9:131375702-131425763	29%	6	0	46
Deletion-34	11q24.3	chr11:123744052-135006516	16%	11	5	13
Deletion-35	15q24.3	chr15:77283978-77377653	21%	5	1	32
Deletion-36	16p12.3	chr16:12988185-25246530	16%	2	0	28
Deletion-37	16q12.1	chr16:48390513-48419402	18%	16	6	11
Deletion-38	16q21	chr16:58564865-58668843	21%	16	7	16
Deletion-39	16q24.3	chr16:89304312-90354753	24%	16	5	23
Deletion-40	22q11.21	chr22:18038787-18079517	23%	1	0	40

GISTIC identified peak regions of amplifications (29 peaks) and deletions (41 peaks) present in more than 15% of samples across all three cohorts.

Table 3.3: Overlapping GISTIC focal peaks.

Event	GISTIC peak cytoband	Cohorts with Overlaps			Overlap level	Minimum common Region (MCR)	Candidate Genes in the MCR
		WASHU	TCGA	CC			
Amplication	6p25.2				Boundaries <1Mb apart		
Amplication	6p22.1				Boundaries <1Mb apart		
Amplication	6q26				Boundaries <1Mb apart		
Amplication	8q24.3				Boundaries <1Mb apart		
Amplication	16p13.3				Boundaries <1Mb apart		
Amplication	19p13.3				Boundaries <1Mb apart		
Deletion	1p35.3				Overlap	chr1:26412170-26460781	PDIK1L
Deletion	2q37.2				Overlap	chr2:241868075-241913552	LOC200772 (ncRNA)
Deletion	6q27				Overlap	chr6:170099399-170123579	WDR27, PHF10, C6ORF120
Deletion	11q24.3				Overlap	chr11:128551690-128575131	FLI1

The GISTIC focal peaks for overlapping regions for three different independent analysis performed on data from 3 different cohorts (WASH-U, TCGA and CC). All peaks boundaries were compared to each other and peaks with complete overlaps or within 1 MB are reported here.

3.1.4 Comparison of Focal CNAs in UM and Pan cancer findings

The Focal CNAs from our GISTIC analysis were compared with those of 3131 tumors spanning nearly 54 tumors types (Beroukhim 2010) [338], in order to distinguish between CNAs that are unique to UMs and those that are found in other tumours . Focal peak boundaries from the current analysis of UM were matched with 158 amplification and deletion peaks reported in the pan cancer copy number study, of which 6/29 (20%) were regions of copy number gain and 12/40 (30%) were regions of copy number loss.

The identification of regions that were common to both the sets was done by annotating UM focal peak boundaries as the reference and categorizing matches as “sub” for UM peaks within and “super” for those spanning those from the Pan cancer study. Additional regions of copy gain and loss exhibited partial overlap and are indicated as “left” or “right” depending on the overlapping boundaries. Shared alterations **(Appendix 4 and Table 3.4)** include gains on chromosomal arms 1q, 6p and 8q and losses on chromosomal arms 1p, 4q, 6q, 7q, 8p, 9p, 11q and 16q suggesting a broader role of these alterations in multiple tumor types. The shared amplification peaks include 1q44, 6p22.3, 6p25.2 (peaks 1 & 2) and 8q24.3 (peaks 1 & 2) and deletion peaks include 1p13.2, 1p36.11, 4q35.1, 6q16.2 (peaks 1 & 2), 6q16.3, 7q35, 8p11.2, 8p21.2, 9p13.2, 11q24.3 and 16q24.3. In contrast 51

UM specific focal CNAs were observed and include 23/29 (79%) copy gains and 28/40 (70%) copy losses **(Table 3.4)**. The median size of shared alterations was 14.7 Mb (range 5.92-23.46 Mb) for amplifications and 4.37 Mb (range 0.77-28.66 Mb) for deletions.

Examining genes that lie in these shared copy number alterations across all 18 peaks and comparing them with known cancer driver genes of the complied lists of Tamborero [339] et al, Vogelstein et al [209] and Leiserson et al [340] revealed three potential cancer causing candidates. These candidate genes were AT-Rich Interaction Domain 1A (*ARID1A)* within the 1p35.3 deletion peak, Non-SMC Condensin II Complex Subunit D3 (*NCAPD3*) within the 11q25 deletion peak and Plectin 1, Intermediate Filament Binding Protein (*PLEC1)* within the 8q24.3 amplification peak. These three genes were queried against the copy number portal Tumourscape (www.broadinstitute.org/tumorscape) to find further information on cancer types with somatic copy number alteration (SCNA) that targets these genes of interest.

PLEC1 is significantly focally amplified in 4/14 independent subtypes including Acute lymphoblastic leukemia (q =0.155, within a peak), lung non-small cell cancer (NSC) ($q=5.89e^{-5}$, within a peak), melanoma (q =0.185, not in a peak) and ovarian cancer (q <=0.25, within a peak). *ARID1A* is significantly focally deleted across the entire dataset of 3131 tumours and located within focal peak region of deletion in 8/14 independent cancer subtypes including colorectal ($q=3.46e^{-5}$, within a peak), lung NSC ($q=5.83e^{-5}$, within a peak), breast ($q=9.7e^{-5}$, within a peak), renal ($q=7e^{-4}$, within a peak), ovarian (q=0.2718, within a peak), myeloproliferative disorder (q=0.0255, within a peak), hepatocellular (q=0.0663, not in a peak) and acute lymphoblastic leukemia (q=0.0775, not in a peak). The Non-SMC Condensin II Complex Subunit D3 (*NCAPD3*) gene is deleted in 5/14 independent cancer subtypes including lung NSC ($q=2.76e^{-5}$, within a peak), breast ($q=1.82e^{-4}$, within a peak), medulloblastoma (q=0.024, within a peak), esophageal squamous (q=0.055, within a peak) and melanoma (q=0.10, not in a peak). One exception to the genes in the overlapping region between UMs and other previously reported compendium of cancers (Beroukhim et al., 2010) was Axis Inhibition Protein 1 (*AXIN1*) which was significantly focally deleted in lung (q=0.018, within a peak) and hepatocellular ($q=1.05e^{-4}$, within a peak) tumours whereas it was amplified in UM samples.

Table 3.4: Comparison of copy number profiles with Pan cancer.

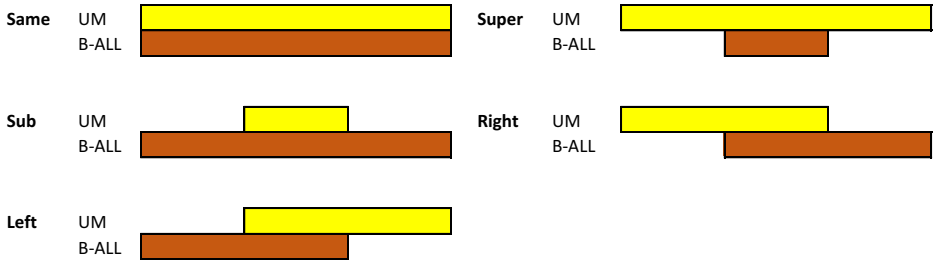
A)

UM cohort			Pan cancer - ALL		
Amplification	Band	Match	Matching Peak band	Extent of overlap	Size (Mb)
Amp-1	1q32.2	No			
Amp-2	1q43	No			
Amp-3	1q44	Yes	1q44	Left	5.94
Amp-4	2p25.2	No			
Amp-5	2q31.1	No			
Amp-6	4p16.3	No			
Amp-7	6p21.33	No			
Amp-8	6p22.1-peak1	No			
Amp-9	6p22.1-peak2	No			
Amp-10	6p22.3	Yes	6p24.1	Sub	23.46
Amp-11	6p25.2-peak1	Yes	6p24.1	Sub	23.46
Amp-12	6p25.2-peak2	Yes	6p24.1	Sub	23.46
Amp-13	7p15.2	No			
Amp-14	7p22.1	No			
Amp-15	7q21.11	No			
Amp-16	8p23.3-peak1	No			
Amp-17	8p23.3-peak2	No			
Amp-18	8q11.21	No			
Amp-19	8q21.3	No			
Amp-20	8q22.1	No			
Amp-21	8q24.22	No			
Amp-22	8q24.3-peak1	Yes	8q24.3	Sub	5.92
Amp-23	8q24.3-peak2	Yes	8q24.3	Sub	5.92
Amp-24	11p15.5-peak1	No			
Amp-25	11p15.5-peak2	No			
Amp-26	16p13.3	No			
Amp-27	17q25.3	No			
Amp-28	21q22.3	No			
Amp-29	22q13.31	No			

B)

UM cohort			Pan cancer - ALL		
Deletion	Band	Match	Matching Peak band	Extent of overlap	Size (Mb)
Deletion1	1p13.1	Yes	1p13.2	Sub	9.09
Deletion2	1p32.2	No			
Deletion3	1p35.3	Yes	1p36.11	Super	1.16
Deletion4	1p36.13	No			
Deletion5	1p36.22	No			
Deletion6	2q37.2	No			
Deletion7	3p21.1	No			
Deletion8	3p21.31	No			
Deletion9	3p25.1	No			
Deletion10	3p25.3	No			
Deletion11	3q11.1	No			
Deletion12	3q22.3	No			
Deletion13	3q28-peak1	No			
Deletion14	3q28-peak2	No			
Deletion15	3q29	No			
Deletion16	4q32.3	No			
Deletion17	4q35.1	Yes	4q35.2	Right	4.58
Deletion18	6q12	No			
Deletion19	6q16.2-peak1	Yes	6q16.1	Sub	28.66
Deletion20	6q16.2-peak2	Yes	6q16.1	Sub	28.66
Deletion21	6q16.3	Yes	6q22.1	Sub	20.58
Deletion22	6q24.3	No			
Deletion23	6q27-peak1	No			
Deletion24	6q27-peak2	No			
Deletion25	7q35	Yes	7q35	Sub	3.95
Deletion26	8p11.21-peak1	No			
Deletion27	8p11.21-peak2	No			
Deletion28	8p11.22	Yes	8p11.22	Sub	2.23
Deletion29	8p21.2	Yes	8p21.2	Sub	7.95
Deletion30	8p23.1	No			
Deletion31	9p13.2	Yes	9p13.2	Right	0.77
Deletion32	9q21.11	No			
Deletion33	9q34.11	No			
Deletion34	11q24.3	Yes	11q25	Super	4.17
Deletion35	15q24.3	No			
Deletion36	16p12.3	No			
Deletion37	16q12.1	No			
Deletion38	16q21	No			
Deletion39	16q24.3	Yes	16q23.3	Super	1.65
Deletion40	22q11.21	No			

C) **Extent of overlap reference to UVM**



Comparison of copy number alterations events shared by our Uveal Melanoma cohorts and those seen in 2433 cancers (Beroukhim et al, 2010); A) Summary of amplification peak overlaps B) Summary of deletion peak overlaps C) Figurative explanation of how overlap is described.

3.1.5 Candidate genes in the focal region of interest

Distinguishing between alterations that are highly selected during tumorigenesis (driver events) from those that do not offer any selective advantage (passenger events) has been a main challenge in the analysis of genetic alterations in cancer. Although GISTIC analysis has identified significant focal aberrations, these regions contain multiple genes and identifying the underlying target/s with limited data is difficult.

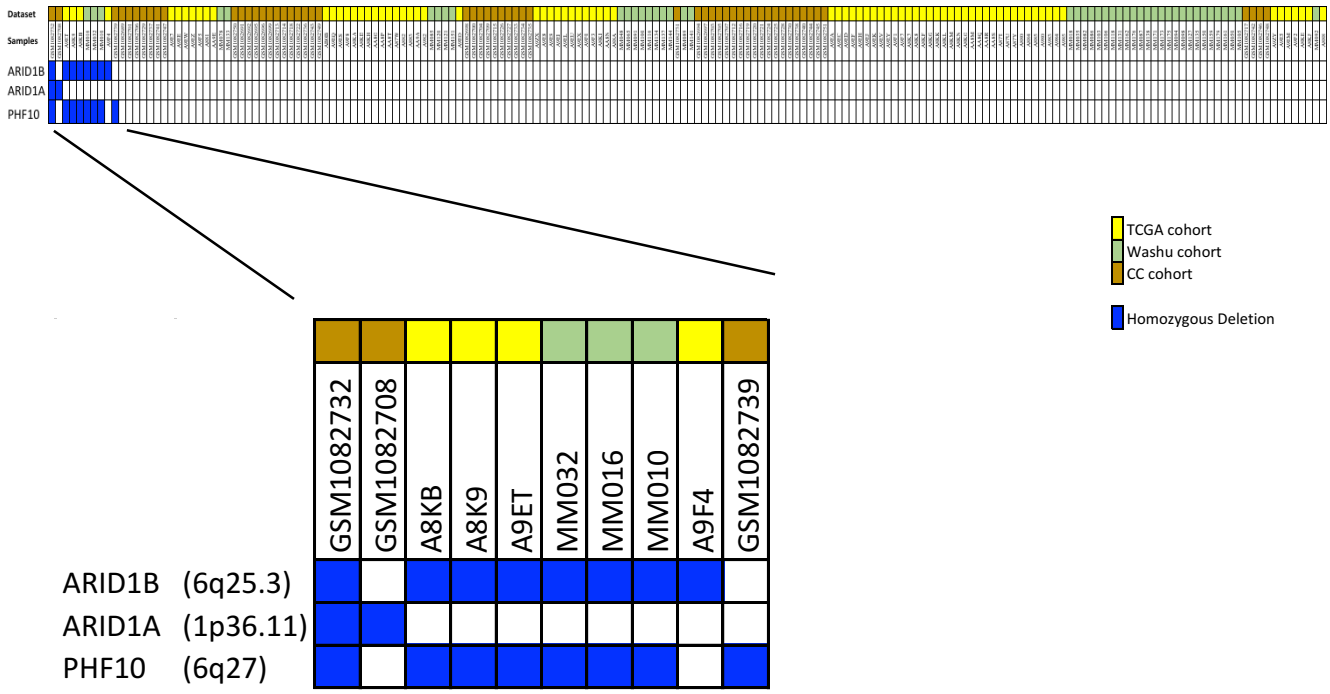
To discover other novel genes involved in UM pathogenesis, known cancer genes reported in earlier studies were queried across 1013 genes spanning 52 out of 70 focal peaks in UM previously not reported in the pan cancer analysis, these included protein coding and non-coding RNAs. Genes residing in the loci of specific amplification and deletion significant peaks identified by GISTIC were assessed to identify potential cancer causing gene (oncogenes and tumour suppressors) using a compiled list of cancer driver genes previously used for comparing shared alterations. Interestingly, none of the focal amplification peaks contain a known candidate cancer gene reported in these studies [209,339,340], whereas 6 out of 29 deletion peaks contained cancer genes previously not described in UM. These included genes *ARHGAP32* (11q24.3), *ARID1B* (6q24.3), *CDON* (11q24.3), *ESR1* (6q24.3), *FAM118B* (11q24.3), *FANCA* (16q24.3), *PTMA* (2q37.2), *SHPRH* (6q24.3), *SMG1* (16p12.3), *SYNE1* (6q24.3), *TJP2* (9q21.11) and *TNFAIP3* (6q24.3). To prioritize genes that are likely to have highest impact, copy number states indicating high amplification (exceeding higher amplicons threshold 0.9) or homozygous deletion (exceeding higher deletion threshold 1.3) events from gene GISTIC analysis were examined.

Gene AT-Rich Interaction Domain 1B (*ARID1B)* is part of the human SWI/SNF chromatin remodelling complex involved in mobilizing nucleosomes and modulates transcription in an ATP dependent manner [341]. In the current dataset, *ARID1B* was found to be in a region of homozygous deletion on chromosome 6 in 8 of 182 (4%) samples, with broader, single copy deletions of *ARID1B* occurring in 28% of other samples. *ARID1A*, maps to 1p36.11 and encodes an alternate subunit of the human SWI/SNF complex [341]. This was found to be in a region of homozygous deletion in 2 of 182 sample **(Figure 3.7)**.

In addition to previously implicated cancer genes (*ARID1A*, *ARID1B*) involved in chromatin remodelling complex, another gene PHD Finger Protein 10 (*PHF10*) which encodes for PBAF, another subtype of the human SWI/SNF chromatin remodelling complex [342] was identified within the 6q27 deletion peak in two independent cohorts **(Table 3.3)**. 6q27

was focally deleted in 30% of samples and homozygous deletions were detected in 8 out of 182 samples (4%) with a broader, single copy deletion seen in 25% of samples **(Figure 3.7)**.

Figure 3.1.9: Copy number alteration of genes in the BAF and PBAF complex.



Focally homozygous deletion observed in three genes from the BAF/PBAF complex, in 10 out of 182 tumours. Sample with the alterations are zoomed in.

The focal alteration on 6q27 harbouring *PHF10* was first identified in the WASH-U cohort after examining the LRR data for samples with a 6q deletion that had undergone metastasis (MM010, MM065, MM016 and MM089). A homozygous deletion (chr6:169756330-170141552) in one of the sample (MM016) spanning *PHF10* was found **(Figure 3.8)**, and later homozygous deletions of a region harbouring this gene were detected in two samples from the CC cohort and three samples of the TCGA cohort **(Figure 3.9).** BAF and PBAF (or the human SWI/SNF) complexes alter the epigenetic landscapes to regulate transcription thus contributing to malignancies and these alterations point to important candidate genes in the chromatin remodelling complexes.

Figure 3.8: Log R ratio data for selected samples from WASH-U cohort.

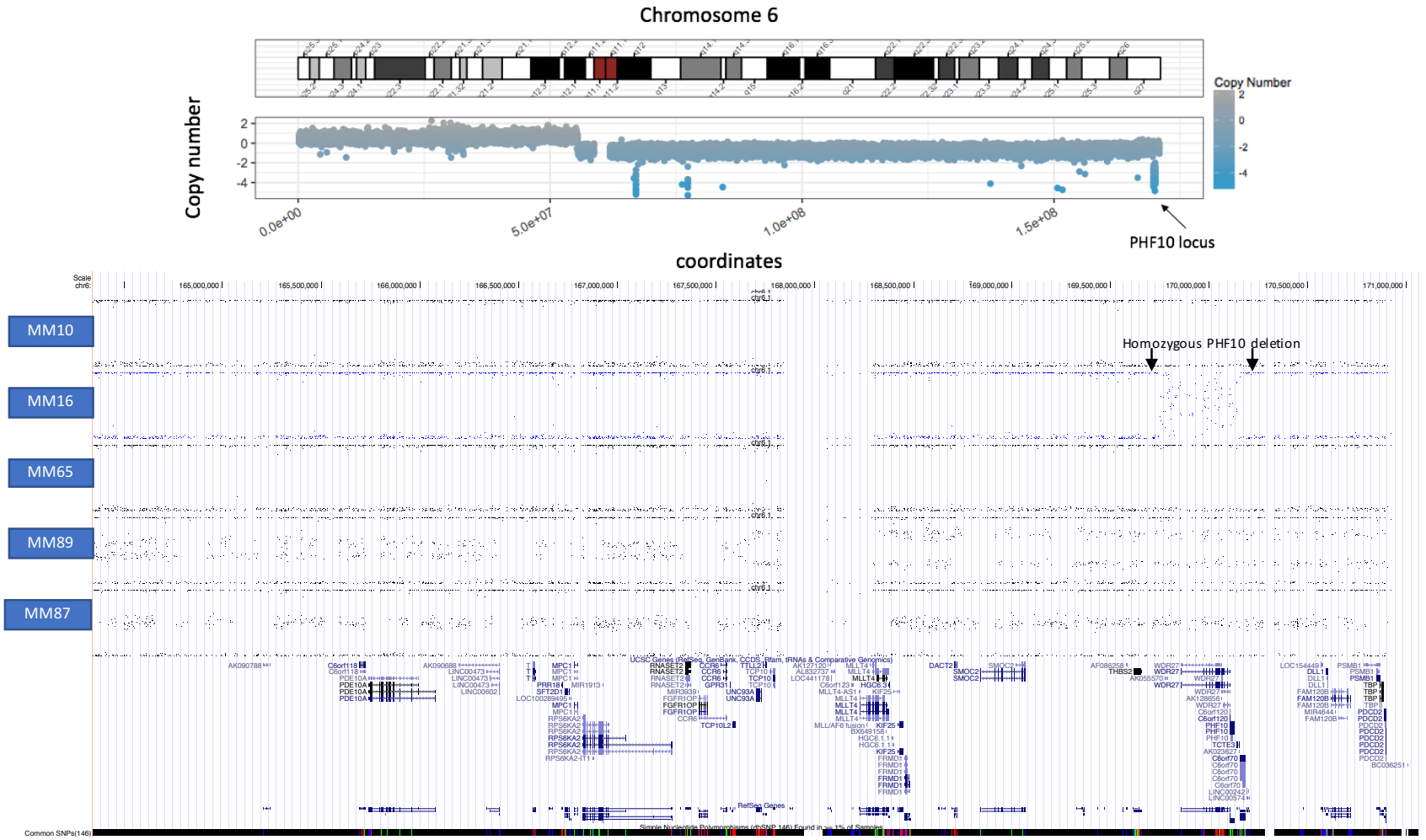
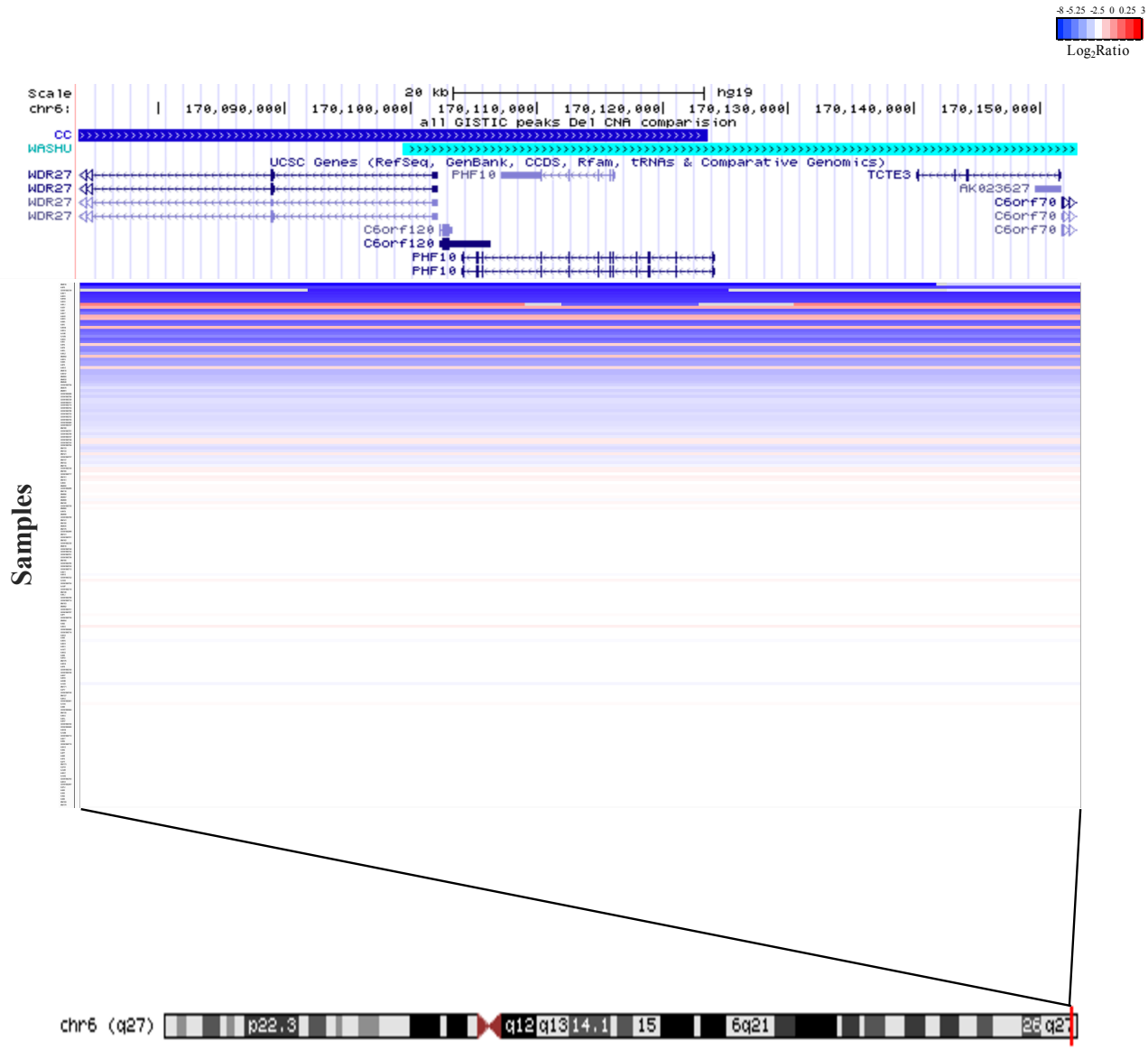


Figure (Top) showing homozygous deletion in *PHF10* gene at 6q27 identified in MM016. Shown (bottom) the zoomed in view of change in log R ratio in comparison to other samples.

Figure 3.9: Homozygous deletion of *PHF10* in IGV



The regions of homozygous deletion at chromosome 6q27 in the CC and WASH-U cohorts showing the region of overlap that harbours *PHF10* and *c6orf120*. The IGV heatmap of 182 segmented tumour profiles is shown below with gains in red and losses in blue. The samples are ordered by their copy number value which represents the focality and amplitude of loss.

3.1.6 Differentially expressed genes in focal peaks and pathway analysis

The findings of the current study have identified CNAs which have been reported in earlier UM studies, have potentially refined the critical locations of some of these CNAs and have identified some novel alterations. However, to understand their role in UM pathogenesis, the consequences of such genomic alterations need to be understood. For a copy number change to confer a selective advantage, one also expects to see a change in gene expression. Refinement of such genes in tumorigenesis has been performed successfully in the case of BAP1, where deletion of a critical tumour suppressor led to loss of expression. However, recent studies have also shown that genomic rearrangements can sometimes destroy topological domains, leading expression of an RNA in a nearby region, sometimes as a consequence of enhancer hijacking. It is also possible that multiple genes residing within a CNA contribute to tumourigenesis [343].

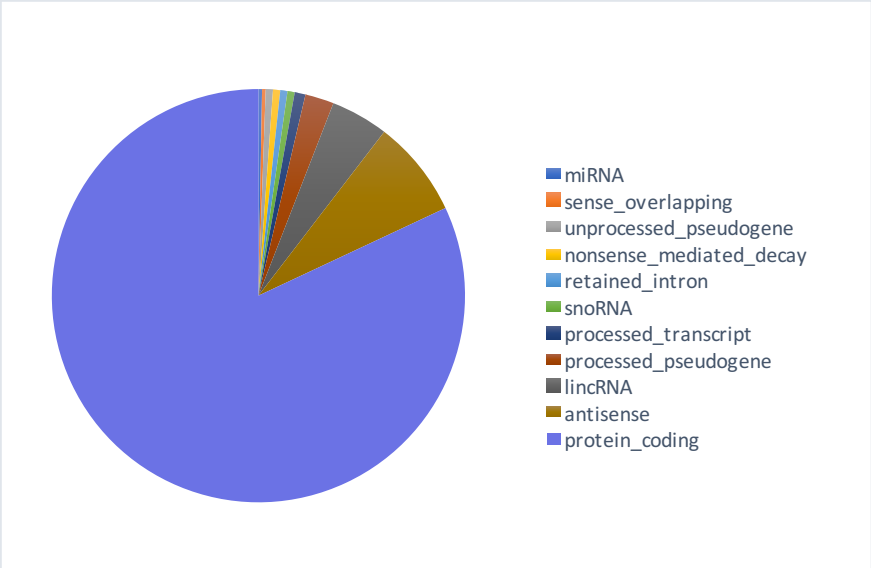
Thus, to prioritize genes that are most likely to have a functional effect due to direct consequence of changes in copy number 80 TCGA samples with matched SNP array and RNA-sequencing data were studied. All genes that mapped to 70 focal amplified and deleted regions were analyzed for association between the presence or absence of a focal alteration and transcript abundance. Here the genes that lie within the significant focal peak boundaries are defined as 'cis-signature' genes. To extract intrinsic biological patterns of these cis-signature genes, gene enrichment analysis was then applied to the all the cis-signature genes that were statistically significantly associated with change in gene expression (FDR q-value < 0.05).

In total 405 cis-genes spanning 70 target regions of interest were identified that were associated with a change in expression. Majority of genes were protein coding mRNAs followed by different RNA species such as antisense RNA, long intergenic non-coding RNA and pseudogenes **(Figure 3.10).** To extract intrinsic biological patterns of these cis-signature genes, 318 out of 405 genes were tested for enrichment of relevant biological categories using collection C2 curated gene sets from the MSigDB repository (Subramanian et al, 2005) [344] using a nominal *p*-value < 5%. The C2 collection provides a significant number of highly enriched genes and are extracted mainly from biomedical literature and biological databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [345] and Reactome [346].

In the current analysis, all the deregulated genes (overexpressed and down regulated) were analyzed together. Although, this approach could reduce power and make interpretation easier, by considering all the genes, positive/negative interactions between genes are

accounted for and could highlight pathways where both positive and negative regulators are present. The second approach was run for comparison, where the genes in the amplified and deleted loci were analyzed separately. Overexpression genes (amplified loci) were enriched only for the WNT signaling pathway and downregulated genes (deleted loci) were enriched for RNA processing and Immune response.

Figure 3.10: Distribution of gene categories for cis genes.



For a list of 405 genes found across 70 focal copy number target regions that are significantly associated with change in expression, the distribution of the categories in which the genes lie. Most of them were protein coding.

The integrative analysis of the union of cis-signature genes and corresponding transcript abundance was performed to characterize functional pathways affected by these consistently amplified and deleted regions. The pathway enrichment of genes perturbed by FCNAs in UMs revealed significantly enriched Reactome and KEGG pathways including processes related to immune response, cell cycle and mRNA regulation (**Table 3.5)**. Some of these gene sets have different names but include common genes targeted by FCNAs. Genes that are present in 3 or more pathways include *RPS6KA1*, *PSMB2*, *RPA2*, *SNRNP40*, *SRSF4*, *NUDC*, *HDAC1*, *RBBP4*, *CDC42*, *DNAJC8*, *HNRNPR*, *SRRM1* on deletion peak 1p35.3, *TIRAP* on 11q24.3 deletion peak, *TAB2*, *NUP43*, *TNFAIP3*, *FBXO5*, *SF3B5* on deletion peak 6q24.3, *RPS6KA2* on deletion peak 6q27, *IKBKB* on deletion peak 8p11.21 and *RIPK2* on amplification peak 8q21.3.

The top significantly enriched gene sets were seen in immune related genes (FDR p-value 0.00244). Alterations that perturbed the genes with immune related function include deletions in the 1p35.3 peak (*RPS6KA1*), the 6q24.3 peak (*TAB2*, *TNFAIP3*), the 8p11.21 peak (*IKBKB*), the 6q27 peak (*RPS6KA2*) and the 11q24.3 peak (*TIRAP*) and amplification of the 8q21.3 peak (Receptor Interacting Serine/Threonine kinase 2: *RIPK2*). *RIPK2 is* focally amplified in majority of UM samples (70%), and plays an important role in modulating the innate and the adaptive immune responses. *RIPK2* activates the NF-kappaB and MAPK pathways by interacting with *NOD1* and *NOD2* via its CARD domain [347,348]. NF-kappaB pathway activation results in the upregulation of transcription and production of inflammatory mediators and has been linked to cancer development [349], One study has already shown this pathway to be active in both primary and metastatic UMs [350]. The kinase gene, *RPS6KA2* on 6q27 deletion (seen in 30% of UMs) has been shown to be a putative tumor suppressor gene in ovarian cancer [351].

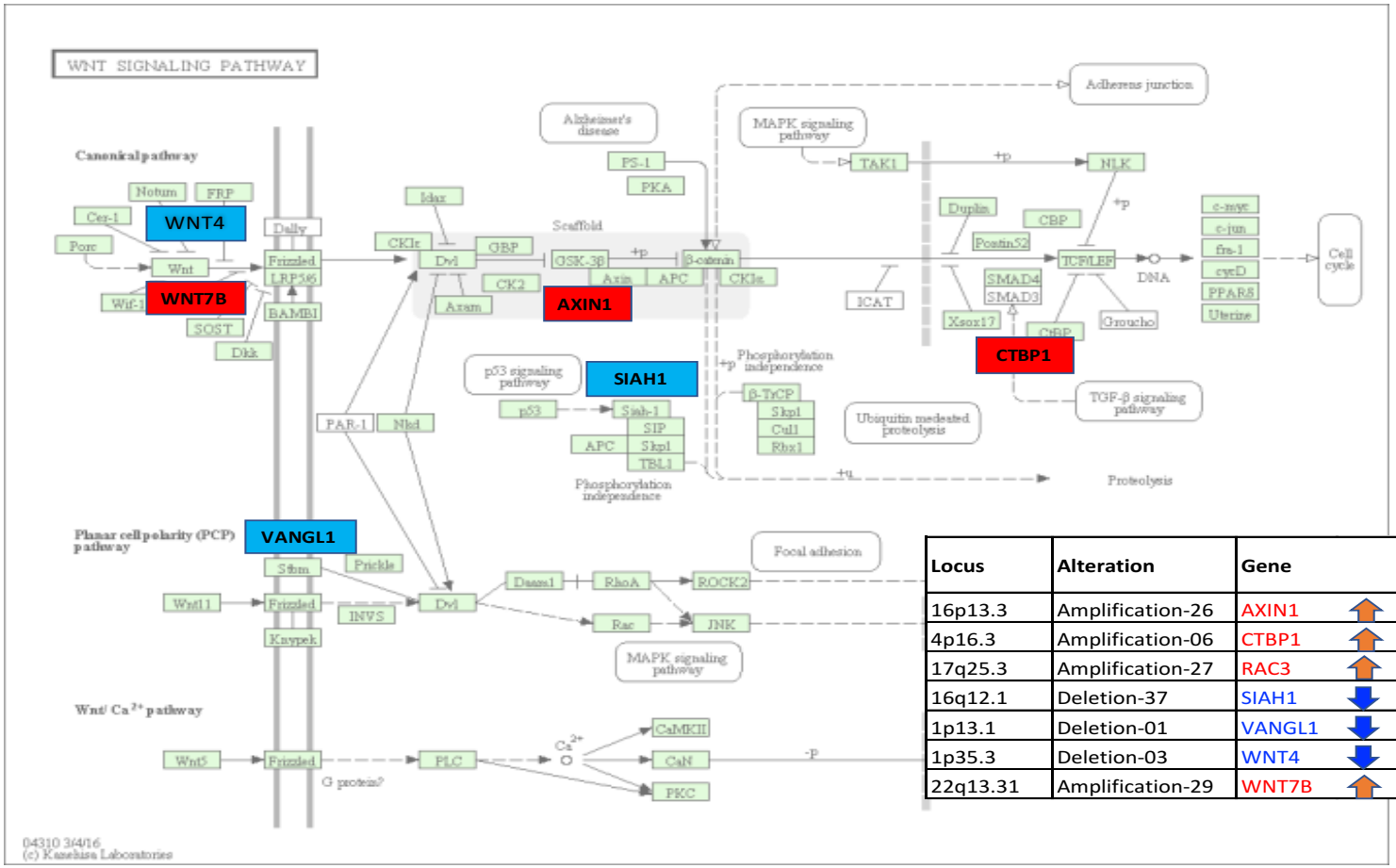
Interestingly, seven genes closely associated with copy loss of 1p13.1 (*VANGL1*), 1p35.3 (*WNT4*) and 16q12.1 (*SIAH1*) and copy gain of 16p13.3 (*AXIN1*), 17q25.3 (*RAC3*), 4p16.3 (*CTBP1*) and 22q13.31 (*WNT7B*) are components of the WNT signaling pathway (FDR p-value 0.0129) (**Figure 3.11**). Although this pathway has not been implicated in UM, the WNT signal transduction pathway maintains the integrity of the stem-cell and promotes self-renewal [352]. Down regulation of WNT pathway regulators, including *CSNK2A2* and *CTNNB1* was previously shown by Onken et al [42,114] in metastasizing UMs. Another study provided evidence for association between immunohistochemical staining of WNT signalling members Wnt5a, MMP7 and beta-catenin, and aggressiveness of UMs [353]. Hence, CNA analysis may have provided important insights into UM pathogenesis by highlighting changes affecting WNT signaling.

Table 3.5: Significant enriched genes sets.

Gene Set Name	Description	Set annotated	# Genes in Gene Set	Total annotated	Set size	p-value	FDR q-value	Hits
REACTOME_IMMUNE_SYSTEM	Genes involved in Immune System	21	933	45956	860	2.84E-06	2.44E-03	ASB1, ATG7, CDC42, DDOST, EIF4G3, IFNGR1, IKBKB, NUP43, PROS1, PSMB2, PSMD1, RAP1GAP, RIPK2,
REACTOME_MYD88_MAL_CASCADE_INITIATED_ON_PLASMA_MEMBRANE	Genes involved in MyD88:Mal cascade initiated on plasma membrane	6	83	45956	860	2.53E-05	1.00E-02	IKBKB, RIPK2, RPS6KA1, RPS6KA2, TAB2, TIRAP
REACTOME_ACTIVATED_TLR4_SIGNALLING	Genes involved in Activated TLR4 signalling	6	93	45956	860	4.82E-05	1.00E-02	IKBKB, RIPK2, RPS6KA1, RPS6KA2, TAB2, TIRAP
REACTOME_NOD1_2_SIGNALING_PATHWAY	Genes involved in NOD1/2 Signaling Pathway	4	30	45956	860	5.35E-05	1.00E-02	IKBKB, RIPK2, TAB2, TNFAIP3
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	Genes involved in Processing of Capped Intron-Containing Pre-mRNA	7	140	45956	860	5.82E-05	1.00E-02	DNAJC8, HNRNPR, NUP43, SF3B5, SNRNP40, SRRM1, SRSF4
KEGG_WNT_SIGNALING_PATHWAY	Wnt signaling pathway	7	151	45956	860	9.37E-05	1.29E-02	AXIN1, CTBP1, RAC3, SIAH1, VANGL1, WNT4, WNT7B
REACTOME_MRNA_SPLICING	Genes involved in mRNA Splicing	6	111	45956	860	1.29E-04	1.29E-02	DNAJC8, HNRNPR, SF3B5, SNRNP40, SRRM1, SRSF4
REACTOME_MRNA_PROCESSING	Genes involved in mRNA Processing	7	161	45956	860	1.40E-04	1.29E-02	DNAJC8, HNRNPR, NUP43, SF3B5, SNRNP40, SRRM1, SRSF4
REACTOME_NFKB_AND_MAP_KINASES_ACTIVATION_MEDIATED_BY_TLR4_SIGNALING_REPE	Genes involved in NFkB and MAP kinases activation mediated by TLR4 signaling	5	72	45956	860	1.47E-04	1.29E-02	IKBKB, RIPK2, RPS6KA1, RPS6KA2, TAB2
REACTOME_INNATE_IMMUNE_SYSTEM	Genes involved in Innate Immune System	9	279	45956	860	1.58E-04	1.29E-02	IKBKB, RIPK2, RPS6KA1, RPS6KA2, TAB2, TNFAIP3, TIRAP, DDOST, PROS1
REACTOME_TRIF_MEDIATED_TLR3_SIGNALING	Genes involved in TRIF mediated TLR3 signaling	5	74	45956	860	1.68E-04	1.29E-02	IKBKB, RIPK2, RPS6KA1, RPS6KA2, TAB2
REACTOME_TOLL_RECEPTOR_CASCADES	Genes involved in Toll Receptor Cascades	6	118	45956	860	1.80E-04	1.29E-02	IKBKB, RIPK2, RPS6KA1, RPS6KA2, TAB2, TIRAP
REACTOME_TRAF6_MEDIATED_INDUCTION_OF_NFKB_AND_MAP_KINASES_UPON_TLR7_8_	Genes involved in TRAF6 mediated induction of NFkB and MAP kinases upon	5	77	45956	860	2.02E-04	1.34E-02	IKBKB, RIPK2, RPS6KA1, RPS6KA2, TAB2
KEGG_SPLICEOSOME	Spliceosome	6	128	45956	860	2.80E-04	1.67E-02	SF3B5, SNRNP40, SRSF4, SRSF10, SYF2, CCDC12
REACTOME_NUCLEOTIDE_BINDING_DOMAIN_LEUCINE_RICH_REPEAT_CONTAINING_RECEPT	Genes involved in Nucleotide-binding domain, leucine rich repeat containing	4	46	45956	860	2.92E-04	1.67E-02	IKBKB, RIPK2, TAB2, TNFAIP3
REACTOME_MITOTIC_G1_G1_S_PHASES	Genes involved in Mitotic G1-G1/S phases	6	137	45956	860	4.03E-04	2.04E-02	PSMB2, PSMD1, RPA2, HDAC1, RBBP4, FBXO5
REACTOME_MAP_KINASE_ACTIVATION_IN_TLR_CASCADE	Genes involved in MAP kinase activation in TLR cascade	4	50	45956	860	4.03E-04	2.04E-02	RIPK2, TAB2, RPS6KA1, RPS6KA2
REACTOME_ADAPTIVE_IMMUNE_SYSTEM	Genes involved in Adaptive Immune System	12	539	45956	860	4.29E-04	2.05E-02	RIPK2, TAB2, PSMB2, PSMD1, IKBKB, ASB1, ATG7, CDC42, RAP1GAP, TRIM63, UBE2F, ULBP3
REACTOME_CELL_CYCLE_MITOTIC	Genes involved in Cell Cycle, Mitotic	9	325	45956	860	4.82E-04	2.18E-02	PSMB2, PSMD1, RPA2, HDAC1, RBBP4, FBXO5, NUP43, CSNK1D, NUDC
REACTOME_TAK1_ACTIVATES_NFKB_BY_PHOSPHORYLATION_AND_ACTIVATION_OF_IKKS_C	Genes involved in TAK1 activates NFkB by phosphorylation and activation of IKKs	3	23	45956	860	5.25E-04	2.26E-02	RIPK2, TAB2, IKBKB
KEGG_MAPK_SIGNALING_PATHWAY	MAPK signaling pathway	8	267	45956	860	5.91E-04	2.42E-02	TAB2, IKBKB, CDC42, RPS6KA1, RPS6KA2, RAC3, STMN1, MAP3K6
REACTOME_CELL_CYCLE	Genes involved in Cell Cycle	10	421	45956	860	7.82E-04	3.06E-02	PSMB2, PSMD1, RPA2, HDAC1, RBBP4, FBXO5, NUP43, CSNK1D, NUDC, SYNE1
KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	NOD-like receptor signaling pathway	4	62	45956	860	9.15E-04	3.42E-02	IKBKB, TAB2, RIPK2, TNFAIP3
REACTOME_MITOTIC_M_M_G1_PHASES	Genes involved in Mitotic M-M/G1 phases	6	172	45956	860	1.32E-03	4.73E-02	PSMB2, PSMD1, RPA2, FBXO5, NUP43, NUDC

All the pathways (FDR <0.05) from KEGG and Reactome databases in MSigDB significantly enriched for copy number target genes which are deregulated in uveal melanoma, and are associated with change in expression. Amplified genes are shown in red and deleted genes in blue.

Figure 3.11: The WNT signaling pathway highlighting genes perturbed by focal copy number alteration.



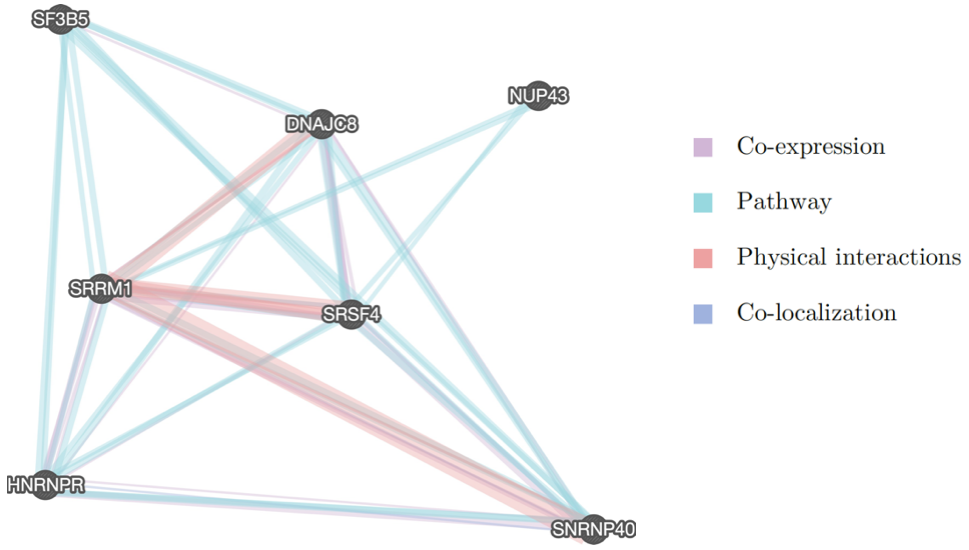
Locus	Alteration	Gene	Frequency (%)
16p13.3	Amplification-26	AXIN1	15
4p16.3	Amplification-06	CTBP1	15
17q25.3	Amplification-27	RAC3	15
16q12.1	Deletion-37	SIAH1	21
1p13.1	Deletion-01	VANGL1	21
1p35.3	Deletion-03	WNT4	29
22q13.31	Amplification-29	WNT7B	17

This includes genes identified by the cis-signature pathway enrichment analysis. *RAC3* which is related to the WNT pathway has not been captured here. Amplified genes are shown in red and deleted genes in blue. The accompanying table indicates the frequency of samples with alterations of specific genes alteration in the focal region of interest. The pathway diagram is adapted from KEGG pathway for Wnt Signalling.

Another significantly enriched gene set is "Processing of Capped Intron-Containing Pre-mRNA pathway" (FDR 0.01). This encompasses genes involved in mRNA splicing and transport of mature mRNAs. Seven genes involved in this pathway include a cluster of 6 genes with copy loss residing in the 1p35.3 peak (*SRSF4, SNRNP40, SRRM1, HNRNPR, DNAJC8* and *NUP43)* and one splicing factor gene (*SF3B5)* in the 6q24.3 peak. The *SRRM1* gene (alias *SRm160*) is a SR-related protein involved in pre-mRNA processing events and splicing activation via constitutive and exonic splicing enhancer (ESE)-dependent mechanisms. It brings together sequence specific SR family proteins (such as *SFRS4*, *SFRS5* and *SFRS10*) and basal small nuclear ribonucleoproteins or snRNPs (such as *SNRP70* and *SNRPA1*) to form the spliceosomes [354–359].

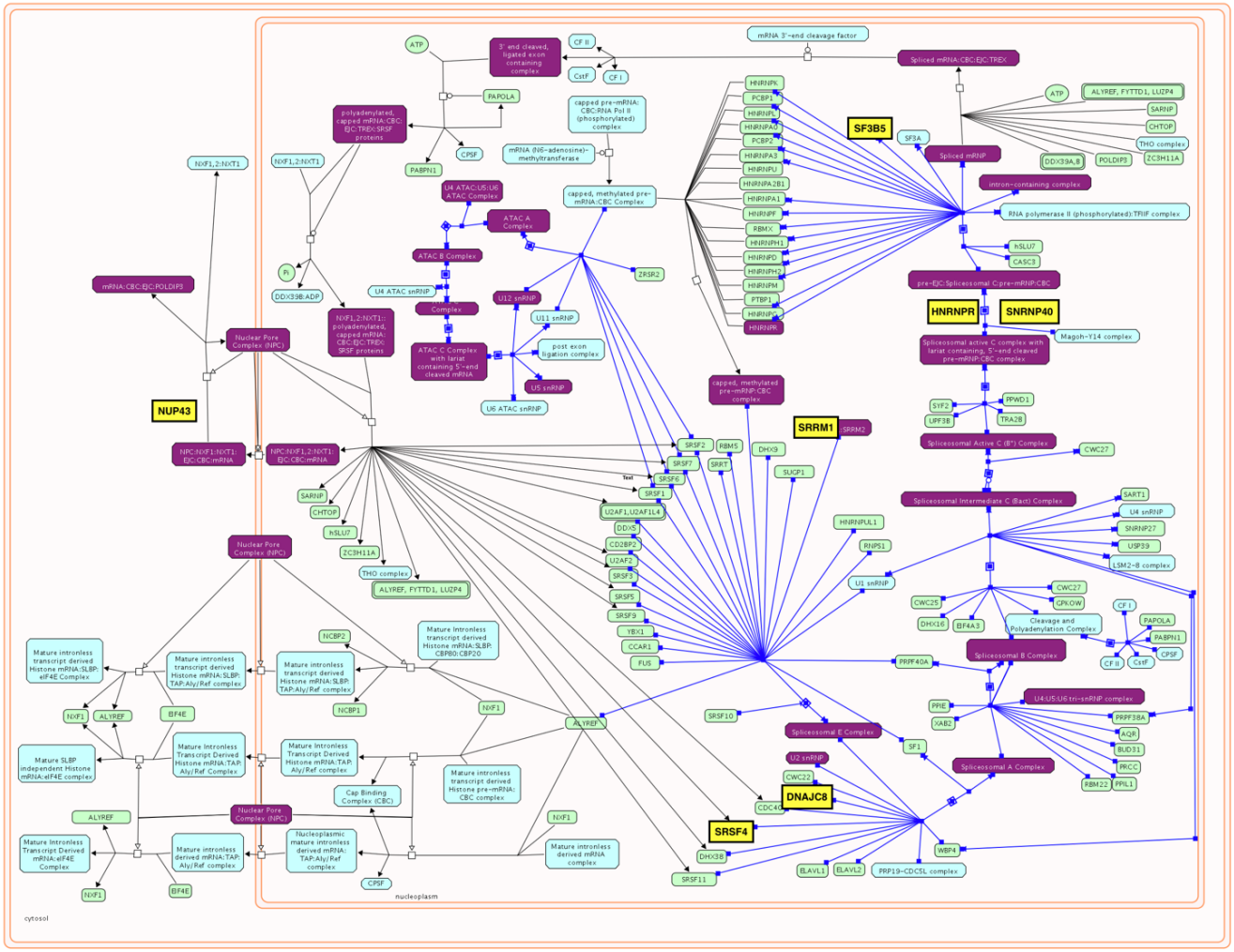
Gene interaction network showed strong physical interaction network between *SRRM1* and the genes *SRSF4*, *SNRNP40* (component of U5 snRNP subunit) and *DNAJC8*, supporting its role in bridging SR proteins and snRNP **(Figure 3.12)**. *SRRM1 also* induces apoptosis and regulate eye and genital development in Drosophila [360]. UMs can harbor mutations in SF3B1, and the alterations lead to change in splicing. These complementary set of genes on FCNAs could represent alternative mechanism of regulating the splicing machinery in UMs **(Figure 3.13)**. Further examination of these genes to see if they are affected by mutation could help in understanding the role of this additional mechanism of transcriptomic regulation and is described in the chapter on exome sequencing.

Figure 3.12: Gene interaction network



The figure shows genetic connections between the genes with altered copy number from the top most enriched pathway which relates to RNA processing and mRNA splicing. Here the genes are connected based on Pathways, co-expression, co-localization and physical interaction. The nodes represent the genes and the links represent the interactions. The weight, represent by the thickness of the link is a measure of how informative the network for these 7 genes is.

Figure 3.13: RNA processing and spliceosome genes.



Adapted from Reactome Processing of Capped Intron-Containing Pre-mRNA pathway highlighting in yellow boxes FCNA genes in the pathway.

3.2 Discussion

Using a combination of high density SNP arrays, gene expression data and pathway analysis, novel candidate cancer driver genes and perturbed pathways in UM have been identified. Although previous studies have described profiling of copy number events in UM, the current study incorporated deep global copy number profiling in large samples size and integrated that with gene expression data. The latest high density SNP platforms that were used also offered more precision in determining the boundaries of recurrent alterations, allowed the examination of additional copy number events such as copy neutral-LOH events and estimated tumour ploidy, which was not possible with the older aCGH arrays.

Clustering of broad copy number revealed four groups characterized mainly by alterations involving 1p loss, 1q gain, monosomy 3, 6p gain, 6q loss, 8p gain/loss 8q gain and 16q loss. Previously, Harbour et al [38] identified prognostically relevant groups using chromosome 3 and 6p status, where they found that 3- and 6p+ occurs in a mutually exclusive manner. However, examining data in the current study revealed a small subset of tumors with both 3- and 6p+. Recently, expression of PRAME tumor antigen has been identified as an independent marker of metastasis and associated with additional chromosomal alterations including, 1q gain, 6p gain and 16q loss. However, these alterations did not reach statistical significance for associated with metastasis independent of PRAME status in the current study. Although, gain of 6p was found in a third of all UMs, as previously shown, the effect of gain of 6p on patient outcome is not conclusive. Although it has been shown to be related to longer survival [35], it co-occurred in tumours with Monosomy 3 in a small subset of samples (11%, 20/182) which could perhaps indicate tumour heterogeneity with ongoing sub-clonal evolution.

The distribution of ploidy and aberrant cell fraction across UM suggests that these tumours are genomically homogeneous with majority of tumors having ploidy close to 2n with a small fraction of tumours exhibiting tetraploid state. Although, there was no significant differences in the ploidy estimates between the class 1 and class 2 subgroups, class 2 tumours tend to exhibit higher ploidy states (tetraploidy). Monosomy 3 which is characteristic feature of class 2 tumors leads to accumulation of aneuploidy and further increase genomic instability [129]. The aberrant cell fraction estimate reflects intratumoural non-aberrant cells and normal cells surrounding the tumour. Here, the class 2 tumors showed a higher percent of non-aberrant cells compared to class 1 tumours. These higher number of

non-aberrant cells could indicate the presence of other cells related to the immune response such tumour-infiltrating macrophages or tumour-infiltrating lymphocytes, which are associated with poor prognosis in UM and can perhaps exacerbate rather than mitigate tumor progression [361].

To uncover the candidate genes that lie in these regions with recurrent chromosomal alteration, the GISTIC algorithm was applied. In this way genes driving the loss or gain of a chromosome or chromosomal arm can be identified. GISTIC analysis of segmented profiles from the 3 cohorts revealed various non-overlapping multiple low frequency copy number peaks and posed the challenge of effectively integrating data from all 3 cohorts genotyped across different platforms. This was resolved by systematically evaluation of each GISTIC peak with platform independent segmentation profiles across all the samples and deriving a frequency based criteria to prioritize focal copy number alterations and associated genes. A further ad-hoc approach was used to increase confidence in candidate genes by combining all the GISTIC peaks into a single dataset and identifying those peaks that had overlapping boundaries and/or boundaries that lie within 1Mb across any two of the three cohorts. These were then selected as high confident peaks. This helped identify focal deletions on 1p35, 2q37.2, 6p25.2, 6p22.1, 6q26, 6q27, 8q24.3, 16p13.3, 11q24.3 and 19p13.3 and revealed a few potential candidate genes for further investigations, including *PDLIM1* (1p35.2), LOC200772 (2q37.2), *WDR27*, *PHF10*/c6orf120 (6q27) and *FLIT1* 11q24.3.

To distinguish between CNAs that are unique to UMs and those that are found in other tumours the focal CNAs from GISTIC analysis were compared with those of 3131 tumors spanning nearly 54 tumors types (Beroukhim 2010) [338]. Examining genes that lie in the shared copy number alterations across all 18 peaks and comparing them with known cancer driver genes revealed three potential cancer genes that could be implicated in UM: *ARID1A* (1p35.3 deletion peak), *NCAPD3* (11q25 deletion peak) and *PLEC1* (8q24.3 amplification peak).

To prioritize genes that are most likely to have a functional effect due to a direct consequence of changes in copy number, 80 TCGA samples with matched SNP array and RNA-sequencing data were studied. CNAs associated with change in gene expression were then identified, followed by pathway analysis of genes in altered regions. This approach revealed a large complementary set of focal copy number events that harbor candidate genes affecting many overlapping pathways. These included NF-kappa B and MAPK signalling, WNT signalling and mRNA splicing/RNA processing. The CNA with the highest frequency

was amplification of a region of chromosome 8q24.3 that contained the *PLEC1* gene. We did not find an important role for c-*MYC* amplification reported in some earlier studies.

An inflammatory phenotype characterized by the increased presence of lymphocytes and macrophages is associated with a worse prognosis in UM [29]. *RIPK2* on 8q21.3 which plays an important role in modulating the immune response by activation of the NF-kappaB pathway is amplified in 70% of UMs. This candidate gene could provide a potential link to recruitment of leukocytes in UMs. Altered components of the WNT signalling pathway include *VANGL1*, *WNT4*, *SIAH1*, *AXIN1*, *RAC3*, *CTBP1* and *WNT7B*. In addition, while recurrent somatic mutations of the splicing factor gene SF3B1 has been well established in UM [242], the current study defined an alternative copy number-dependent mechanism of deregulating mRNA processing pathways.

Another interesting theme that emerged from the current analysis included the identification of novel candidates in the chromatin remodeling complexes (*ARID1A*, *ARID1B* and *PHF10*). The systematic identification of these genes in a large cohort (N=182) as examined here, have previously not been shown in UM. These alterations included homozygous deletions seen in 1% and 4.39% of samples for *ARID1A* and *ARID1B* respectively in a mutually exclusive pattern and 4.39% of samples for *PHF10* that are associated with change in gene expression (cis-signature genes). *ARID1A* and *ARID1B* encode for mutually exclusive subunits of the BAP complex (human SWI/SNF) chromatin remodeling complex and *PHF10* is a component of the PBAF complex that is frequently inactivated in cancers. [342] It has a diverse function in the regulation of transcription. Loss of *PHF10* has been shown to lead to loss of cell proliferation in normal fibroblast cells [342] and knockdown of PHF10 in gastric cancer cells led to significant induction of caspase-3 expression at both the RNA and protein levels [362]. Thus, *PHF10,* in addition to *ARID1B* presents a candidate gene for a subset of tumours with loss of 6q.

In summary, this chapter presents a comprehensive investigation of copy number profiling in UMs using three independent cohorts and requiring an analysis of data generated on different independent platforms. Previous studies, including one recently published on the TCGA cohort[46], mainly reported broad copy number alterations, and those findings were replicated here. An analysis of SNP arrays in the current study permitted the identification of smaller genomic alterations than have previously been identified with chromosomal methods. This included the identification of focal deletions at a genome wide level that could be replicated in the different studies. This study also presents the first instance where copy number data have been integrated with gene expression data. This revealed changes in a

number of biological processes such as chromatin organization, RNA processing and WNT signalling. These require experimental follow up.

Chapter 4. Mutational landscape in uveal melanoma

4.1 Results

4.1.1 Landscape of somatic mutation in coding regions in uveal melanoma

In total, exome sequencing data from 120 patients (104 paired and 16 unpaired tumour samples) were used in the current study. Exome data was compiled from two different sources: The Cancer Genome Atlas (TCGA, 80 cases) and samples recruited at Washington University in St Louis (WASH-U, 40 cases). Exome sequencing of 120 samples resulted in an average of 3.94 Gb of sequence per tumour sample. Mean coverage of the exome captured region was 64x per tumour sample, with 97% covered with at least 30X coverage.

Coding regions of the genomes of 120 UMs contained 19,467 somatic variants after filtering all the non-coding and population level polymorphic variants **(Figure 4.1)**, of which 1200 (~6%) were insertions/deletions (INDELs) and 18267 (~94%) were single nucleotide variants (SNVs). Of the SNVs, 5075 (26%) were predicted to be synonymous; 11810 (61%), missense; 1290 (7%), nonsense and 194 (1%) splice site. Among the INDELs, the most common variants were 439 frameshift deletions (2%), followed by 373 frameshift insertions (2%), 174 non-frameshift insertions (1%) and 112 non-frameshift deletions (1%). A subset of the samples (n=3), underwent low depth of coverage whole exome sequencing. The transition/transversion (Ti/Tv) ratio for the cohort (N=120) was 2.06. The lower than expected value of 2.8 for whole exome sequencing indicates a mix of non-exonic and exonic data. The transition and transversions, revealed the presence of a high degree of C>A transversions in the hypermutated sample (MM049) **(Figure 4.2)**. Excluding hypermutated (n=1) and poor coverage samples, the average number of all types of somatic variants per samples for SNVs and INDELs, were 117.70 (range 45-1302) and 10.52 (range 1-286) respectively **(Figure 4.3)**. A comparison of somatic mutations across all the samples with the COSMIC database [363] yielded 749 mutations across 685 mutated genes reported previously. Of the 11,810 missense mutations, 7972 had a CADD score > 20, indicating they have a high probability of being deleterious.

The six classes of base substitutions were extracted, showing increased C>T transitions **(Figure 4.4)**. A mutational signature analysis revealed three mutational signatures enriched in UM and similar to signatures reported by Alexandrov et al [203] **(Figure 4.5)**.

Signature 1 & 3 was enriched in cytosine to adenine (C>A) transversion and was similar to validated signature 4 by Alexandrov et al [203] . Signature 2 (from the current study) was the cytosine to thymine transition (C>T) at the C_pG dinucleotide, which is found in all cancer types and is related to ageing [203]. Consistent with previous reports [180,198], no UVR signatures were found in UM. Comparing the number of somatic mutations (SNV and INDELs) between the gene expression (GEP) based class 1 and class 2 tumours revealed no statistical difference between the two classes of tumors (p-value 0.4167). The top 10 most frequently mutated genes include *RGPD3*, *GNAQ*, *GNA11*, *BAP1*, *TTN*, *SF3B1*, *WHAMM*, *HYDIN*, *EIF1AX* and *NUP50* **(Figure 4.6).**

Figure 4.1: Filtered variants in MutSigCV analysis.

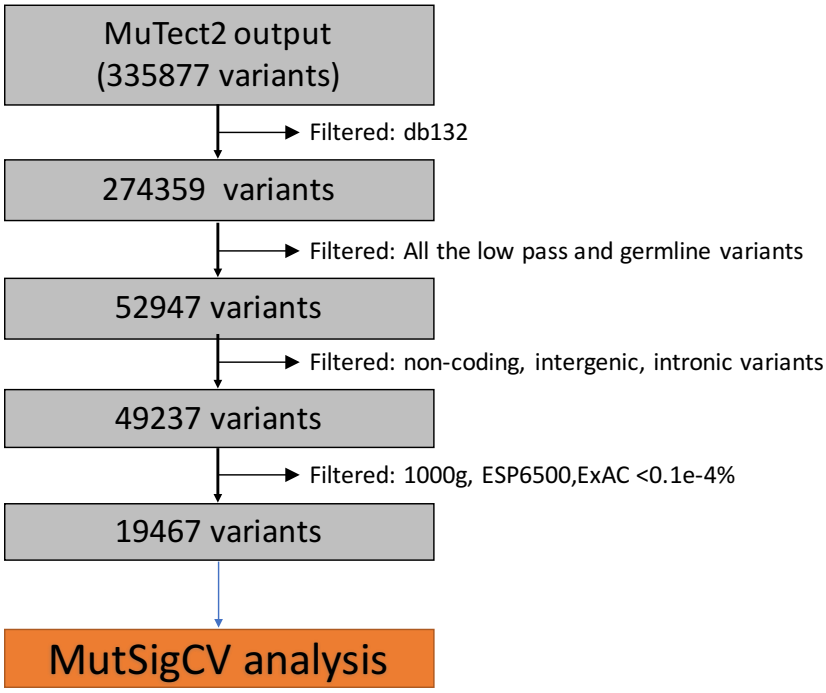
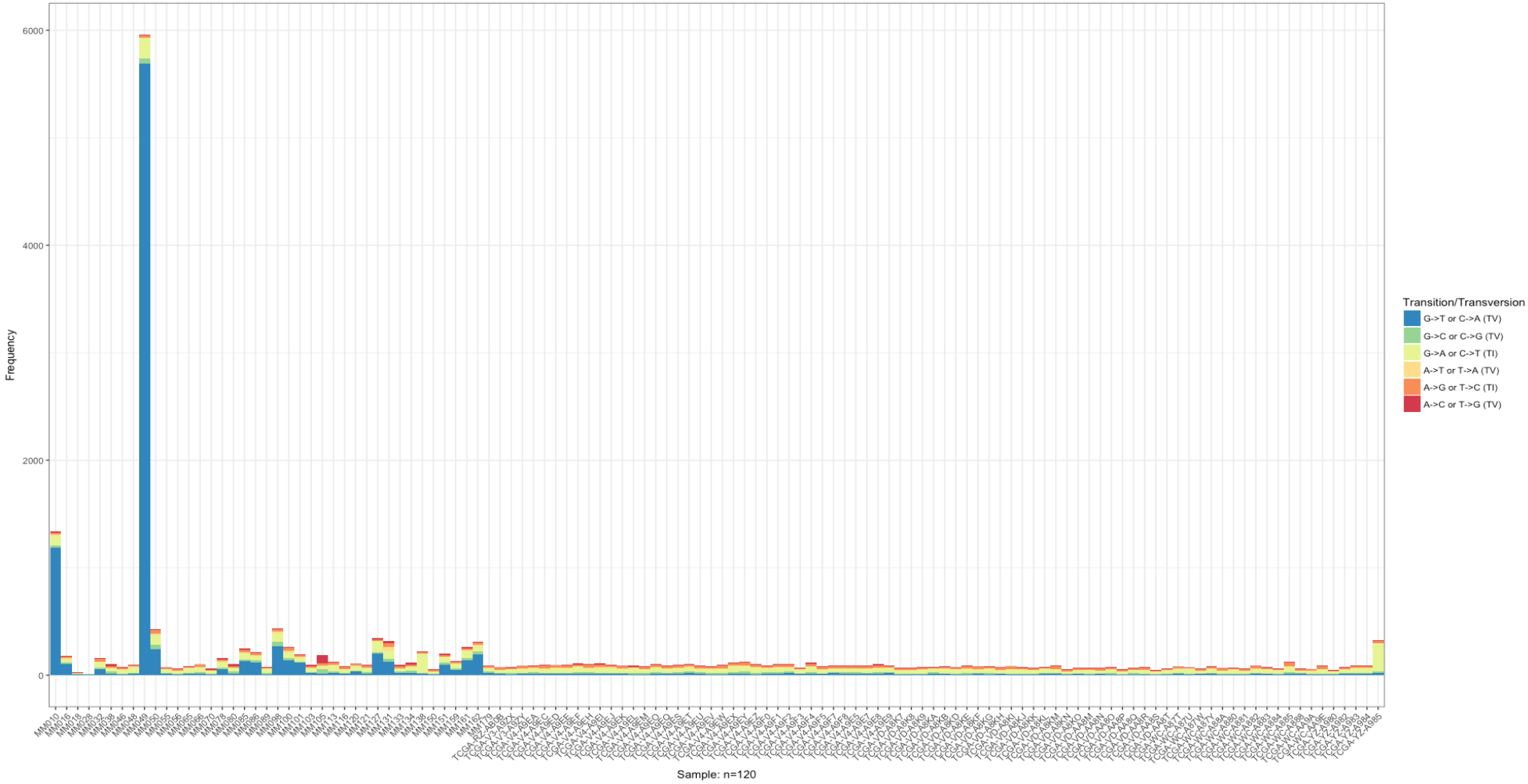


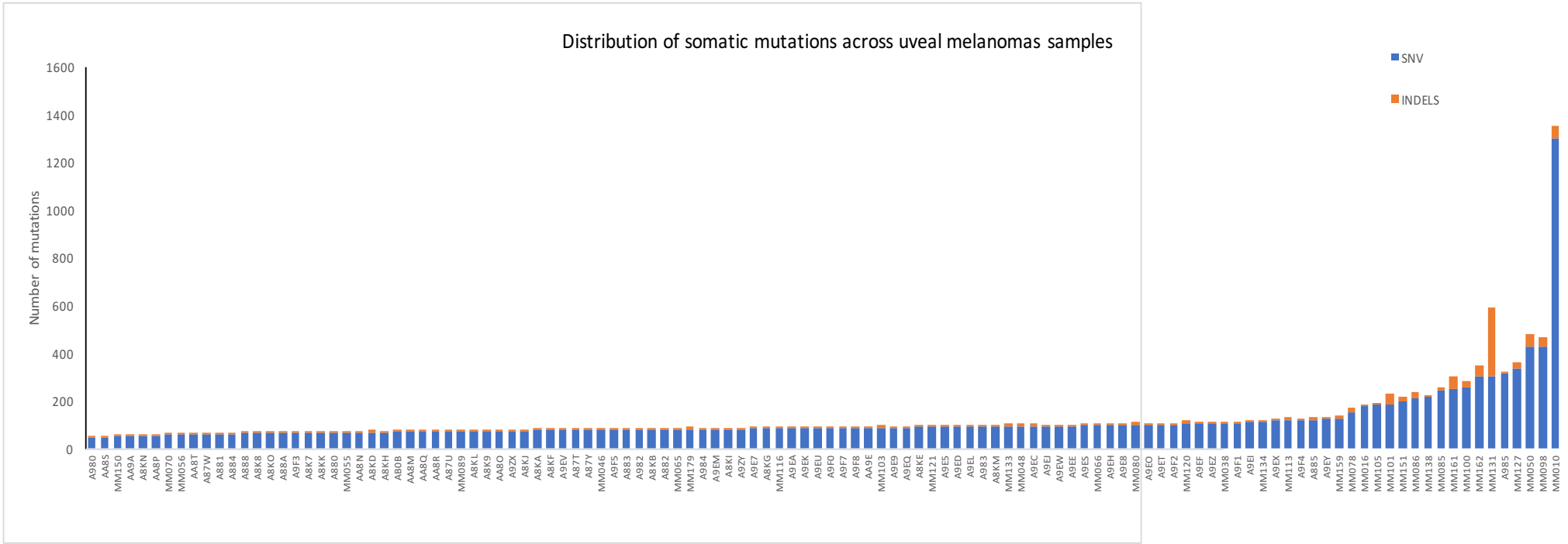
Figure showing the number of variants at each stage of the filtering steps prior to running MutSigCV analysis.

Figure 4.2: Transitions and transversions for the cohort of samples in the current study.



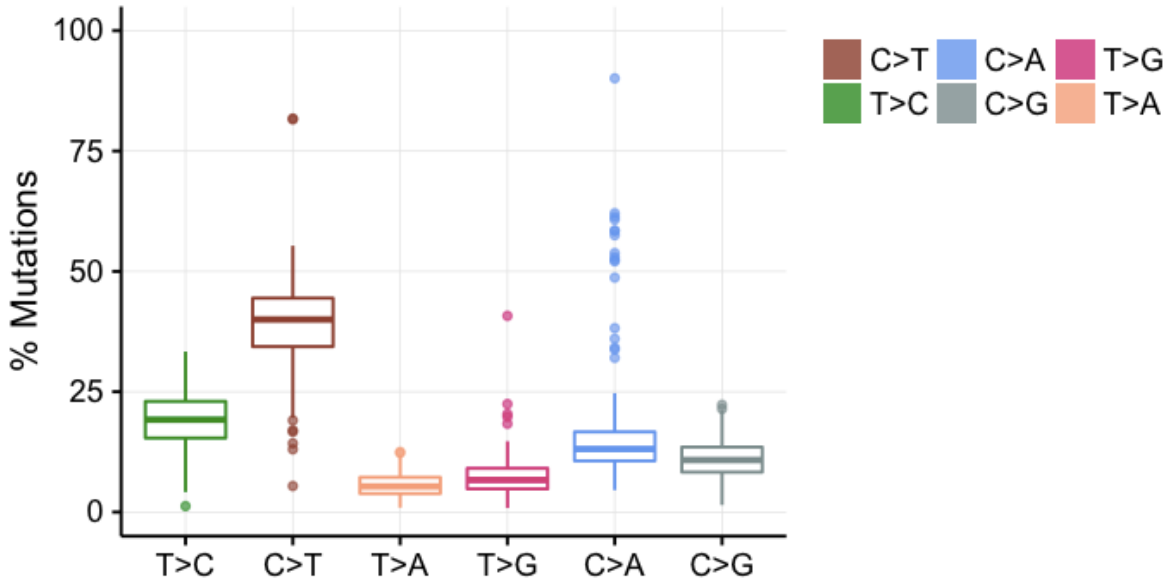
Hypermutated sample (MM049, an SF3B1 mutant (p. R625L) sample) showed an increased C>A transversion rate.

Figure 4.3: Histogram of SNV and INDELS across all the samples.



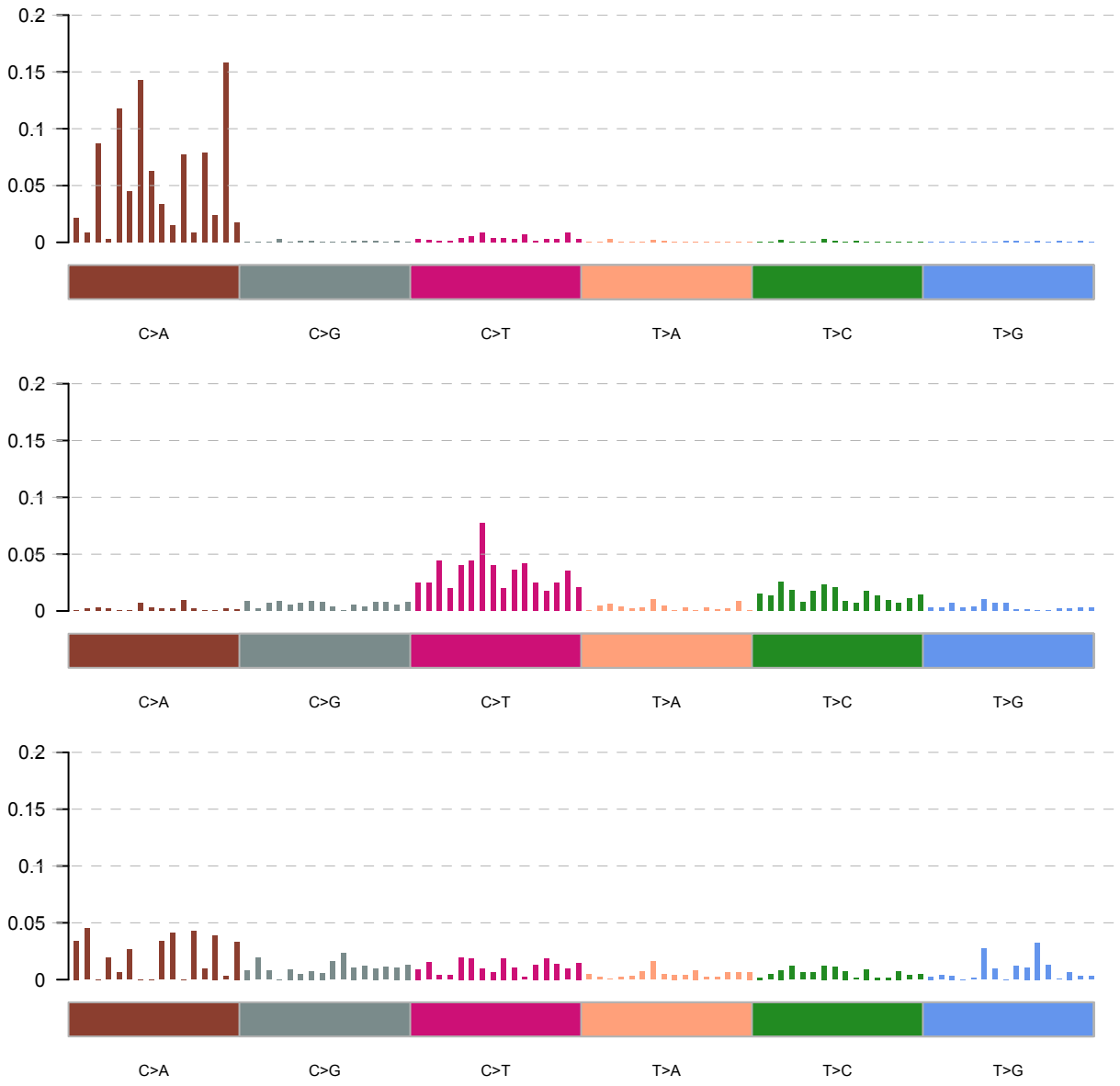
Here, the low coverage and hypermutated samples (n=4) were excluded.

Figure 4.4. Distribution of SNV substitutions.



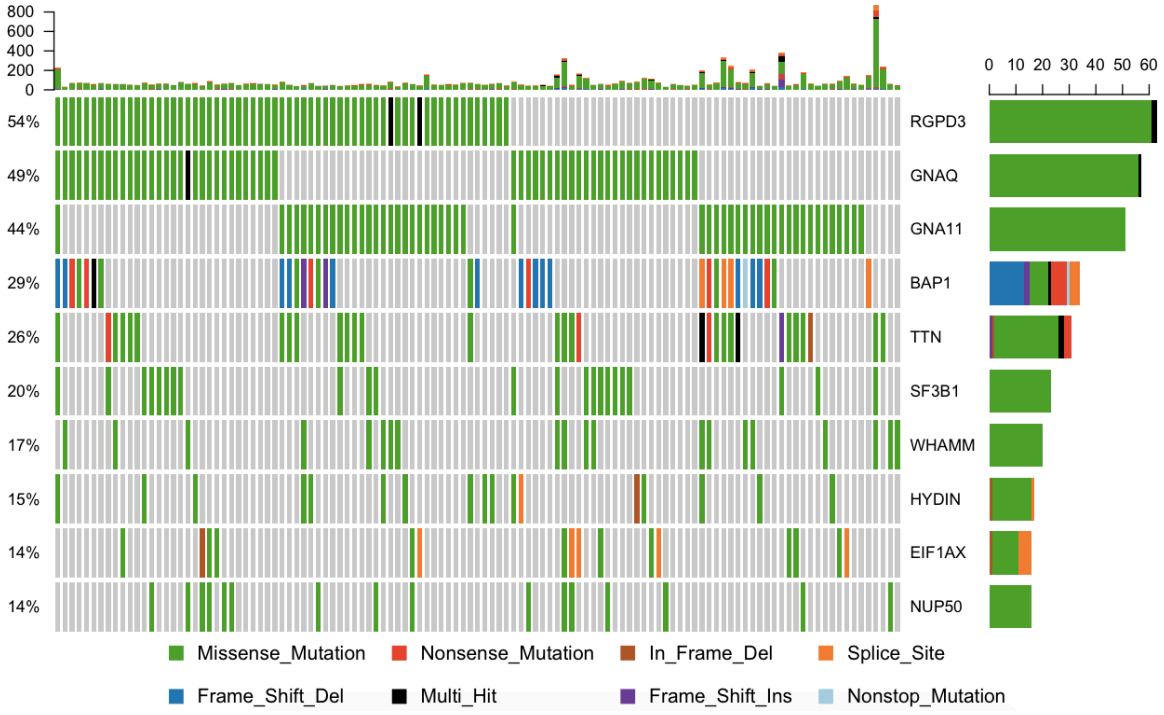
Boxplots showing overall distribution of six different classes of SNV substitutions showing increased C>T transitions.

Figure 4.5: Mutational signatures in UM.



SNVs were extracted from exome sequencing data and subjected to A mutational signature analysis with NMF. This decomposes the bases into 96 different combination of substitutions. It then takes the base change and the nucleotide context (flanking bases) to look for enrichment of a signature. Three signature were found to be enriched in UM. These signatures were enriched in C>T transitions and C>A transversions.

Figure 4.6: Highly frequently mutated genes.



The top ten most frequently mutated genes identified in the current whole exome sequencing data (N116) prior to applying MutSigCV analysis.

4.1.2 Significantly mutated genes in uveal melanoma

The genes reported above were purely based on mutation frequency. To perform a more refined analysis and to see what is truly significantly mutated at a higher rate an analysis was performed with MutSigCV [325]. This identifies mutations that are significantly higher than the background mutation rate after adjusting for gene length and base composition. MutSigCV analysis of 116 Primary UMs revealed 18 significantly mutated genes (after FDR correction q-value < 0.1) occurring in 100% of cases **(Table 4.1)** and included, in the order of FDR corrected significant p-value (percent of patients with alterations): *BAP1* (29%), *GNA11* (44%), *RGPD3* (54%), *SF3B1* (19%), *EIF1AX* (14%), *GNAQ* (48%), *HAUS6* (13%), *NUP50* (14%), *WHAMM* (16%), *AOC3* (9%), *LCE2A* (4%), *MFF* (5%), *HDGFRP3* (5%), *RBMX2* (4%), *KIAA2013* (12%), *FAM133B* (3%), *DNAJB7* (5%) and C7orf49 (3%) **(Figure 4.7 and Table 4.1)**. Although, such an analysis has not been performed for UM before, the identification of established known UM drivers (*GNAQ*, *GNA11*, *BAP1*, *SF3B1* and *EIF1AX*) with such an approach, validated it for the identification of significant mutations. Other UM driver genes recently reported did not reach an FDR cut-off level of significance and these

includes the *CYSLTR2* (p-value 0.0005) and *PLCB4* genes (p-value 0.08). Other novel recurrent mutations (>5%) were further evaluated to assess their role as drivers or technical artefacts. Manual inspection of variants from the novel significant genes (*RGPD3*, *NUP50*, *WHAMM*, *AOC3*, *LCE2A*, *MFF*, *HDGFRP3*, *RBMX2*, *KIAA2013*, *FAM133B*, *DNAJB7* and C7orf49) lead to discarding all but one, an augmin complex gene called HAUS6 (all discarded variants and reasons for not considering them further are listed in the **Appendix 3**). The most recurrent and significantly mutated gene *RGPD3*, was discarded based on low percentage of identical read counts (~5%, likely artefact) and low CADD score (less than 10). These variants were not deleterious and less likely to affect the protein function in a detrimental manner. Other variants (from genes *AOC3*, *C7orf49*, *DNAJB7*, *FAM133B*, *HDGFRP3, KIAA2013*, *LCE2A* and *RBMX2*) were discarded since they were present in only one of the two cohorts (either WASH-U or TCGA, but not both). These variants are likely to represent sequencing error artefacts due to batch effects. *HAUS6*, which harboured deleterious mutations in 14 out of 120 (11.66%) exomed samples was found to be a good candidate gene for follow up. Two *HAUS6* mutations with a CADD score greater than 10 were found, including p.E743D and a hotspot p.L553F alterations. However, both the hotspot mutations did not lie within the HAUS_6 domain of the protein **(Figure 4.8)** and majority of the HAUS6 variants were present at a low allele frequency (<15%) so it is not clear what the implications of these alterations are.

To get an accurate frequency estimate of mutations in the five previously reported driver genes of UM (*GNAQ*, *GNA11*, *CYSLTR2*, *SF3B1*, *BAP1* and *EIF1AX*), 133 primary UM samples which had been subjected to exome (including low coverage and hypermutated samples) and RNA sequencing were pooled together. Overall, the mutation frequencies for the reported drivers were: *GNAQ* (48.09%), *GNA11* (46.56%), *CYSLTR2* (4.58%), *SF3B1* (24.42%), *BAP1* (40.45%) and *EIF1AX* (16.03%). Recurrent mutations affecting codon 630 (D630N, D630Y) were found in the *PLCB4* gene in two samples (A985, A8KD), consistent with a previously report from another study [198]. Additionally, three novel deleterious *PLCB4* mutations (CADD score > 20) (p.G517V, p.M315V and p.V357A) were identified bringing the total frequency of *PLCB4* mutations to 3.18% **(Figure 4.9)** (**Table 4.3**). These mutations were not mutually exclusive to *GNAQ*/*GNA11* mutation, except in one case (A8KD). A novel non-Q209 hotspot mutation (p.G48V) in the *GNAQ* gene, was identified in one sample (A8KI) which did not have any other oncogenic driver **(Table 4.3)**. Two previously unreported non-Arg625 mutations (p.T663P and p.H662R) in the SF3B1 gene were identified

in samples (A9EH and A881) which did not harbour mutations in other secondary driver genes (R625- *SF3B1*, *EIFIAX* or *BAP1)*. While hotspot mutations affecting codon 209 in either a *GNAQ* or *GNA11* gene were present at an expected frequency of nearly ~90% of tumours in a mutually exclusive manner **(Figure 4.6)** and the frequency of *SF3B1* and *EIF1AX* mutations were close to those reported in the literature [198,218] , closer examination revealed four false negative cases where *BAP1* mutations had been identified by us earlier with Sanger sequencing [115,218], but which were not identified in the current analysis **(indicated in Table 4.3)**. Thus, to get a more accurate estimate of the frequency of *BAP1* mutations and other potentially missed mutations in *SF3B1* and *EIF1AX*, additional complementary analyses were performed to account for samples with these missing drivers.

First, the variant filtered bin of germline variant and low confident (spurious) variants were queried to find these missing drivers. This resulted in rescuing a BAP1 mutation in one sample (MM127). This variant was below the Mutect2 tumour LOD score (log ratio of likelihood of an event occurring in a tumour versus sequencing error). However, the mutant allele was supported by three reads (out of 115 reads). Additionally, a mutation in *EIF1AX* (p.E128) was identified by looking through a low confident mutation bin in one sample without secondary driver (A8KG). Secondly, the tier-1 mutation data generated using 4 different callers by the Broad Institute TCGA Genome Data Analysis Center (GDAC) Firehose analysis were downloaded and queried. This revealed additional *BAP1* mutations in two samples (A8K8 and A9EL) that were not picked up in the current analysis. Additionally, another novel non-625Arg *SF3B1* (p.R614S) was identified in one sample without a secondary drivers (A9EH). However, this variant was listed as a low confident variant **(Table 4.2)**. As a third approach, the ExomeSeq and RNAseq data on selected BAP1-wildtype samples were manually queried using the Integrative genome viewer [364] (version 2.3.92) visualizer tool. This helped to identify intronic deletions in two samples (AA8P and A9EQ) **(Figure 4.10 A & B)**.

Finally, examining copy number data revealed identification of a deep homozygous deletion in one sample (AA8O) that had been missed by other mutation callers **(Figure 4.11)**. This added three more *BAP1* variants not previously reported **(Table 4.3)**. Thus, the use of different mutational callers, complementary RNAseq data, copy number data and manually curating the data using visualization tools improved the detection of mutations in BAP1 and other driver genes. Germline *BAP1* mutations were identified in two samples (MM066 and A8KN). After accounting for the newly identified variants, the mutational frequency of *BAP1* increased from 40.45% to 44.27%, of *SF3B1* from 22.13% to 24.42% and

of *EIF1AX* from 16.03% to 16.79%. This also helped to identify a subset of UM samples (~20%) without any known secondary driver mutations **(Figure 4.12)**. Interestingly, *HAUS6* gene mutations were statistically significantly associated with this subgroup of tumors, where a secondary driver events had been absent (p-value 0.009). However, there was no statistical difference between the presence of *HAUS6* mutations in class 1 and class 2 tumours (P-value 0.2558).

Figure 4.7: Significant mutations identified with the MutSigCV algorithm.

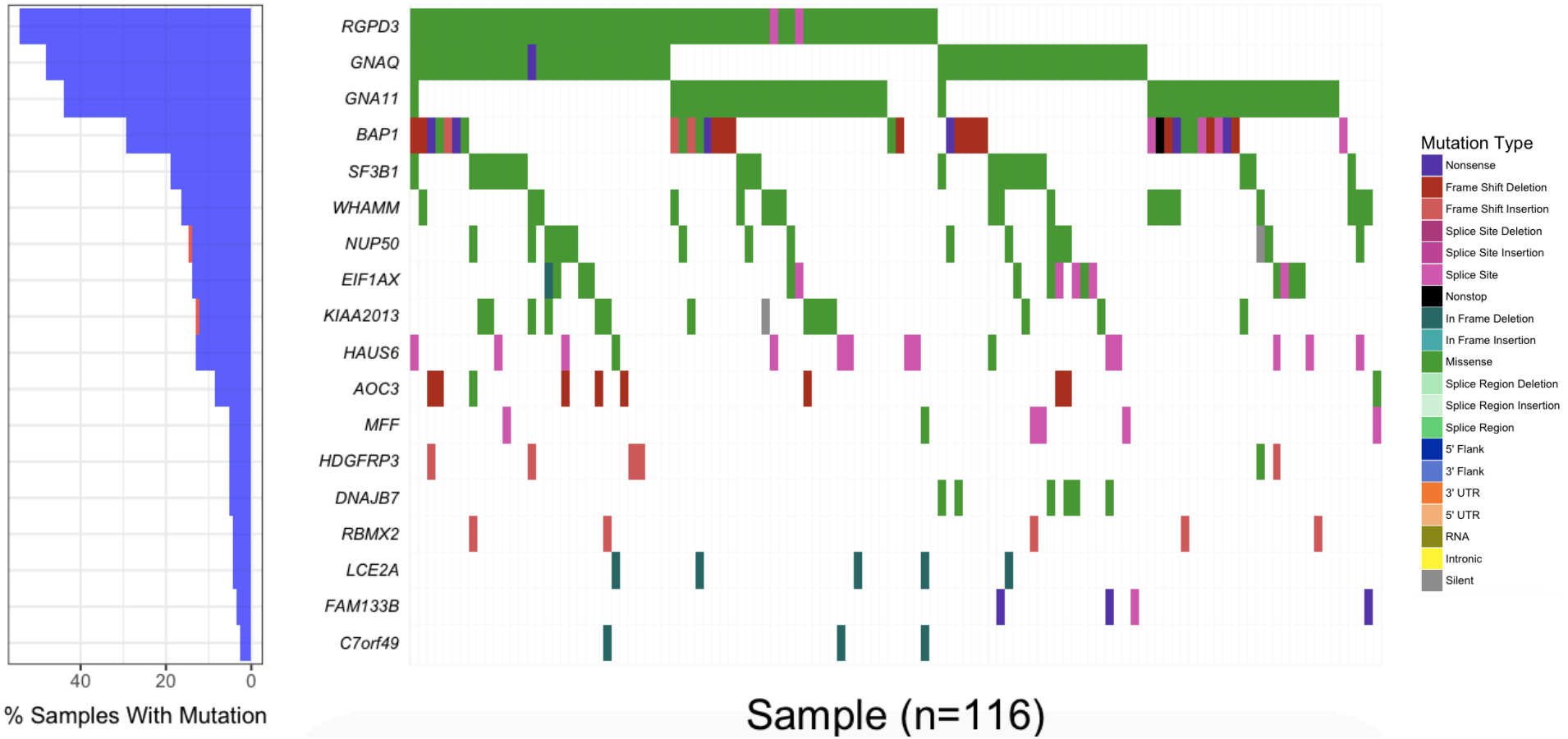


Table 4.1: MutSigCV results for the top 25 genes ranked according to p-value.

gene	nnei	x	X	p-value	FDR q value	Significant at FDR < 0.1
BAP1	40	8	4590261	0	0	Yes
GNA11	50	14	6851052	0	0	Yes
RGPD3	44	25	11971323	0	0	Yes
SF3B1	50	4	7429734	0	0	Yes
EIF1AX	50	5	6358599	2.55E-15	9.63E-12	Yes
GNAQ	20	3	2587923	4.11E-15	1.29E-11	Yes
HAUS6	50	8	7386444	1.97E-14	5.30E-11	Yes
NUP50	50	4	6570954	6.63E-10	1.56E-06	Yes
WHAMM	50	5	5707728	1.04E-09	2.18E-06	Yes
AOC3	50	7	7796529	1.37E-09	2.58E-06	Yes
LCE2A	38	6	4590027	5.41E-08	9.28E-05	Yes
MFF	50	5	6931548	9.23E-08	1.45E-04	Yes
HDGFRP3	50	17	7662447	1.63E-07	2.37E-04	Yes
RBMX2	50	5	6217146	8.23E-07	1.11E-03	Yes
KIAA2013	50	9	6147648	1.51E-06	1.90E-03	Yes
FAM133B	50	9	9042228	1.00E-05	1.18E-02	Yes
DNAJB7	50	31	13094640	2.35E-05	2.61E-02	Yes
C7orf49	50	6	4945005	5.77E-05	6.04E-02	Yes
TPTE	50	5	4364100	3.50E-04	3.23E-01	No
FGFR1OP	50	4	6118164	3.55E-04	3.23E-01	No
PLEKHF2	50	4	6176898	3.59E-04	3.23E-01	No
HNRNPA3	50	14	7005726	5.08E-04	4.23E-01	No
CYSLTR2	50	5	7726797	5.15E-04	4.23E-01	No
GATA3	50	8	6936345	6.55E-04	5.15E-01	No

Columns after gene indicate the covariate used to calculate the background mutation rate for a given gene and include, nnei, the number of neighboring genes pooled together for each gene, x, number of silent or non-coding mutated base in the neighboring genes and X, total number of bases related to these neighboring genes.

Table 4.2: Variants from novel genes identified by MutsigCV.

Gene	Variant classification	Variant type	Genome_Change	cDNA change	Codon_Change	Protein change	Cosmic variant	CADD score	Polyphen2 prediction	SIFT prediction
AOC3	Missense_Mutation	SNP	g.chr17:41003704G>C	c.344G>C	c.(343-345)aGg>aCg	p.R115T		8.491	Benign	Tolerated
AOC3	Missense_Mutation	SNP	g.chr17:41008337G>A	c.2062G>A	c.(2062-2064)Gca>Aca	p.A688T		19.14	Probably damaging	Tolerated
AOC3	In_Frame_Del	DEL	g.chr17:41003601_41003603delCCA	c.241_243delCCA	c.(241-243)ccadel	p.P81del				
AOC3	Frame_Shift_Del	DEL	g.chr17:41003602_41003603delCA	c.242_243delCA	c.(241-243)ccafs	p.P81fs				
C7orf49	In_Frame_Del	DEL	g.chr7:134851406_134851408delTCC	c.429_431delGGA	c.(427-432)gaggaa>gaa	p.143_144EE>E				
C7orf49	In_Frame_Ins	INS	g.chr7:134851408_134851409insTCC	c.428_429insGGA	c.(427-429)gag>gaGGAg	p.143_143E>EE				
DNAJB7	Start_Codon_SNP	SNP	g.chr22:41257997A>G	c.2T>C	c.(1-3)aTg>aCg	p.M1T		13.65	Probably damaging	Deleterious
DNAJB7	Missense_Mutation	SNP	g.chr22:41257836C>T	c.163G>A	c.(163-165)Gta>Ata	p.V55I		24.5	Probably damaging	Tolerated
DNAJB7	Missense_Mutation	SNP	g.chr22:41257836C>G	c.163G>C	c.(163-165)Gta>Cta	p.V55L		27.4	Probably damaging	Deleterious
FAM133B	Nonsense_Mutation	SNP	g.chr7:92195342G>A	c.643C>T	c.(643-645)Cga>Tga	p.R215*		19.59	.	Tolerated
FAM133B	Splice_Site	SNP	g.chr7:92195326G>T		c.e10+1			18.93		Deleterious
FAM134B	Missense_Mutation	SNP	g.chr5:16616978C>T	c.103G>A	c.(103-105)Gca>Aca	p.A35T		4.51	Benign	Tolerated
FAM134B	Silent	SNP	g.chr5:16616805C>T	c.276G>A	c.(274-276)ccG>ccA	p.P92P				
HAUS6	Missense_Mutation	SNP	g.chr9:19102559C>A	c.91G>T	c.(91-93)Gcc>Tcc	p.A31S		2.492	Benign	Tolerated
HAUS6	Missense_Mutation	SNP	g.chr9:19058431T>G	c.2334A>C	c.(2332-2334)gaA>gaC	p.E778D		14.52	possibly damaging	Deleterious
HAUS6	Splice_Site	SNP	g.chr9:19060087C>G	c.1764G>C	c.(1762-1764)ttG>ttC	p.L588F		16.52	possibly damaging	Deleterious
HDGFRP3	Missense_Mutation	SNP	g.chr15:83819996T>C	c.577A>G	c.(577-579)Act>Gct	p.T193A		7.809	Benign	Tolerated
HDGFRP3	Frame_Shift_Ins	INS	g.chr15:83819992_83819993insAAAAAAAA	c.580_581insTTTTTTTT	c.(580-582)tcafs	p.S194fs				
HDGFRP3	Frame_Shift_Ins	INS	g.chr15:83819992_83819993insAAAAAAA	c.580_581insTTTTTTT	c.(580-582)tcafs	p.S194fs				
KIAA2013	Missense_Mutation	SNP	g.chr1:11985396A>G	c.899T>C	c.(898-900)cTc>cCc	p.L300P		11.11	possibly damaging	Tolerated
KIAA2013	Silent	SNP	g.chr1:11983371C>A	c.1209G>T	c.(1207-1209)ggG>ggT	p.G403G				
LCE2A	In_Frame_Del	DEL	g.chr1:152671515_152671556delCAGCTCTGGGGGCTGCTGCGGCTCCAGCTCTGGGGGCTGCTG	c.138_179delCAGCTCTGGGGGCTGCTGCGGCTCCAGCTCTGGGGGCTGCTG	c.(136-180)cccagctctgggggctgctgcggctccagctctgggggctgctgc>ccc	p.SSGGCCGSSSGGCC47del	Yes			
LCE3A	In_Frame_Ins	INS	g.chr1:152595392_152595393insGGCAAC	c.187_188insGTTGCC	c.(187-189)cag>cGTTGCCag	p.62_63insRC				
LCE4A	Silent	SNP	g.chr1:152681848C>T	c.297C>T	c.(295-297)tgC>tgT	p.C99C				
MFF	Missense_Mutation	SNP	g.chr2:228197195C>T	c.242C>T	c.(241-243)cCc>cTc	p.P81L		22.5	possibly damaging	Tolerated
MFF	Splice_Site	INS	g.chr2:228197303_228197304insAATCCGAGC	c.350_351insAATCCGAGC	c.(349-354)gaaatc>gaAATCCGAGCaatc	p.118_119insRAI				
MFF	Splice_Site	INS	g.chr2:228197303_228197304insAATCCG	c.350_351insAATCCG	c.(349-354)gaaatc>gaAATCCGaatc	p.118_119insRI				
MFF	Splice_Site	INS	g.chr2:228205096_228205097insACCTGTGTTGCGTGGTGGGTCTGCTGCC	c.440_440insACCTGTGTTGCGTGGTGGGTCTGCTGCC	c.(439-441)ctt>cACCTGTGTTGCGTGGTGGGTCTGCTGCCtt	p.L147fs				
MFF	Splice_Site	INS	g.chr2:228205096_228205097insACCTGTGTTGC	c.440_440insACCTGTGTTGC	c.(439-441)ctt>cACCTGTGTTGCtt	p.L147fs				
NUP50	Missense_Mutation	SNP	g.chr22:45574145A>G	c.367A>G	c.(367-369)Acc>Gcc	p.T123A	Yes	0.204	Benign	Tolerated
NUP50	Missense_Mutation	SNP	g.chr22:45574520G>A	c.742G>A	c.(742-744)Gac>Aac	p.D248N		14.01	Benign	Tolerated
NUP50	Missense_Mutation	SNP	g.chr22:45580363C>T	c.1234C>T	c.(1234-1236)Ccc>Tcc	p.P412S		16.83	Benign	Tolerated
NUP50	Silent	SNP	g.chr22:45580380G>A	c.1251G>A	c.(1249-1251)acG>acA	p.T417T				
RBMX2	Frame_Shift_Ins	INS	g.chrX:129545492_129545493insAAAAAGGA	c.474_475insAAAAAGGA	c.(475-477)aaafs	p.-159fs				
RBMX2	Frame_Shift_Ins	INS	g.chrX:129545492_129545493insA	c.474_475insA	c.(475-477)aaafs	p.K159fs				
RGPD3	Missense_Mutation	SNP	g.chr2:107039738T>C	c.4685A>G	c.(4684-4686)aAc>aGc	p.N1562S		5.547	Benign	Tolerated
RGPD3	Missense_Mutation	SNP	g.chr2:107041612A>C	c.2811T>G	c.(2809-2811)agT>agG	p.S937R		5.022	Benign	Tolerated
RGPD3	Missense_Mutation	SNP	g.chr2:107042530G>C	c.2620C>G	c.(2620-2622)Cct>Gct	p.P874A		9.375	Benign	Tolerated
RGPD3	Splice_Site	INS	g.chr2:107042546_107042547insGA		c.e19-2					
RGPD3	Splice_Site	INS	g.chr2:107042546_107042547insAAA		c.e19-2					
WHAMM	Missense_Mutation	SNP	g.chr15:83481955T>C	c.710T>C	c.(709-711)aTg>aCg	p.M237T		0.01	Benign	Tolerated
WHAMM	Missense_Mutation	SNP	g.chr15:83499532A>G	c.1823A>G	c.(1822-1824)aAg>aGg	p.K608R		6.165	Benign	Tolerated
WHAMM	Missense_Mutation	SNP	g.chr15:83482024T>A	c.779T>A	c.(778-780)aTt>aAt	p.I260N		20.8	Probably damaging	Deleterious
WHAMM	Missense_Mutation	SNP	g.chr15:83495511C>A	c.1554C>A	c.(1552-1554)gaC>gaA	p.D518E		17.39	possibly damaging	Tolerated

Table showing variants from novel genes (excluding the known secondary drivers in UMs) that were significantly mutated genes identified from the MutsigCV analysis.

Table 4.3: UM driver genes.

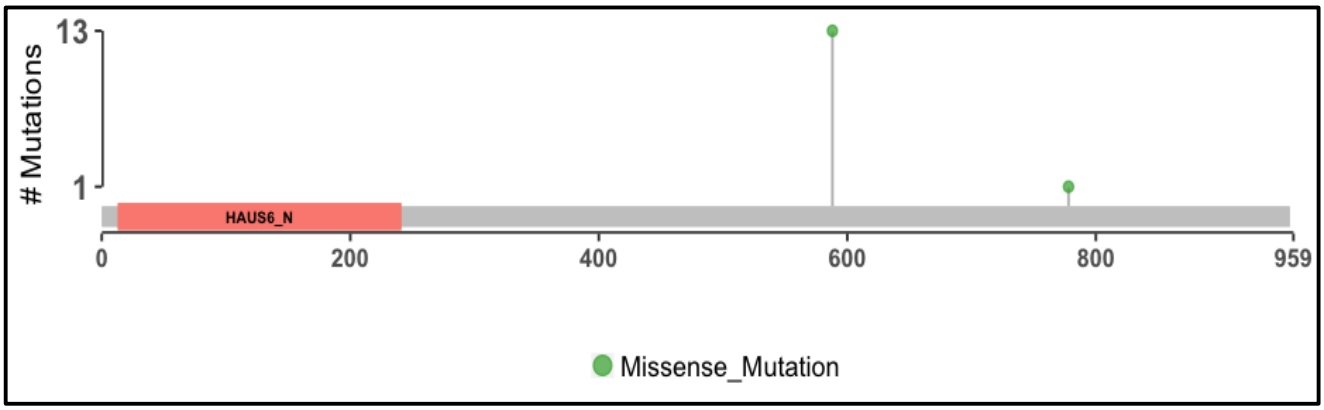
List of all mutations in all previously reported driver genes in uveal melanoma across 131 primary tumor samples subjected to Exome and RNA sequencing. Asterisk indicate where the mutations were identified from: * - was identified from Sanger sequencing, * identified from RNAseq and ** novel variants identified using complementary approaches.

Sample	Sample Source	GEP class	Data type	GNAQ	GNA11	CYSLTR2	PCLB4	BAP1	SF3B1	EIF1AX	Missing known on driver	Missing known sec driver	BAP1 identification by curr complementary analysis
MM137	WASHU	class 2	RNA sequencing		p.Q209L			c.82-121del, g.chr3:52443889-52443927delATTCATCTTCCCGCGGGGCGGCC*CCTC/					Yes
MM175	WASHU	class 2	RNA sequencing	p.Q209L				c.867_887del					No (Sanger)
A9ED	TCGA	class 2	Exome sequencing			p.L129Q		del.g.chr3:52443758.GCCTGGGTGGGGCGACAAGA GGCAGGCGCGTCCCGGGCCCATCCGGCCTCCCCAGC					Yes
A8KD	TCGA	class 2	Exome sequencing				p.D630N	del.g.chr3:52443788.ATGGTCAGGCAGGCGCGTCC CCAGCCCCTGGCCCTCCCGGTCCCC					Yes
AA8O	TCGA	class 2	Exome sequencing			p.L129Q		Homozygous deletion of exon 1					Yes (IGV-SNP array)
A9EE	TCGA	class 2	Exome sequencing	p.Q209L				p.-145fs					Yes
MM103	WASHU	class 2	Exome sequencing		p.Q209L			p.*730R					Yes
A9E7	TCGA	class 2	Exome sequencing	p.Q209L				p.146_147del					Yes
A9E8	TCGA	class 2	Exome sequencing		p.Q209L			p.485_486del					Yes
MM070	WASHU	class 2	Exome sequencing	p.Q209L				p.APSH323fs					Yes
MM100	WASHU	class 2	Exome sequencing		p.Q209L			p.ASSCR13fs					Yes
MM135	WASHU	class 2	RNA sequencing		p.R183C			p.C91W	p.R625C				Yes
A9F8	TCGA	class 2	Exome sequencing	p.Q209L				p.D68G					Yes
A9EI	TCGA	class 2	Exome sequencing		p.Q209L			p.DEFI672fs					Yes
A9ZX	TCGA	class 2	Exome sequencing		p.Q209L			p.E136fs					Yes
AA8T	TCGA	class 2	Exome sequencing		p.Q209L			p.E498fs					Yes
A9EO	TCGA	class 2	Exome sequencing	p.Q209L				p.E685V					Yes
MM179	WASHU	class 2	Exome sequencing		p.Q209L			p.E7*					Yes
MM151	WASHU	class 2	Exome sequencing			p.L129Q		p.EGP194del					Yes
MM116	WASHU	class 2	Exome sequencing	p.Q209L				p.F118fs					Yes
MM141	WASHU	class 1	RNA sequencing		p.Q209L			p.F50fs					Yes
MM144	WASHU	class 2	RNA sequencing		p.Q209L			p.F50Lfs*22	p.R625H				No (Sanger)
AB0B	TCGA	class 2	Exome sequencing		p.Q209L			p.F81fs					Yes
MM121	WASHU	class 2	Exome sequencing			p.L129Q		p.G128R					Yes
AA9A	TCGA	class 2	Exome sequencing		p.Q209L			p.G185R					Yes
MM066	WASHU	class 2	Exome sequencing		p.Q209L			p.H169Q					Yes
MM055	WASHU	class 2	Exome sequencing		p.Q209L			p.H169Q					No (Sanger)
MM127	**WASHU**	**class 2**	**Exome sequencing**		**p.Q209L**			**p.K25N**					**Yes**
A984	TCGA	class 2	Exome sequencing					p.K421fs			Yes		Yes
A8KN	TCGA	class 2	Exome sequencing	p.Q209L				p.K453fs					Yes
A8K8	TCGA	class 2	Exome sequencing		p.Q209L			p.K61E					No (GDAC pipeline)
MM046	WASHU	class 2	Exome sequencing	p.Q209L				p.KC637fs					Yes
MM120	WASHU	class 2	Exome sequencing		p.Q209L			p.L49fs					Yes
A9EV	TCGA	class 2	Exome sequencing		p.Q209L			p.LMAVVPDRRIKYEA230fs					Yes
AA8N	TCGA	class 2	Exome sequencing		p.Q209L			p.N78S					Yes
A8KK	TCGA	class 2	Exome sequencing	p.Q209L				p.P124fs					Yes
A980	TCGA	class 2	Exome sequencing		p.Q209L			p.P339fs					Yes
A9EL	TCGA	class 2	Exome sequencing	p.Q209L				p.P391Afs*5					No (GDAC pipeline)
MM091	WASHU	class 2	RNA sequencing		p.Q209P			p.Q253fs					Yes

A9EX	TCGA	class 2	Exome sequencing	p.Q209L				p.Q40*					Yes
A9EU	TCGA	class 2	Exome sequencing		p.Q209L			p.Q441*					Yes
A9F0	TCGA	class 2	Exome sequencing	p.Q209L				p.Q456fs					Yes
MM173	WASHU	class 2	RNA sequencing		p.Q209L			p.R146M*					No (Sanger)
MM161	WASHU	class 2	Exome sequencing		p.Q209L			p.R300fs					Yes
A88A	TCGA	class 2	Exome sequencing		p.Q209L			p.RRSRRKVSTLV56fs					Yes
A9F3	TCGA	class 2	Exome sequencing		p.Q209L			p.S98Rfs*28					Yes
A985	TCGA	class 2	Exome sequencing	p.R183Q	p.R183C		p.D630N	p.SREKTGMVRPG609fs	p.R625C				Yes
A8KL	TCGA	class 2	Exome sequencing		p.Q209L			p.V27fs					Yes
MM056	WASHU	class 2	Exome sequencing	p.Q209L				p.W196*					Yes
A8KF	TCGA	class 2	Exome sequencing		p.Q209L			p.W196*					Yes
A8KI	TCGA	class 2	Exome sequencing	p.G48V**				p.X571_splice					Yes
A888	TCGA	class 2	Exome sequencing		p.Q209L			p.Y173C					Yes
A9EF	TCGA	class 2	Exome sequencing		p.Q209L		p.V357A	p.Y223*					Yes
A9F1	TCGA	class 2	Exome sequencing	p.Q209L				p.Y33*					Yes
AA8P	TCGA	class 2	Exome sequencing		p.Q209L			Putative intronic deletion (exon 4-5)					Yes (IGV-DNAseq)
A9EQ	TCGA	class 2	Exome sequencing		p.Q209L			Putative intronic deletion (exon 7-8)					Yes (IGV-RNAseq)
MM162	WASHU	class 2	Exome sequencing		p.Q209L			splice_c.e13+1					Yes
MM098	WASHU	class 2	Exome sequencing		p.Q209L			splice_c.e16+1					Yes
A881	TCGA	class 1	Exome sequencing		p.Q209L				p.H662R**				Yes
A9E5	TCGA	class 1	Exome sequencing	p.Q209L					p.K666T				Yes
MM089	WASHU	class 1	Exome sequencing	p.Q209L					p.K666T*				Yes
A9EH	TCGA	class 1	Exome sequencing		p.Q209L				p.R614S**				No (GDAC pipeline)
MM134	WASHU	class 2	Exome sequencing	p.Q209L	p.Q209L				p.R625C				Yes
A9EA	TCGA	class 1	Exome sequencing	p.Q209L					p.R625C				Yes
A9EW	TCGA	class 1	Exome sequencing	p.Q209L					p.R625C				Yes
A8K9	TCGA	class 1	Exome sequencing	p.Q209L					p.R625C				Yes
A8KB	TCGA	class 1	Exome sequencing	p.Q209L					p.R625C				Yes
MM065	WASHU	class 1	Exome sequencing	p.Q209L					p.R625C				Yes
MM133	WASHU	class 2	Exome sequencing	p.Q209L					p.R625C				Yes
MM032	WASHU	class 1	Exome sequencing	p.Q209R					p.R625C				Yes
A8KA	TCGA	class 1	Exome sequencing		p.Q209L				p.R625C				Yes
MM010	WASHU	class 1	Exome sequencing			p.L129Q			p.R625C				Yes
MM176	WASHU	class 1	RNA sequencing						p.R625C				Yes
MM064	WASHU	class 1	RNA sequencing	p.Q209L					p.R625H				Yes
MM028	WASHU	class 1	Exome sequencing	p.Q209L*					p.R625H*				No
MM131	WASHU	class 1	Exome sequencing		p.Q209L				p.R625L	p.G9D*			Yes
A9E9	TCGA	class 1	Exome sequencing	p.Q209L					p.R625L				Yes
A9EZ	TCGA	class 1	Exome sequencing	p.Q209L					p.R625L				Yes
AA9E	TCGA	class 1	Exome sequencing	p.Q209L					p.R625L				Yes

A9F4	TCGA	class 1	Exome sequencing	p.Q209L					p.R625L				Yes
A885	TCGA	class 1	Exome sequencing	p.Q209L					p.R625L				Yes
MM101	WASHU	class 1	Exome sequencing	p.Q209L					p.R625L				Yes
A9ES	TCGA	class 2	Exome sequencing	p.R183Q					p.R625L				Yes
A9EJ	TCGA	class 1	Exome sequencing		p.Q209L				p.R625L				Yes
A8KH	TCGA	class 1	Exome sequencing		p.Q209L				p.R625L				Yes
MM049	WASHU	class 1	Exome sequencing		p.Q209L				p.R625L				Yes
A9EC	TCGA	class 1	Exome sequencing	p.Q209L					p.T665P	p.G9V			Yes
A8KO	TCGA	class 1	Exome sequencing	p.Q209L						p.7_8KG>R			Yes
A8KG	**TCGA**	**class 1**	**Exome sequencing**		**p.Q209L**					p.E128			**Yes**
A87U	TCGA	class 1	Exome sequencing	p.Q209L						p.G15D			Yes
MM132	WASHU	class 1	RNA sequencing	p.Q209L						p.G15D			Yes
A9F7	TCGA	class 1	Exome sequencing		p.Q209L					p.G15D			Yes
MM086	WASHU	class 1	Exome sequencing	p.Q209L			p.G517V			p.G6D			Yes
A880	TCGA	class 1	Exome sequencing	p.Q209L						p.G6D			Yes
AA8R	TCGA	class 1	Exome sequencing		p.Q209L					p.G6D			Yes
A9EY	TCGA	class 1	Exome sequencing	p.Q209L						p.G8R			Yes
A884	TCGA	class 1	Exome sequencing		p.Q209L					p.G8R			Yes
MM159	WASHU	class 1	Exome sequencing		p.Q209L					p.G9A			Yes
A8KE	TCGA	class 1	Exome sequencing	p.Q209L						p.G9R			Yes
MM082	WASHU	class 1	RNA sequencing	p.Q209L						p.G9V			Yes
MM078	WASHU	class 2	Exome sequencing	p.Q209L						p.G9V			Yes
MM050	WASHU	class 1	Exome sequencing	p.Q209L						p.K3E			Yes
MM094	WASHU	class 2	RNA sequencing		p.Q209L					p.K7R			Yes
MM018	WASHU	class 1	Exome sequencing	p.Q209P						p.N4Y			No
A9ZY	TCGA	class 1	Exome sequencing		p.Q209L					p.W70R			Yes
MM038	WASHU	class 1	Exome sequencing	p.Q209L						spice_c.e2-1			Yes
MM105	WASHU	class 1	Exome sequencing		p.Q209L					splice_c.e2-2			Yes
A8K7	TCGA	class 1	Exome sequencing	p.Q209L								Yes	No
A983	TCGA	class 1	Exome sequencing	p.Q209L								Yes	No
AA8Q	TCGA	class 1	Exome sequencing	p.Q209L								Yes	No
AA8S	TCGA	class 1	Exome sequencing	p.Q209L								Yes	No
A87T	TCGA	class 1	Exome sequencing	p.Q209L								Yes	No
AA8M	TCGA	class 1	Exome sequencing	p.Q209L								Yes	No
A9F2	TCGA	class 1	Exome sequencing	p.Q209L								Yes	No
A9EK	TCGA	class 1	Exome sequencing	p.Q209L								Yes	No
A9EM	TCGA	class 1	Exome sequencing	p.Q209L								Yes	No
MM016	**WASHU**	**class 1**	**Exome sequencing**	**p.Q209L**								**Yes**	No
MM171	**WASHU**	**class 1**	**RNA sequencing**	**p.Q209L**								**Yes**	No
A8KM	TCGA	class 2	Exome sequencing	p.Q209L								Yes	No
A883	TCGA	class 2	Exome sequencing	p.Q209L								Yes	No
MM080	**WASHU**	**class 2**	**Exome sequencing**	**p.Q209L**								**Yes**	No
MM150	**WASHU**	**class 2**	**Exome sequencing**	**p.Q209L**								**Yes**	No
A9F5	TCGA	class 2	Exome sequencing		p.Q209L							Yes	No
A882	TCGA	class 2	Exome sequencing		p.Q209L							Yes	No
MM113	**WASHU**	**class 1**	**Exome sequencing**		**p.Q209L**		**p.M315V**					**Yes**	No
A9ET	TCGA	class 1	Exome sequencing		p.Q209L							Yes	No
A87W	TCGA	class 1	Exome sequencing		p.Q209L							Yes	No
MM085	**WASHU**	**class 1**	**Exome sequencing**		**p.Q209L**							**Yes**	No
MM048	**WASHU**	**class 1**	**Exome sequencing**		**p.Q209L**							**Yes**	No
A87Y	TCGA	class 2	Exome sequencing		p.Q209L							Yes	No
MM138	**WASHU**	**class 2**	**Exome sequencing**		**p.R183C**							**Yes**	No
A982	TCGA	class 1	Exome sequencing			p.L129Q						Yes	No
A8KJ	TCGA	class 1	Exome sequencing								Yes	Yes	No

Figure 4.8: Mutations in the HAUS6 gene.



Hotspot mutations identified in the *HAUS6* gene in ~12% of uveal melanomas. These mutations (p.L553F & p.E743D) do not lie within the HAUS6_N terminal domain of the protein.

Figure 4.9: Reported missense mutations identified in the PLCB4 gene

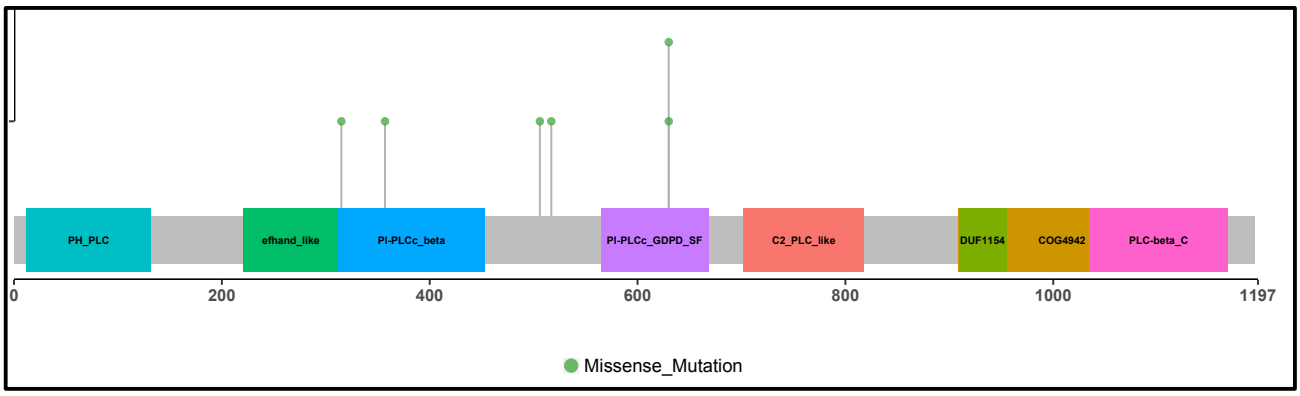
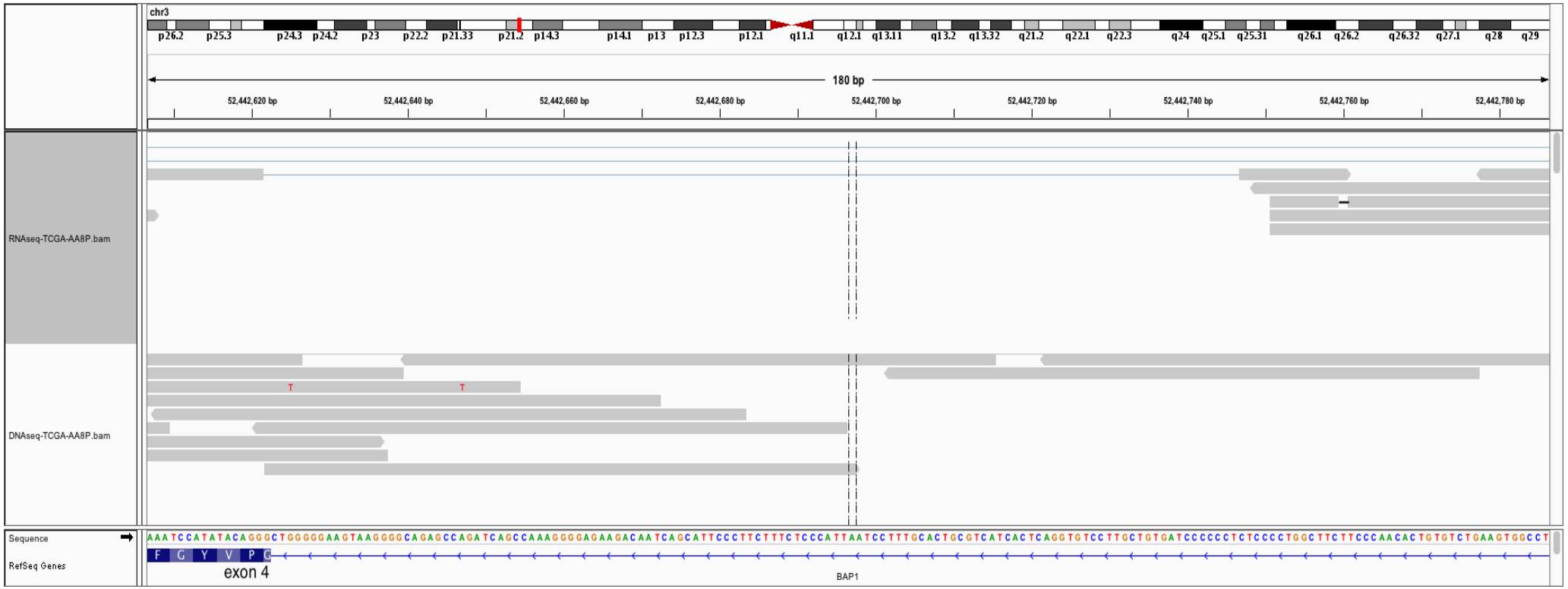


Figure 4.10 (A & B): IGV images of BAP1 mutations.



(A) BAP1 mutations identified by manually curating the RNASeq and Exome BAM files in IGV. (A) A splice deletion mutation identified between exon 4 and exon 5 in sample AA8P

(B) Intronic BAP1 deletion identified between exon 7 and exon 8 in sample A9EQ using the exome BAM file.

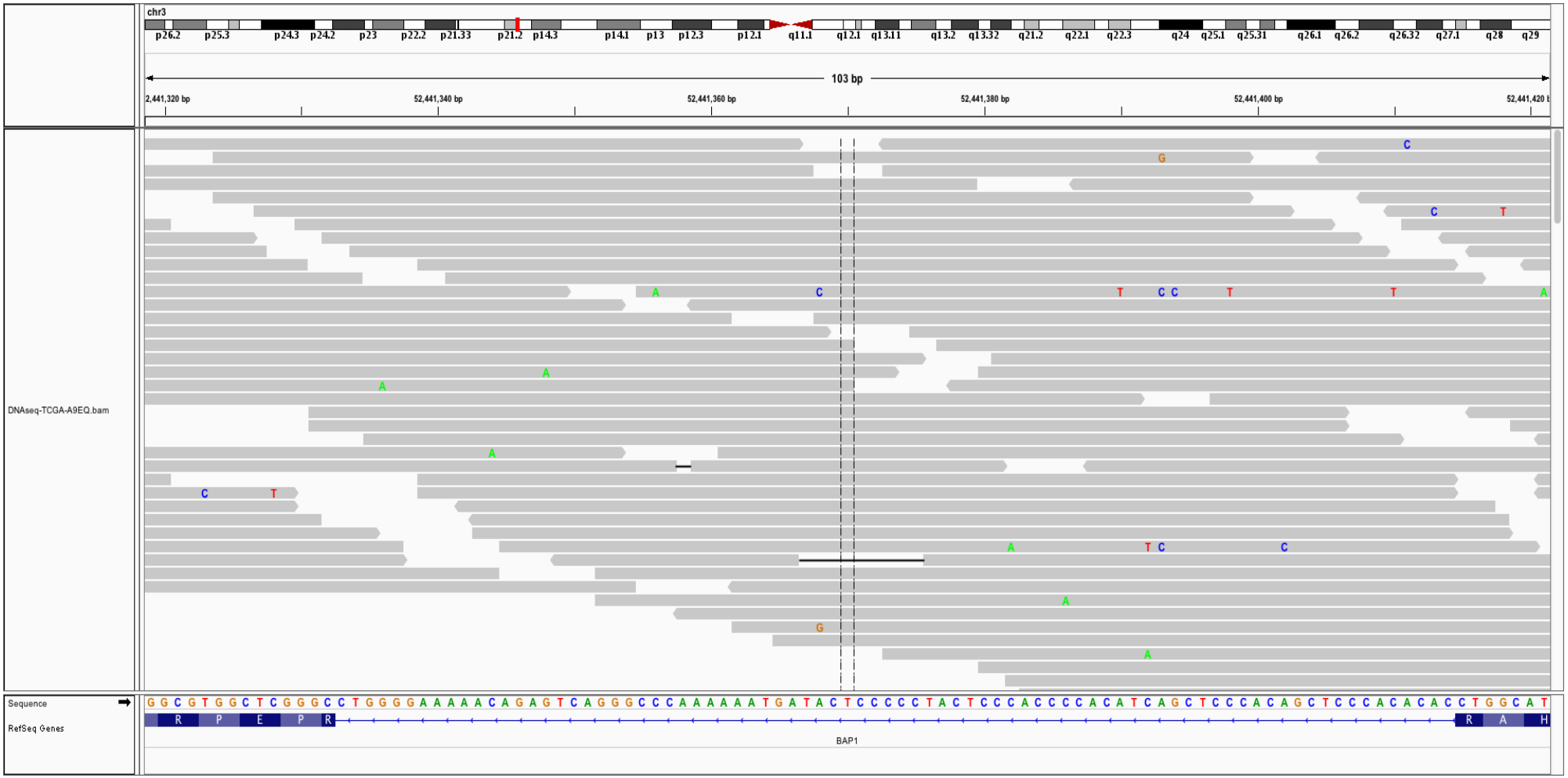
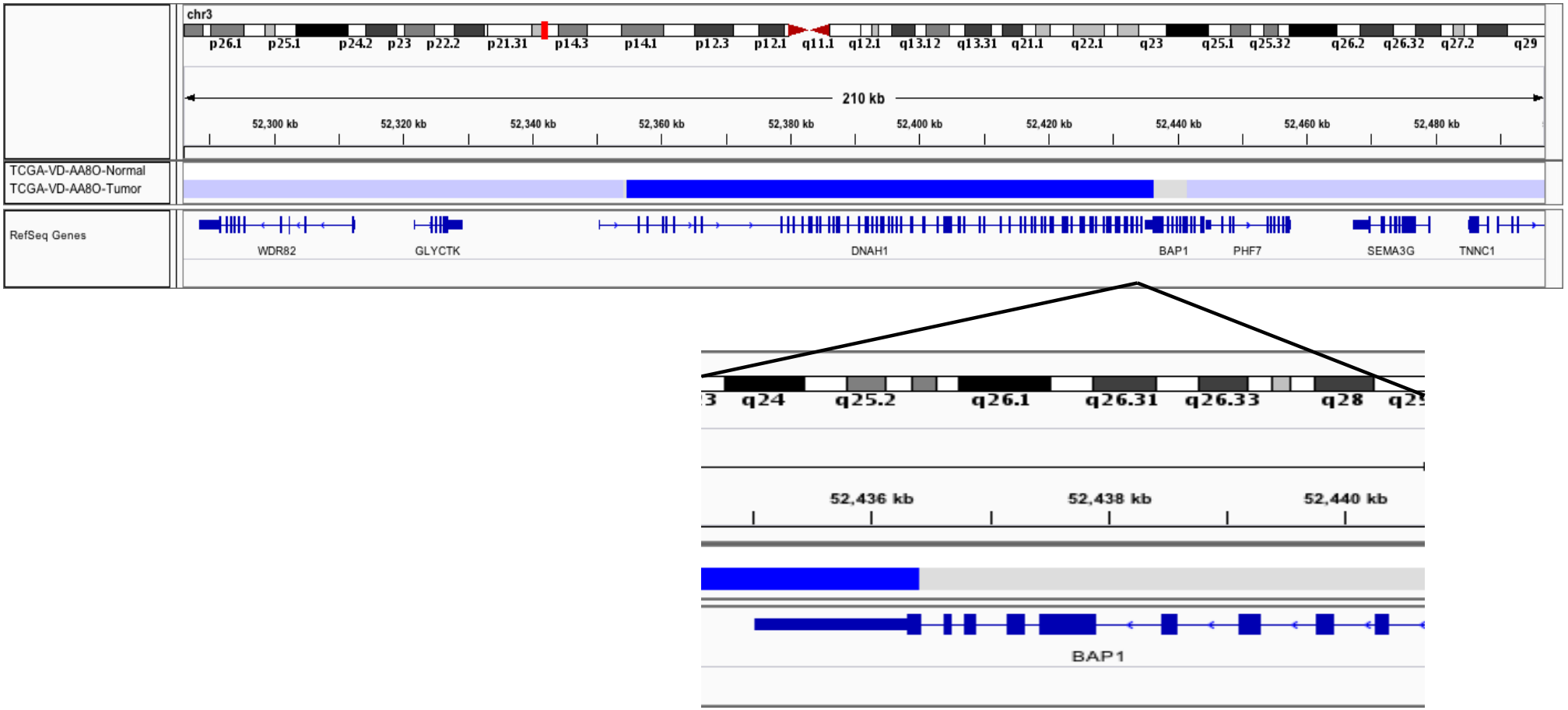
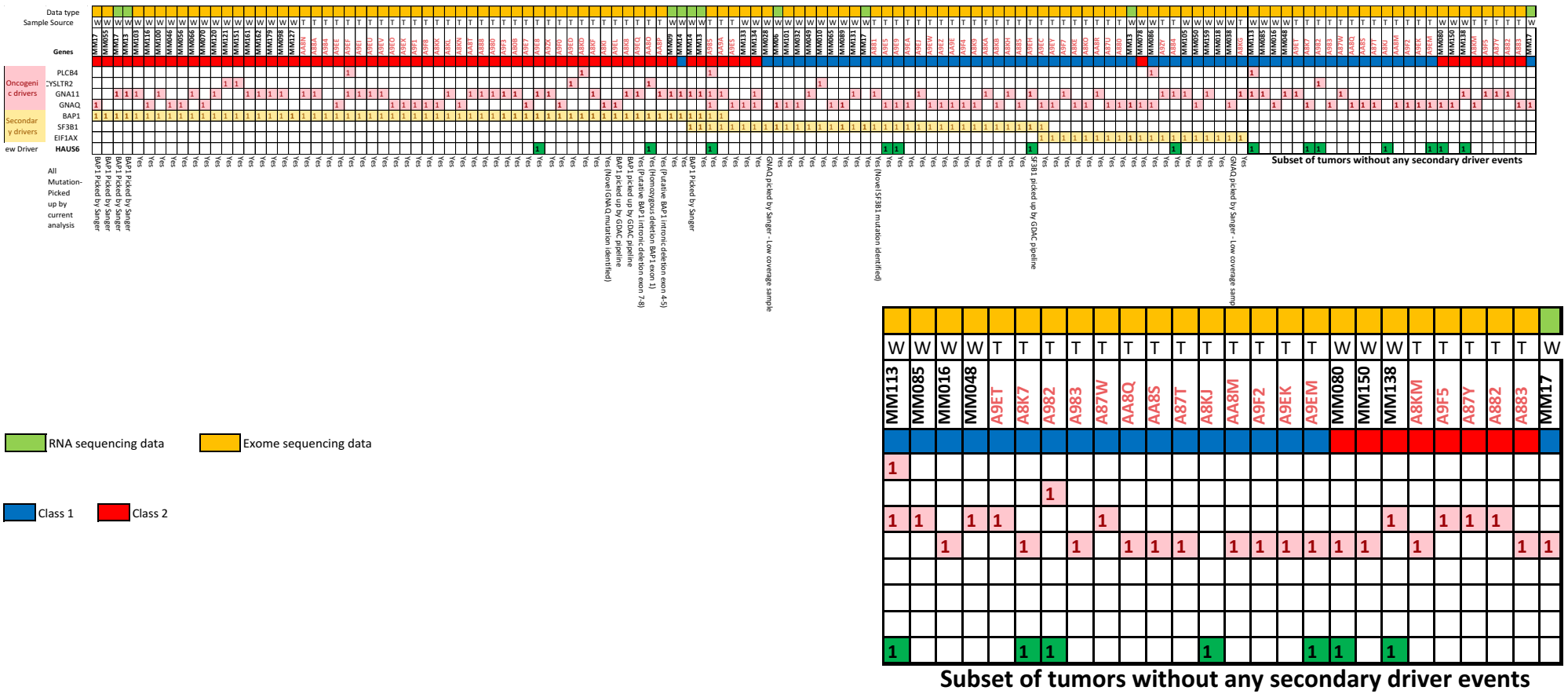


Figure 4.11: IGV images of PHF10 homozygous deletion



Screenshot from IGV showing deep homozygous deletion was identified in the exon 1 of the BAP1 gene using copy number segmentation data. The normal sample showed no copy number change.

Figure 4.12: Summary of UM driver mutations.



All the UM driver mutations identified across 131 primary uveal melanomas. The mutations identified are identified by the current analysis and other sources are indicated below. A subset of tumors do not harbor any mutation in the known secondary driver (BAP1, SF3B1 and EIF1AX) is zoomed bel

Table 4.4: RNA splicing and processing pathways gene mutations

Gene	Cytoband	Sample	hg19 coodinates	Variant_Classification	Protein_Change	CADD score	Matching SNP array	Copy number Alteration
DNAJC8	1p35	A9EY	chr1:28555506-28555506	Missense_Mutation	p.V36A	18.46	Yes	No
EIF4G3	1p36.12	MM049	chr1:21188801-21188801	Missense_Mutation	p.Q961K	28.6	No	
HNRNPR	1p36.12	MM049	chr1:23648134-23648134	Missense_Mutation	p.P233H	26.5	No	
HNRNPR	1p36.12	MM049	chr1:23648135-23648135	Missense_Mutation	p.P233T	16.19	No	
RBM10	Xp11.3	A9EC	chrX:47039307-47039312	In_Frame_Del	p.ST311del		Yes	No (Female)
RBM10	Xp11.3	A87W	chrX:47039363-47039393	Frame_Shift_Del	p.SNVRVIKDKQT329fs		Yes	No (Male)
SF3A1	22q12.2	MM010	chr22:30737701-30737701	Missense_Mutation	p.Q351K	23.7	Yes	No
SF3A1	22q12.2	MM049	chr22:30737821-30737821	Missense_Mutation	p.Q311K	24.3	No	
SF3A2	19p13.3	A985	chr19:2243426-2243426	Frame_Shift_Del	p.F3fs		Yes	No
SF3A2	19p13.3	A9EF	chr19:2247812-2247813	Frame_Shift_Ins	p.S222fs		Yes	No
SNRNP40	1p35.2	A9ZX	chr1:31769472-31769472	Missense_Mutation	p.A43T	16.1	Yes	No
SRRM1	1p36	A9EQ	chr1:24987276-24987276	Missense_Mutation	p.E461G	25.4	Yes	No
SRRM1	1p36	MM049	chr1:24996760-24996760	Missense_Mutation	p.P785Q	20.4	No	
SRRM1	1p36	MM049	chr1:24996760-24996760	Missense_Mutation	p.P785L	20.5	No	
SRRM1	1p36	MM055	chr1:24996760-24996760	Missense_Mutation	p.P785Q	20.4	No	
SRRM1	1p36	MM055	chr1:24996760-24996760	Missense_Mutation	p.P785L	20.5	No	
SRSF2	17q25.2	A888	chr17:74732373-74732390	In_Frame_Del	p.173_179SSSVSRS>S		Yes	No
SRSF5	14q24	MM010	chr14:70235391-70235391	Splice_Site	p.R63M	22.1	Yes	No
SRSF5	14q24	MM101	chr14:70235572-70235572	Nonsense_Mutation	p.Y82*	18.44	No	
SRSF7	2p22.1	MM049	chr2:38977329-38977329	Missense_Mutation	p.K12N	17.5	No	
SRSF9	12q24.31	MM010	chr12:120901872-120901872	Missense_Mutation	p.G135W	22.2	Yes	No

Mutations targeting RNA splicing and processing pathways identified from exome analysis of 131 primary uveal melanomas.

Table 4.5: Chromatin remodeling complex gene mutations.

Chromatin remodeling complex	Gene	Coordinates	Mutation	Sample	Mutation type	CADD score
actin-like 6A	ACTL6A	chr3:179294664-179294664	p.P244H	MM055	Missense_Mutation	28.3
actin-like 6A	ACTL6A	chr3:179305718-179305718	p.G404C	MM049	Splice_Site	24.5
AT rich interactive domain 1A (SWI-like)	ARID1A	chr1:27056330-27056331	p.G443fs	MM049	Inframe insertion	
AT rich interactive domain 1A (SWI-like)	ARID1A	chr1:27087515-27087515	p.P697T	MM010	Missense_Mutation	32
AT rich interactive domain 1A (SWI-like)	ARID1A	chr16:27087885-27087885	p.M724I	MM176	Missense_Mutation	17.87
AT rich interactive domain 1B (SWI1-like)	ARID1B	chr6:157222625-157222625	p.P631Q	A8K9	Missense_Mutation	13.16
AT rich interactive domain 1B (SWI1-like)	ARID1B	chr6:157488218-157488218	p.S975F	MM105	Missense_Mutation	27.7
AT rich interactive domain 1B (SWI1-like)	ARID1B	chr6:157522220-157522220	p.D1498Y	MM049	Missense_Mutation	18.09
AT rich interactive domain 1B (SWI1-like)	ARID1B	chr6:157528337-157528337	p.R2021M	MM010	Missense_Mutation	18.46
AT rich interactive domain 2 (ARID, RFX-like)	ARID2	chr12:46205245-46205245	p.D110V	A8K7	Missense_Mutation	23.1|23.1
AT rich interactive domain 2 (ARID, RFX-like)	ARID2	chr12:46230757-46230757	p.G336*	MM049	Nonsense_Mutation	40
B-cell CLL/lymphoma 7A	BCL7A	chr12:122481864-122481864	p.P115L	MM127	Missense_Mutation	28.8
B-cell CLL/lymphoma 7C	BCL7C	chr16:30904269-30904269	p.E58*	MM049	Splice_Site	44
bromodomain containing 7	BRD7	chr16:50357574-50357574	p.P456Q	MM049	Missense_Mutation	18.45
bromodomain containing 7	BRD7	chr16:50359708-50359708	p.L428I	MM049	Missense_Mutation	22.1
bromodomain containing 7	BRD7	chr16:50388781-50388781	p.D104V	A9ZY	Missense_Mutation	20.2
bromodomain containing 7	BRD7	chr16:50388781-50388781	p.D104V	A9EH	Missense_Mutation	20.2
bromodomain containing 7	BRD7	chr16:50388781-50388781	p.D104V	AA8N	Missense_Mutation	20.2
bromodomain containing 7	BRD7	chr16:50388781-50388781	p.D104V	A88A	Missense_Mutation	20.2
bromodomain containing 7	BRD7	chr16:50388781-50388781	p.D104V	A984	Missense_Mutation	20.2
bromodomain containing 7	BRD7	chr16:50359707-50359707	p.L428*	MM049	Nonsense_Mutation	39
PHD finger protein 10	PHF10	chr6:170115819-170115819	p.F226fs	MM133	Frameshift deletion	
SWI/SNF related	SMARCA2	chr9:2081837-2081837	p.Q730H	A884	Missense_Mutation	16.11
SWI/SNF related	SMARCA2	chr9:2123881-2123881	p.R1309S	MM049	Missense_Mutation	23.2
SWI/SNF related	SMARCA2	chr9:2182199-2182199	p.L1455H	MM050	Missense_Mutation	17.06
SWI/SNF related	SMARCA4	chr19:11152089-11152089	p.T1396I	MM032	Missense_Mutation	23.3
SWI/SNF related	SMARCA4	chr19:11152169-11152169	p.P1423T	MM010	Missense_Mutation	23.7
SWI/SNF related	SMARCA4	chr19:11141465-11141465	p.E1148X	MM098	stopgain	41
SWI/SNF related	SMARCA4	chr19:11132513-11132513	p.T910M	A985	Missense_Mutation	16.22
SWI/SNF related	SMARCC1	chr3:47651677-47651678	p.974_974S>SQQAHQHS	A984	Inframe insertion	
SWI/SNF related	SMARCC2	chr12:56575537-56575537	p.N264T	MM105	Missense_Mutation	12.87
SWI/SNF related	SMARCD1	chr12:50492504-50492504	p.T467S	A9ET	Missense_Mutation	22.5
SWI/SNF related	SMARCD3	chr7:150945590-150945590	p.E20A	A881	Missense_Mutation	15.07

Mutations targeting genes which are components of the chromatin remodeling complex identified from an analysis of exome or RNA-data from 131 primary uveal melanomas

4.1.3 Integration of copy number and exome data

GISTIC analysis refined the recurrent broad arm level alteration and defined significant common regions of gains and deletions that harbour multiple candidate cancer genes such as *PLEC1* (8q24.3 amplification), *ARID1A* (1p35.3 deletion) and *NCAPD3* (11q25 deletion). Alterations involving chromosomal loss of 3p, 3q, 6q, 8p, 11p and gain of 8q are significantly associated with metastasis and other alterations associated with PRAME positive tumours (independent marker of metastasis) include gain of 1q and loss of 6q and 16q [46]. Interrogation of amplified and deleted focal peaks on these chromosomes for candidate genes, revealed significant enrichment of cis signature genes associated with MAPK, WNT signalling, NF-kappa B and RNA processing and splicing pathways (described in Chapter 1). Genes from these regions were then queried for mutations identified through exome sequencing in an attempt to identify novel drivers.

Protein damaging mutations affecting *PLEC1* were found in four samples (MM049, MM089, A982 and A983). To further refine the mechanism of tumorigeneisis in UM, mutations in genes altered by copy number changes and other related genes in the same pathway were examined. Protein damaging genes that mapped to genes underlying these pathways and highly recurrent copy number altered regions (identified in chapter 3) were examined. This lead to the identification of additional novel mutations relating to RNA processing pathways including, mutations in *SNRNP40*, *SRRM1*, *DNAJC8*, *EIF4G3*, *HNRNPR*, *SRSF4*, *SRSF5*, *SYF2*, *SF3A1*, *SRSF7* and *RBM10* genes **(Figure 4.13)**. These mutations were present in tumours with secondary drivers and none of the samples harbouring mutations exhibited changes in copy number **(Table 4.4)**.

Following up on findings from copy number analysis, and using a conservative criterion that at least one of the genomic aberrations should be unequivocally focal (homozygous deletion), the human SWI/SNF genes, *ARID1A* (1p36.11) and *ARID1B* (6q24.3), identified from the GISTIC analysis (results - chapter 1) were investigated for additional mutations. *ARID1A* and *ARID1B* were deleted homozygously in 2/182 and 8/182 of samples interrogated on a global SNP array. Additionally, *PHF10*, part of the PBAF complex was also found to be deleted homozygously in 8/182 of SNP arrayed samples. Mining the exome sequencing data on 120 primary UMs to identify additional mutations

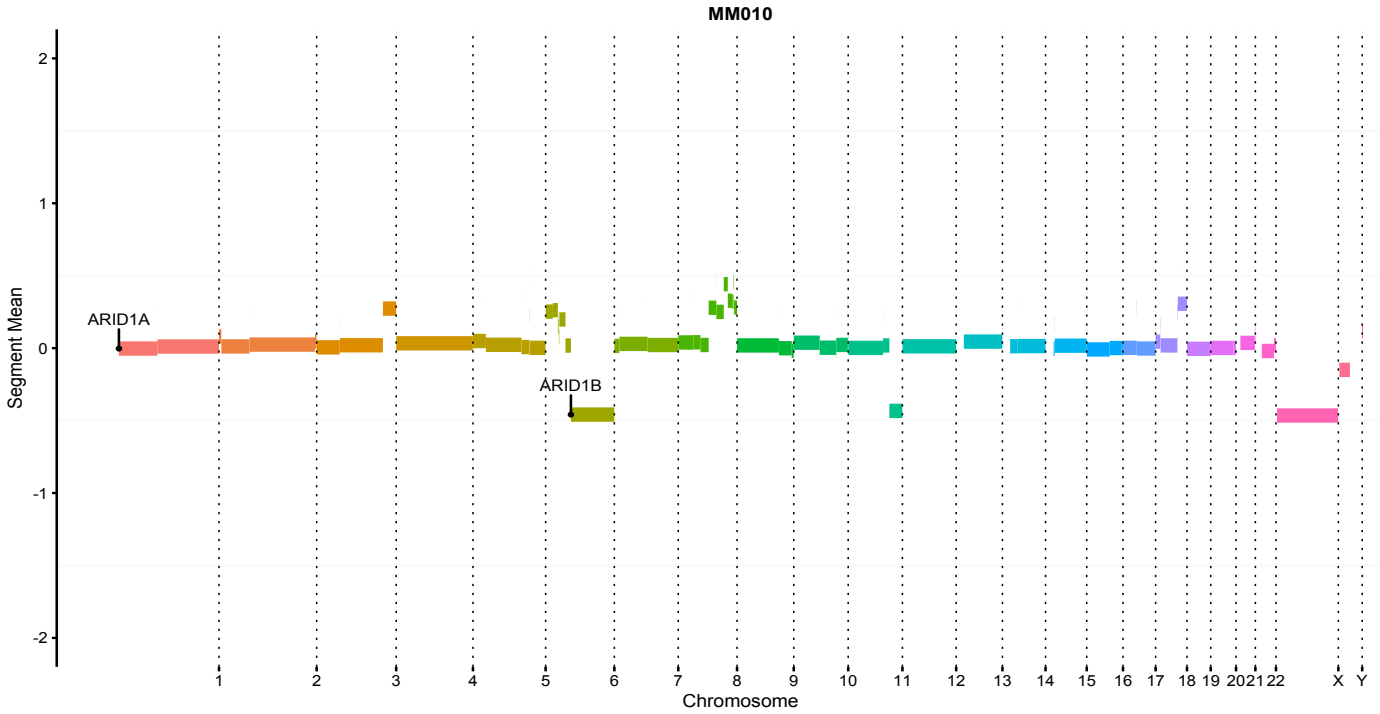
Figure 4.13: Mutations targeting RNA splicing and processing machinery in UM.

	A9EY	A9EC	MM010	MM101	MM134	MM049	A985	A9EF	MM055	A9ZX	A9EQ	A888	A87W	MM080	Pathway
BAP1							■	■	■	■	■	■			
SF3B1	■	■													
EIF1AX		■	■	■	■	■	■								
SNRNP40										1					mRNA splicing
SRRM1					1	1			2		1				pre- & post- splicing
DNAJC8	1														mRNA splicing
EIF4G3						1									mRNA cap recognition
HNRNPR						2								1	pre-mRNA processing
SRSF5			1	1											mRNA splicing
SF3A1			1			1									mRNA splicing
SF3A2							1	1							mRNA splicing
SRSF7						1									mRNA splicing
SRSF2												1			mRNA splicing
RBM10		1											1		mRNA splicing

1 Mutation with no copy number change (number of mutation)

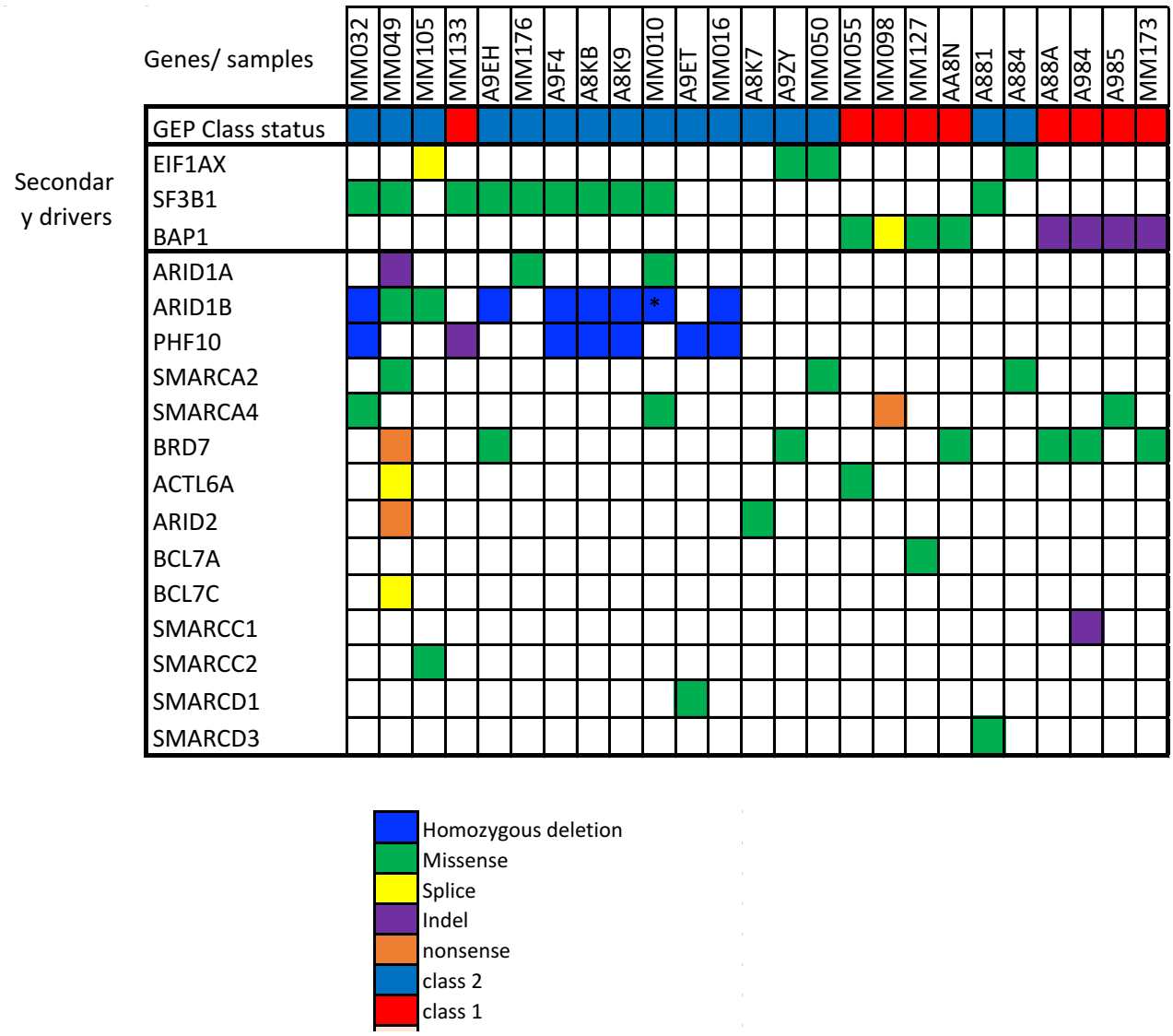
Figure showing all the Protein damaging mutations targeting RNA splicing and processing machinery in UM

Figure 4.14: **Whole genome copy number segmented profile for MM010**.



The mutations in ARID1B located in the region of shallow deletion., is indicated by the black dot.

Figure 4.15: Tumours with mutations in components of the mammalian chromatin remodeling BAF/PBAF complex.



In total 131 primary uveal melanomas, originally interrogated were following by integration of exome, RNAseq and copy number data affecting these three genes, revealed six samples with DNA mutations (six missense mutations and two INDELs) **(Table 4.5)**. One of the samples with a homozygous deletion in *ARID1B* (MM010) also harboured a mutation **(Figure 4.14)** indicating presence of potential tumour heterogeneity. Using RNA-seq data, one additional mutation in *ARID1B* (p.M724I) in MM176) was identified. Based on the combined mutation and copy number data, 9% of samples harbour a deleterious mutations or homozygous deletion in *ARID1A*, *ARID1B* or *PHF10* genes **(Figure 4.15)**. These mutations were nearly always present in tumours with a secondary driver mutations (*BAP1*, *SF3B1* and

EIF1AX) and there was no statistical association between the ARID1A/B/PHF10 mutants and the secondary drivers (p-value 0.059). However, when class 1 and class 2 tumors, were compared there was statistical evidence for the ARID1A/B/PHF10 mutants to be associated with class 1 tumors (p-value 0.01). In addition to *ARID1A*, *ARID1B*, other enzymatic subunits of the SWI/SNF complex were mined in the exome sequencing data. This resulted in the identification of 25 deleterious mutations encompassing the subunits of the PBAF and BAF complex [342], including: the enzymatic ATPase subunits, *SMARCA2* and *SMARCA4;* and other components *ACTL6A, ARID2, BCL7A/C, BRD7, SMARCC1, SMARCC2, SMARCD1*and *SMARCD3* **(Table 4.6).** A recurrent mutation in BRD7 (p. D104V) was identified in six samples, however it was not found to be significantly mutated by the MutsigCV analysis, and was found exclusively in the TCGA cohort, thus weaking confidence in this variant. Although, other components of the complex carried a small burden of mutation, determining the significance of these low frequency aberrations will require a larger sample size.

4.2 Discussion

Here a comprehensive analysis of whole exome sequencing data from of 120 uveal melanomas from two different cohorts (WASH-U and TCGA) is described. The mutational signature analysis revealed the absence of a UV radiation induced damage related signature and DNA break strand repair or BRCA signature, contrary to a previous report [198]. 19,467 somatic SNV and INDEL mutations were identified and 10,014 genes were mutated in at least one of the tumours. This number could be relatively large because 40% of the samples from the WASH-U cohort did not have matched normal samples (there were 16 unpaired tumours), and although in the current pipeline, Mutect2 caller was run with "panel of normal option", there could be many potential private germline mutations that were not filtered. A frequency based statistical method (the MutSigCV algorithm) was applied to this initial list of variants to identify significantly mutated genes (SMGs) which were mutated at a much higher rate than the calculated background mutation rate (BMR) [325]. Controlling for false positives is a challenge owing to a heterogenous BMR across samples. Since mutation rates have been shown to correlate with levels of histone H3K9me3 modification, gene expression due to transcription coupled repair and size of genomic footprint for a gene, these factors are accounted as covariates in the MutSigCV model [325]. However, as tools keep improving, a major challenge still lies in distinguishing drivers from passenger events, and to assess if the

identified SMG truly have functional roles in tumorigenesis. In a sequencing study, the identified SMG can either be technical artefacts. For example, they could lie in a hypermuted gene locus or they could harbour driver mutations that confer a functional advantage to the tumour cell. A total of 18 genes reach statistical significance at a FDR<0.05, however, on closer examination of the variants based on criteria such as, known driver, deleteriousness and potential artefact based on sequencing error, read depth/quality that were missed during filtering most of these genes (11 out of 18) were discarded as artefacts. Six of the top 10 highly significant genes include all the known UM drivers (*GNAQ*, *GNA11*, *BAP1*, *SF3B1* and *EIF1AX*) [18–20,22,228].

The identification of these established known UM driver genes in the current analysis has validated this approach for the identification of significant mutations. The only other high frequent event involved a gene known as HAUS Augmin like Complex Subunit 6 (*HAUS6).* This maps to chromosome 9p22.1 and has not previously been reported in UM. Two hotspot mutations (p.L553F and p.E743D) were identified in ~12% of samples. *HAUS6* mutations were found to be significantly associated with the subgroup of tumours (20%) which did not harbour any known secondary driver mutations (*BAP1*, *SF3B1* or *EIF1AX*) and mutations were found in 14 out 131 samples. *HAUS6* is a subunit of augmin protein complex, which plays a role in maintenance of chromosome integrity, microtubule attachment to the kinetochore and central mitotic spindle assembly formation [366]. The N terminal domain of HAUS6 is involved in the mitotic spindle assembly, which interacts with microtubule polymerisation promoting, NEDD1-gamma-tubulin complex and recruits the complex to the spindle [367]. However, these hotspot mutations were present at low allelic frequency and additionally do not lie in the N-terminal domain of the subunit 6 (HAUS_6_N) and its effect on the protein function will have to be functionally followed up.

The current analysis has validated and established the mutational frequencies of the known driver mutations in UM (*GNAQ*, *GNA11*, *CYSLTR2*, *SF3B1*, *BAP1* and *EIF1AX*). Further, examining the *PCLB4* gene mutations revealed previously described hotspot mutations, affecting residue codon 630 and additional non-hotspot mutations (p.G517V, p.M315V and p.V357A). Although these mutations were previously shown to occur in tumours that did not harbour GNAQ and GNA11 mutations, in the current study, this was not found to be the case in the current study. Four out of five samples with *PLCB4* mutation (including one with the recurrent hotspot mutation) also harboured either a *GNAQ* or a *GNA11* mutation. While the D630Y alteration affects the Y-domain of the conserved catalytic core of the *PLBC4* controlling signal transduction [198], the novel p.M315V and

p.V357A mutations targeting the PLC-beta domain is also part of the catalytic core of the enzyme [368]. *PLCB4* is one of the canonical downstream targets of Gq subfamily of G-protein coupled receptors, and could potentially play a role either independently or in tandem with *GNAQ /GNA11* mutations to mediate constitutive activation of downstream pathways of Gαq and Gα11 proteins.

The current pipeline could identify 19 out of 23 true *BAP1* mutations verified with Sanger sequencing (published). One of these variants that was not picked up by the callers suffered from coverage issue. The other three *BAP1* mutations that were not picked up by the mutation calling pipeline were from samples (MM144 (p.F50LfsX22*), MM173 (p.R146M*) and MM175 (c.867_887del*) where only RNA-sequencing data were available. This may not be surprising since loss of function mutations in *BAP1* in the form of nonsense or frameshift mutations will very likely lead to nonsense mediated decay and loss of the transcript. To account for additional the GEP based class 2 samples with a missing *BAP1* mutation, a complementary approach was used by querying through filtered bins and integrating matched copy number and RNAseq data and integrating calls from different variant callers. This led to the identification of five additional novel *BAP1* mutations that were missed by most variant callers due to poor coverage or due to design of the exome target baits which can miss out on large deletions that span the intron (list here). In addition, three novel mutations in *SF3B1* (A881, A9EC and A9EH) and one in *EIF1AX* (A8KG) were identified. Final frequency estimates of the established UM drivers were thus calculated in the current study of 131 samples as: *GNAQ* (48.09%), *GNA11* (46.56%), *CYSLTR2* (4.58%), *SF3B1* (24.42%), *BAP1* (40.45%) and *EIF1AX* (16.03%).

Integrating mutation calls and copy number data, helped in the identification of several novel alterations that are part of RNA splicing and processing pathways. Mutation data was also queried for 'cis-genes' (as previously describe in results - chapter 1) identified by integrating copy number and gene expression data in chapter 2. In addition, other components of the spliceosome were also mined in the exome data. This led to the identification of mutations in *SNRNP40*, *SRRM1*, *DNAJC8*, *EIF4G3*, *HNRNPR*, *SRSF4*, *SRSF5*, *SYF2*, *SF3A1*, *SRSF7* and *RBM10*. Interestingly, these mutations do not lie in samples with hemizygous loss of the gene locus, and were present in samples which already had a mutation in the SF3B1, gene. This suggests either that the tumour environment is selecting for splicing defects, or the activity of the mutant *SF3B1* requires mutations in additional members of the spliceosome complex.

The mammalian SWI/SNF chromatin remodelling complex has been described previously in pancreatic [369] and gastric cancers [362]. They have been proposed to affect expression of RB1, thereby affecting regulation of the cell cycle and in this way can be thought of as tumour suppressors [370]. In this study, by integrating SNP array based copy number data, exome sequencing and RNA-sequencing, additional genetic alterations targeting the SWI/SNF complex genes have been identified. These mutations were nearly always present in tumours with a secondary driver mutations (*BAP1, SF3B1* and *EIF1AX*), however are more likely to be found in class 1 compared to class 2 tumours. Overall, loss of components of the SWI/SNF complex through focal homozygous deletion and damaging mutations accounted for only 0.8-7% of tumours. However, when taken together, genomic aberrations targeting this complex account were found in 19% of tumours.

Previous studies [46,222] on the TCGA cohort did not report the novel mutations described in the current study. This includes mutations in *HAUS6*, and mutations in genes encoding components of the chromatin remodeling complex and RNA processing. Although the two studies had comparable number of exome sequenced samples, the pipeline used in the current study differed with respect to the alignment and variant calling tools. One of the novel aspects of the method employed in the current study involved the integrated approach that was used where copy number, gene expression and mutation were combined to identify novel genes.

While the major high frequency mutations in *GNAQ*, *GNA11*, *BAP1*, *SF3B1*, and *EIF1AX* constitute "mountains" in the landscape of cancer genome, the identification of additional infrequent mutations and copy number alterations of epigenetic regulators and pre-mRNA splicing and processing could constitute the "hills" in the cancer genome. Further evaluation of these events with functional work will help to elucidate their importance in tumour pathogenesis.

Chapter 5. Transcriptomic investigation of gene fusions in uveal melanoma

5.1 Results

5.1.1 Fusion events detected across all the samples

Twenty-two primary UM tumour samples were used in this study. Paired end RNA-seq data consisted of two different types of libraries: Those prepared from PolyA+ mRNA (12 samples) and those prepared from total RNA (10 samples) **(Table 2.3)**. A total of 495,750,223 reads (~495.8 million) were available with an average of 13.80 million reads per sample for the PolyA+ selected data and an average of 33 million reads per samples for the total RNA data. In the current study the JAFFA [330] gene fusion detection pipeline was then used to analyze paired end RNA-seq data to identify all the fusion transcripts. After filtering of chimeric RNAs supported by less than 2 spanning reads and filtering of non-protein coding or ribosomal RNAs, 191 chimeric RNAs with a unique fusion junction were identified. There were 80 inter-chromosomal fusions, involving two different chromosomes and 111 intra-chromosomal fusions. These fusion transcripts were further categorized as polygamous, if either the 3' or 5' gene were detected in more than 1 sample and as private fusions if the fusion transcript was only detected in one tumour sample. Approximately 70% of all tumours expressed at least a single polygamous transcript, with a range of 1-8 polygamous transcript per tumour **(Table 5.1)**. Among the polygamous fusions transcripts, twelve were uniquely present in gene expression profile (GEP) based class 1 tumours and 8 were present in only class 2 tumours **(Figure 5.1)**.

Chromosomal mapping of genes that were part of the private and polygamous fusions revealed a non-random distribution with the largest number of fusion transcripts being derived from chromosomes 1, 3 and 5 and the fewest number of partner genes being derived chromosomes 22, X and Y **(Figure 5.2)**. Due to the small numbers of samples, it is difficult to infer any thorough conclusion regarding the tumour subclass specific distribution of fusion transcripts. However, a circos plot of GEP based class 1-specific and class 2-specific polygamous fusion gene partners indicated few class specific chromosomal differences with class 2 tumours tending to exhibit a wider range of fusion transcripts involving more chromosomes **(Figure 5.3)**. There is also a chromosome 5-1 fusion involving a CANX-

TMED5 fusion in a class 1 tumour (MM171) an additional CANX-NDUFAF1 fusion in a class 2 tumour (MM135). Other fusions with recurrent 5' transcripts include TTC3:CTNNBL1 and TTC3:SUPT20H in class 1 tumours MM082 and MM094 respectively; RCC1 fused to two different genes (CSE1L and HMGB1) in MM094 and ABCB5 fused to ATP6V0A2 in MM171 and EPS15 in MM173.

Table 5.1: Summary of gene fusion transcripts in UM.

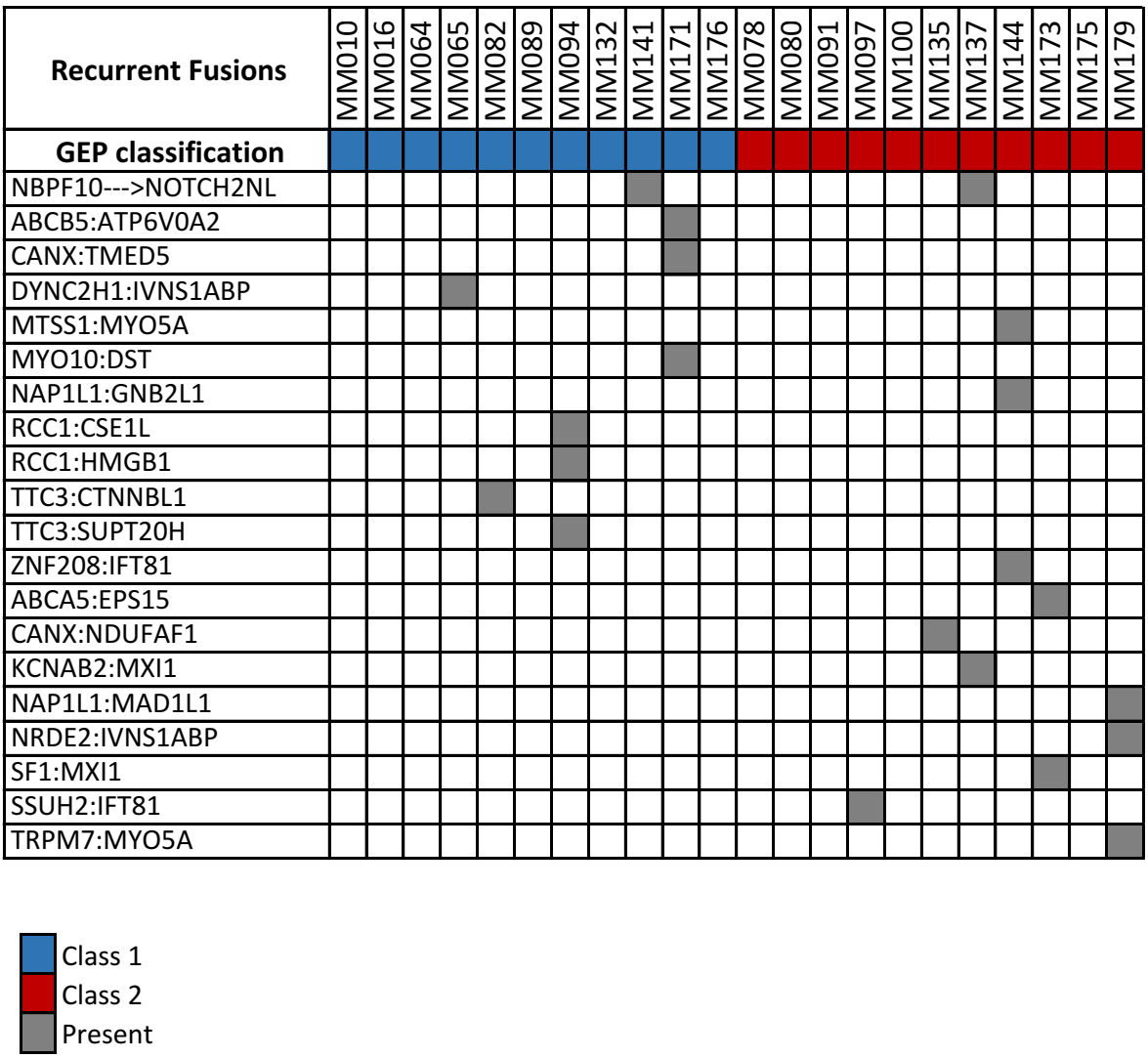
Tumours	Private fusions	Private fusions range /tumour	Redundant fusions	Redundant fusions range/tumour
All (N=22)	152	1-25	39	1-6
Class 1	86	1-25	23	1-6
Class 2	66	1-13	16	1-5
BAP1 mutant tumours*	83	2-19	22	1-5
SF3B1 mutant tumours*	10	1-5	3	1-2
EIFIAX mutant tumours	17	1-11	7	2-5
Wildtype tumours	32	2-25	6	1-6

*** excluded sample with both BAP1 and SF3B1 mutation**

Distribution of fusion transcripts among all primary uveal melanoma samples by tumour class and mutation status of known driver genes.

All the gene fusions were then queried against previously identified gene fusions in the TCGA Fusion gene Data Portal (http://54.84.12.177/PanCanFusV2/). This revealed only a single overlapping gene fusion between the current study and those reported in TCGA. This is the *TFG-GPR128* fusion which has been reported in patients with atypical myeloproliferative neoplasm and has been identified in the healthy population at a frequency of 0.02 [371]. This fusion was observed in the MM080 tumour sample. Other polygamous gene fusions predicted by Oncofuse, to have a high probability of being drivers were seen as private fusions in the current set include: *PQLC1-SMAD2* (sample: MM179), *RABL3-STXBP5L* (sample: MM171), *SASH1-APP* (MM144), *TRPM7-MYO5A* (MM179), *UBA2-DLC1* (MM171) and *YWHAE-CRK* (MM173).

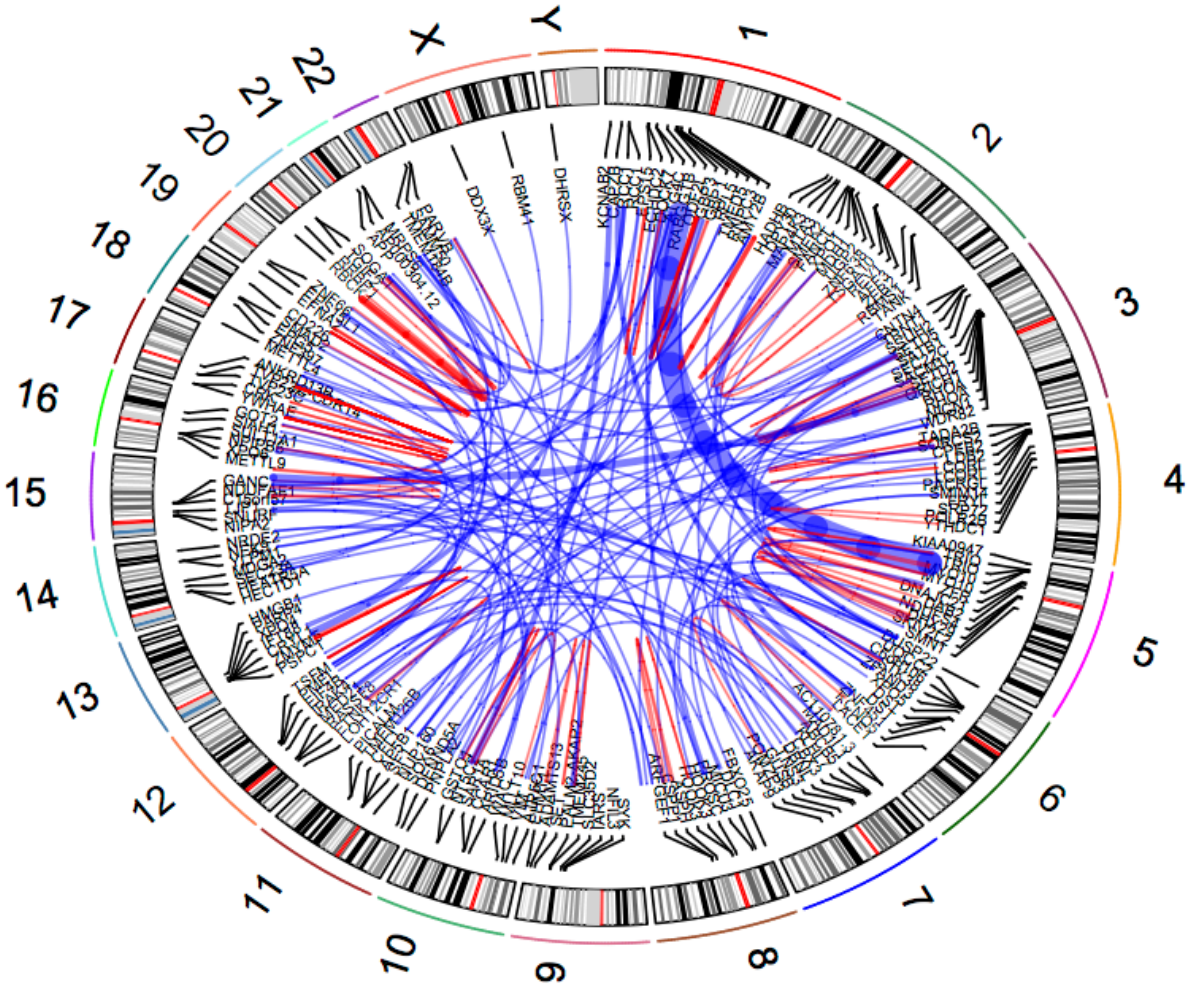
Figure 5.1: Polygamous fusion transcripts identified in primary uveal melanomas.



The GEP based tumour class is indicated on top and the grey shaded boxes show distribution of the fusion event across all samples.

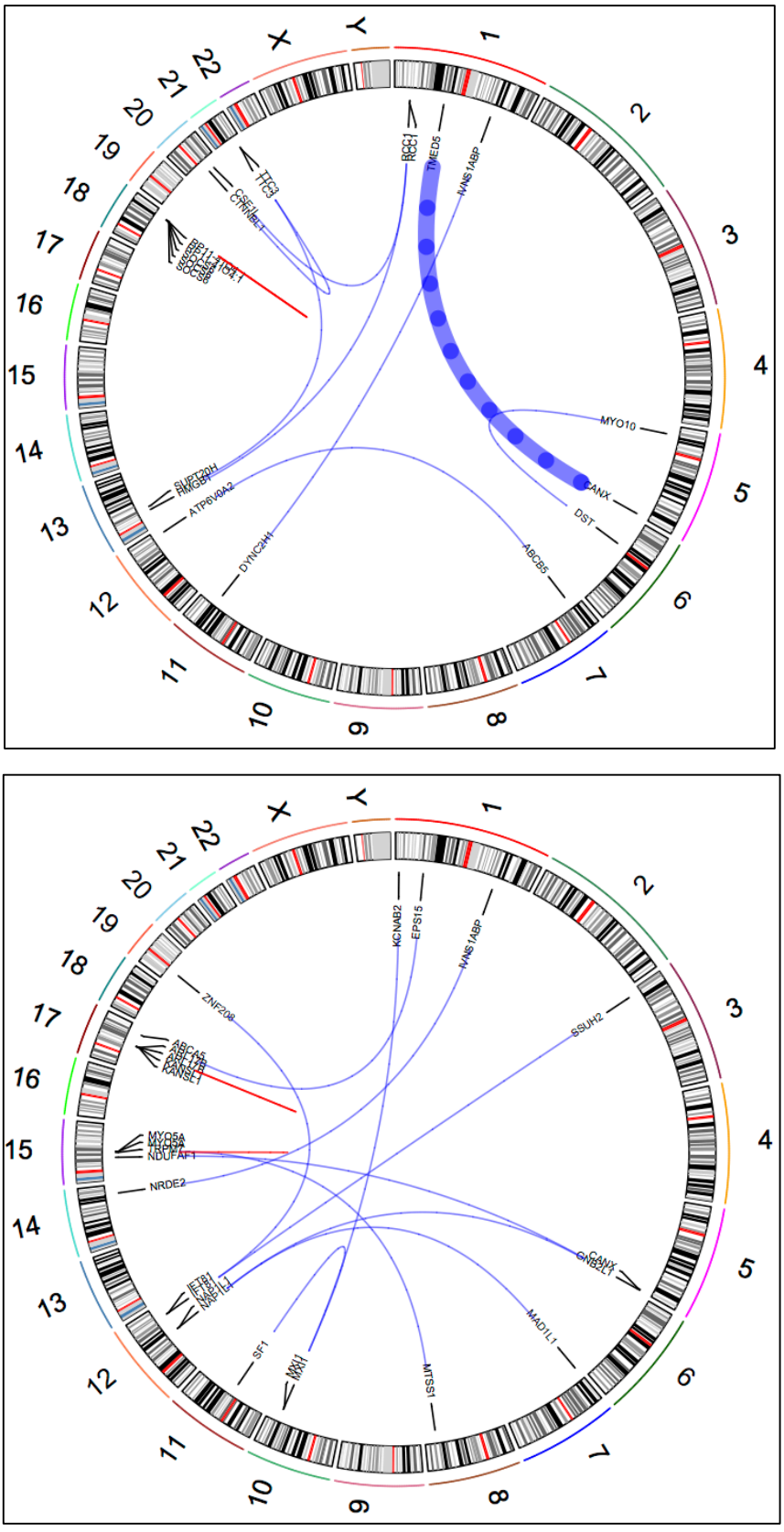
Figure 5.2: Circos plot showing chromosomal connections for all the fusion transcripts.

Chromosomes	chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	chr13	chr14	chr15	chr16	chr17	chr18	chr19	chr20	chr21	chr22	chrX	chrY
Number of Genes	32	19	28	16	30	16	15	19	9	16	12	20	11	9	18	12	12	10	9	11	6	4	2	1



The intra-chromosomal fusions are shown in red and the inter-chromosomal fusions are indicated in blue. Chromosomes are shown outside the circle and the genes that make up the fusions are indicated within the circle. Chromosomal distributions of fusion genes are shown above the plot. The thickness of the lines corresponds to number of reads supporting the event.

Figure 5.3: Circos plots of GEP class specific polygamous fusion events.



Upper: Class 1 tumours; Lower: Class 2 tumours.

5.1.2 Recurrent polymorphic gene fusion in uveal melanoma

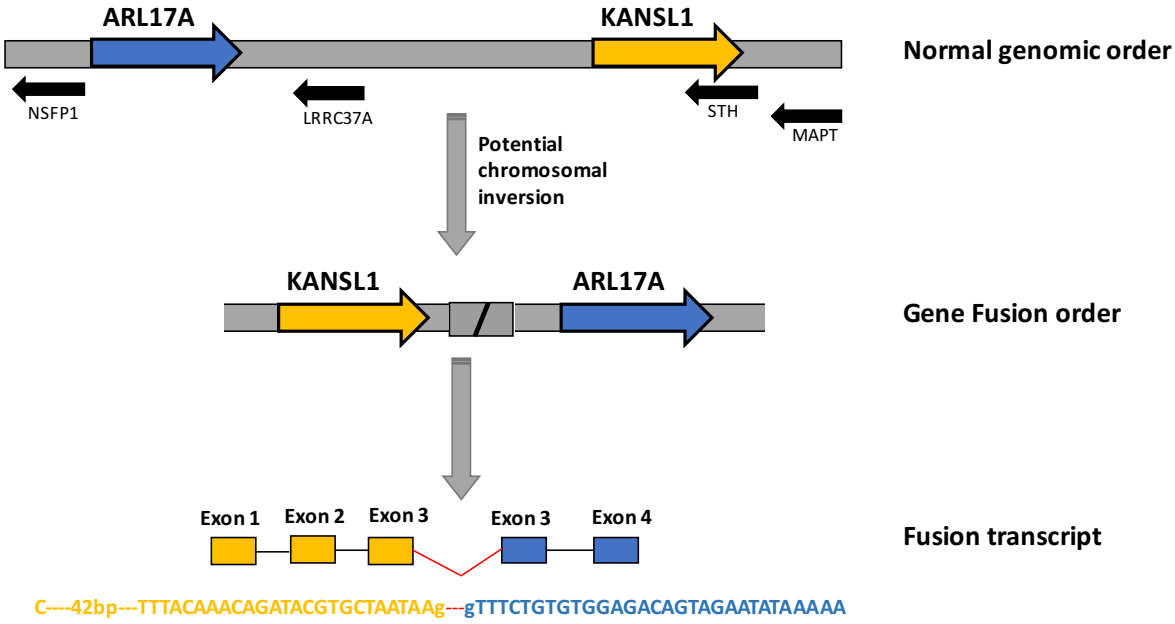
After performing all the filtering steps (described under chapter 2, Gene Fusion analysis section) to remove the polymorphic and low confident fusion events, no recurrent fusion (present in more than one sample) was found. Thus, to prevent missing relevant fusions that might be filtered out due to the stringent filtering criteria some of the filtered candidates were screened. This led to the identification of a single recurrent fusion occurring at high frequency in our primary UM RNA-seq samples and involved the KAT8 Regulatory NSL Complex Subunit 1 (*KANSL1*) gene which binds to its 3' partner gene ADP Ribosylation Factor Like GTPase 17A (*ARL17A*) to form the *KANSL1-ARL17A* gene fusion.

KANSL1-ARL17A gene fusion was seen in 36% (8/22) of samples. Both *KANSL1* and *ARL17A* are located on the minus (reverse) strand of chromosome 17q21.31 with a gap of 476 kilobases between them. The *KANSL1* and *ARL17A* genes encode 15 and 4 exons respectively and the breaks occur downstream of the second coding exon in the *KANSL1* gene (the last base of this exon maps to chr17:44171926 (hg19)) and upstream of exon3 of *ARL17A* so that its spliced transcript starts at chr17:44648235 (hg19) **(Figure 5.4, Table 5.2)**. In the normal genomic structure, the order of the genes with respect to the directionality of the mRNA is *ARL17A* → *KANSL1*, however the *KANSL1-ARL17A* gene fusion shows an inverted gene order indicating involvement of a potential chromosomal inversion event at play **(Figure 5.4)**. The resulting *KANSL1-ARL17A* gene fusion transcript encodes a putative peptide with 483 amino acids where majority of the peptide sequences are derived from *KANSL1* gene **(Figure 5.5)**. The putative fusion peptide retains the coiled coil domain of the *KANSL1* gene while losing the WDR5 interacting motif, zinc finger domain and the PEHE domain [372]. Univariate analysis of *KANSL1-ARL17A* fusion with other UM specific genetic mutations (*GNAQ*, *GNA11*, *BAP1*, *SF3B1* and *EIFIAX*), copy number alterations (monosomy 3, 6p gain, 6q loss, 8p/q gain, 8p loss and 16q loss) and gene expression profile class status, indicated no statistically significant association with any other UM specific genetic event, suggesting they are independent events. This fusion is a common inversion polymorphism at chromosome 17q21.31 where one allele is an inverted haplotype (termed H2) [373]. This H2 haplotype is common in Europeans, occurring at a frequency of ~20% and predisposes to a deletion which causes a rare congenital disorder (Koolen-deVries syndrome)[372]. Its possible relevance to UM is unclear, however UMs are found predominantly in individuals of European ancestry and the role of *KANSL1-ARL17A* fusion as a predisposing event could be explored.

Table 5.2: Recurrent gene fusions in UM

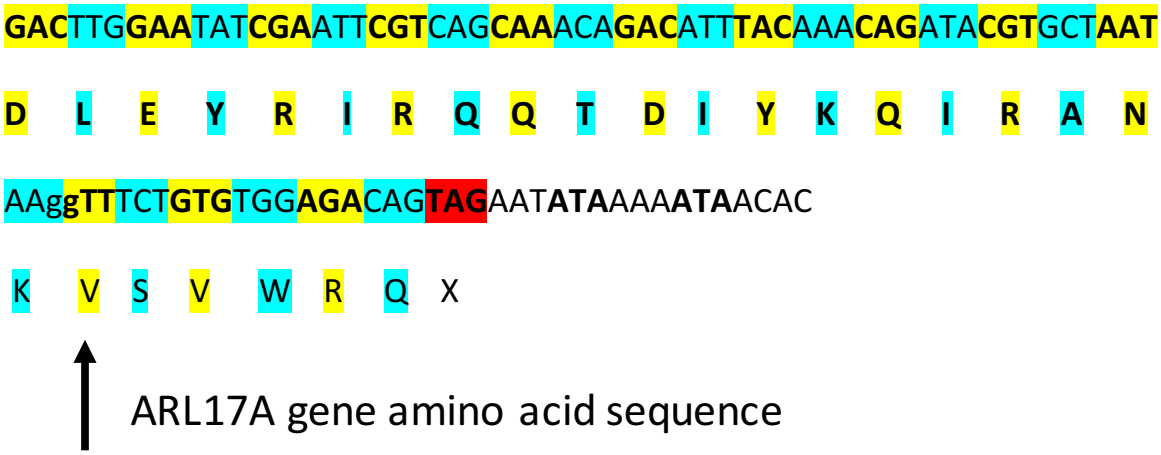
Sample	GEP class	fusion genes	chrom1	base1	strand1	chrom2	base2	strand2	gap (kb)	spanning pairs	spanning reads
MM010	1	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	2	7
MM065	1	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	2	6
MM091	2	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	0	7
MM097	2	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	1	2
MM144	1	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	0	4
MM171	1	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	0	10
MM175	2	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	1	2
MM179	2	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	1	1

Figure 5.4: ***KANSL1-ARL17A*** **gene fusion.**



Schematic diagram showing the normal and potential rearranged genomic structure of the *KANSL1-ARL17A* gene fusion on chromosomal band 17q21.31 and the fusion transcript that was identified.

Figure 5.5: Putative truncated protein sequence of the *KANSL1-ARL17A* gene fusion.



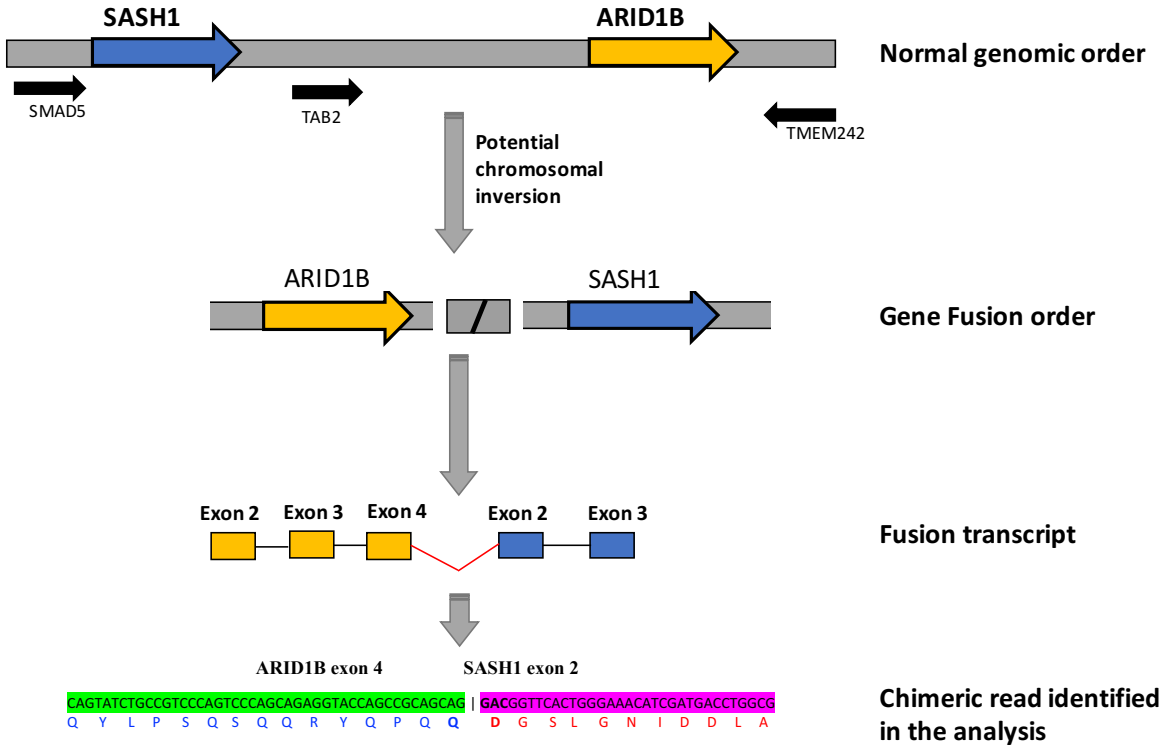
The codons are highlighted in yellow and blue and the amino acid corresponding to each codon is indicated below.

5.1.3 Other potential candidates of unknown significance

Although most UMs had at least one gene fusion event, there were some polygamous fusions involving genes encoding proteins involved in chromatin modification that were filtered due to low read counts (1 spanning reads). These fusions were predicted by JAFFA as medium confident calls and were found to have a high probability of being a driver. They included *ARID1B-SASH1* and *ARID1B-RNF149* fusions **(Table 5.3)**. In previous chapters, mutation and copy number aberrations affecting the *ARID1B* gene was reported **(Table 5.4)**. Additionally, another *ARID1B-SASH1* gene fusion transcript, involving a putative breakpoint located downstream from exon 3 or 4 depending on the isoform of *ARID1B* and upstream of exon 2 of the *SASHI* gene was identified. This would result in a putative fusion peptide of 1837 amino acids (**Figure 5.6)**. This putative peptide was predicted by Oncofuse to have a high probability of being a driver event. For the *ARID1B-RNF149* fusion transcripts, the 5' *ARID1B* region is retained on chromosome 6q with breakpoints located downstream from exon 4 or 5 depending on the isoform, with a second breakpoint before exon 2 within the *RNF149* gene on chromosome 2. However, this leads to an out of frame partial transcript of *RNF149*, ultimately resulting in a fusion peptide where *RNF149* is truncated and where the mRNA is likely to undergo nonsense mediated decay **(Figure 5.7)**. Hence, this fusion could inactivate *ARID1B*. Functional domains of the wild type *ARID1B*, which would be potentially disrupted in the putative peptide include the Zinc finger, RING-type domain (*ARID1B-RNF149*) and variant SH3 domain (*ARID1B-SASH1*).

ARID1B is a member of the human SWI/SNF transcriptional complex (or BAF complex) which is involved in regulating chromatin structure [341]. It has been implicated as a driver of tumourigenesis in a subset of hepatocellular carcinomas, breast cancers, ovarian cancers, medulloblastomas and pancreatic cancers [369,374–377]. Two known functional domains in *ARID1B* comprise an "AT-rich Interactive Domain" (ARID) and a "Domain of Unknown function" (DUF3518). The ARID domain (~100 amino acid residues) binds to DNA without sequence specificity [341] and the DUF3518 domain interacts with the helicase subunits *BRG1* and *BRM* which are also members of the BAF complex, [342]. In both the *ARID1B* fusions with the different 3' partners reported here, both the ARID and DUF3518 functional domains of *ARID1B* including are likely to be lost **(Figure 5.8)**. *SASH1* has also been described as a potential tumor suppressor that may negatively regulate proliferation, apoptosis, and invasion of cancer cells, and reduced expression of this gene has been observed in multiple human cancers[378,379].

Figure 5.6: Potential chromosomal rearrangement of *ARID1B-SASH1* gene fusion.



Schematic diagram showing the normal and potential rearranged genomic structure of the *ARID1B-SASH1* gene fusion on chromosome 6 and the fusion transcript that was identified

Figure 5.7: *ARID1B* fusions in uveal melanoma.

***ARID1B-SASH1* Fusion transcript**

ARID1B exon 4 **SASH1 exon 2**

CAGTATCTGCCGTCCCAGTCCCAGCAGAGGTACCAGCCGCAGCAG | GACGGTTCACTGGGAAACATCGATGACCTGGCG

Q Y L P S Q S Q Q R Y Q P Q Q D G S L G N I D D L A

***ARID1B-RNF149* Fusion transcript**

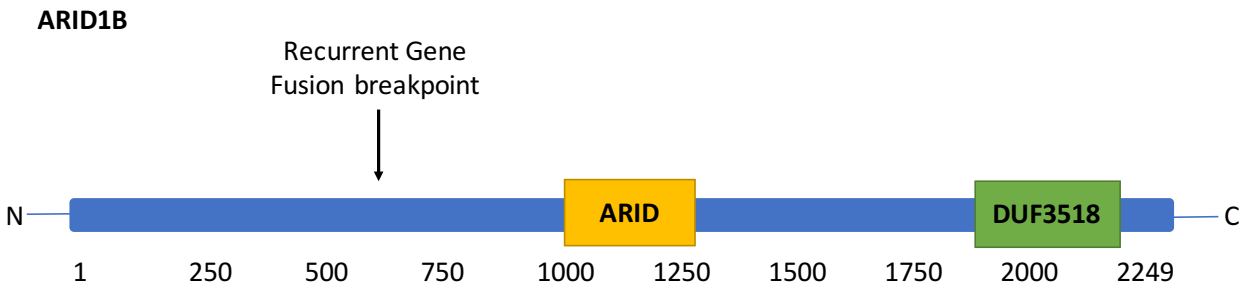
ARID1B exon 4 **RNF149 exon 2**

TTGAACTTAATACAGCAAGAAAGACCATCAAGTTTACCA | GAACAGGAAATATAGTGGTCATTATGAT

L N L I Q Q E R P S S L P E G E I X

Shown above are the two fusion events where a recurrent 5' partner gene *ARID1B* is fused to different 3' partners (*SASHI* and *RNF149* genes) and the resulting fused putative peptide sequence at the break.

Figure 5.8: Structure of the *ARID1B* protein.



Schematic structure of the *ARID1B* protein showing the functional domains which are lost in the putative gene fusion peptides.

Table 5.3: Gene fusions with ARID1B partner

Sample	5' Partner Gene	5' chromosome	5' location	3' Partner Gene	3' chromosome	3' location
MM179	ARID1B	chr6	157222659	SASH1	chr6	148711270
MM016	ARID1B	chr6	157256710	RNF149	chr2	101911643

Potentially relevant gene fusions with low spanning reads (=1). MM179 is class 2 and MM016 is class 1 sample, both of these tumours had metastatic disease.

Table 5.4: List of ARID1B mutations identified by exome and copy number analysis

Datatype	Samples	Alteration
Exome analysis	MM105	Missense_Mutation
Exome analysis	MM049	Missense_Mutation
Exome analysis	MM010*	Missense_Mutation + Homozygous deletion
Copy number	MM032	Homozygous deletions
Copy number	A9EH	Homozygous deletions
Copy number	A9F4	Homozygous deletions
Copy number	A8KB	Homozygous deletions
Copy number	A8K9	Homozygous deletions

5.2 Discussion

Gene fusions leading to chimeric transcripts are frequently observed in blood born cancers and are often a result of chromosomal level structural alterations, including translocations, inversions and other complex rearrangements. With the advent of the new sequencing technology the landscape of structural alterations in solid tumours is beginning to emerge. This has led to the discovery of fusions between *EWSR1* and the *ETS* family of transcription factors in bone sarcomas [380], fusions between *TMPRSS2* and *ETS* transcription factor *ERG*, *ETS* and *ETV4* in prostate cancers [174,175], an *FGFR3-TACC3* fusion in bladder cancer [381], an *EML4-ALK* fusion in lung adenocarcinomas [382], *PAX8-PPARG* fusions in follicular thyroid cancers [383] and many more. This study has attempted to identify novel gene fusions events in UM. These were detected using RNA-seq data, and comprised both canonical and non-canonical chimeras, resulting from read-through or trans-splicing events. While it is a challenge to identify, chimeric transcripts produced solely due to rearrangements at the DNA level without the availability of whole genome sequencing data, some of the non-canonical chimeric RNAs due to read-through (cis-splicing) events were eliminated by using the minimum gap criteria between two fused genes.

Overall, examining the chimeric transcripts, revealed more intra-chromosomal gene fusions compared to inter-chromosomal events. Of the 191 fusions events identified, 152 were private fusions restricted to a single sample and 39 were polygamous fusions, where a recurrent gene partner was found in multiple samples. Among the polygamous fusions, 12 were unique to the GEP based class 1 tumours and 8 were found in the class 2 tumours. Recurrent polymorphic *KANSL1-ARL17A* fusions were found across both class 1 and class 2 samples. Although *KANSL1-ARL17A* fusion on chromosome 17q21.3 has been reported previously (called *KANSARL*) as a cancer predisposition fusion in the European population [384], this study focusses on fusions that are unlikely polymorphisms. In whole, somatic fusions events in UM were not as common as in case of other solid tumors such as prostate cancer and lung cancers [174,382]. Hence, gene fusions are unlikely to be a major driver of UM. However, the identification of low frequency fusions involving chromatin modifiers such as ARID1A is of interest. In one instance the downstream partner (SASH1) has already been implicated as a tumour suppressor. A change in its structure could affect its activity and could increase risk of metastasis and that it will be interesting to look at the frequency of such fusions in metastasizing samples, to see if they could have played a role in metastasis.

Interestingly *ARID1B* fusions involving breakpoints located on either downstream from exon 4 or 5 and leading to fusions with the *ZNF384* gene have been reported in pediatric acute lymphoblastic leukemia [385]. Additionally, recurrent genome-wide rearrangement of *ARID1B* have been reported in neuroblastoma [386]. Further, gene fusions event involving *ARID1B* as the 5' partner have been reported in the TCGA fusion database from breast cancer, glioblastoma multiforme, brain lower grade glioma and lung adenocarcinoma [182]. This suggests that the two fusions reported here in UM are less likely to be random events, despite their low abundance.

Although one of the key drawbacks in the current study is the lack of orthogonal validation of the reported gene fusion events with Sanger sequencing of cDNA due to the unavailability of samples for experimental follow up, the gene fusions identified in this study can still be considered plausible candidates, based on their recurrence and biological significance. To gather further support to establish the occurrence of these events, additional RNA-seq data could be screened for the reported fusions such as the one involving the *ARID1B* gene, on a larger number of samples. Ideally such recurrent fusions should then be validated with qRT-PCR and Sanger sequencing. The effects of such possibly pathogenic fusions on gene expression can then be evaluated by transfection into human cells, followed by an investigation of their effect cell signaling, transcriptional activation, proliferation, invasion and migration. Some of these fusions could be therapeutic targets and could help in understanding the underlying mechanism of tumourigenesis or metastasis.

Chapter 6. Conclusion

Uveal melanomas (UM) is the most common eye cancer which manifests into a highly aggressive form affecting nearly half the patients with the tumour. This study describes the first integrated genomic and transcriptomic investigation performed using different high-throughput datasets in UM. To understand the genomic architecture of these tumours, a comprehensive investigation was performed with whole genome SNP-arrays, whole exome sequencing and RNA-sequencing technologies on a total of 207 primary tumour samples (including samples with matched data across different platforms) spanning across three different cohorts (WASH-U, TCGA and CC). Analysis of the data revealed that UM exhibits a less complex pattern of genomic aberrations, characterized by recurrent chromosomal alterations and low mutational burden, compared to most solid tumours [200,387]. With the exception of few samples, the tumour genomes displayed a homogeneous low burden of single nucleotide variants and notable there was absence of the canonical ultra-violet radiation (UVR) signature, related to DNA damage at pyrimidine dinucleotides, thus, negating the role of UVR in the etiology of UMs.

Landscape of copy number and structural aberrations in uveal melanoma

UMs exhibit unique characteristic highly recurrent aneuploidy of few selected chromosomes, including chromosomes 1, 3, 6, 8, 11 and 16. Chromosomal alterations that are associated with metastasis include monosomy 3, loss of 6q, loss of 8p and gain of 8q, this is consistent with previous studies [32–34]. Recurrent chromosomal loss of 1p, was previously shown to be associated with decreased disease-free survival [37,155], however in the current study it was not found to be associated with metastatic outcome. Somatic polyploidy is a characteristic of human tumours and are frequently in the form of aneuploidy, a key feature in UMs. Majority of UMs have ploidy close to 2n, with a few cases exhibiting tetraploid state which tends to occur more in the gene expression based highly aggressive class 2 tumour compared to low metastatic competent class 1 subgroup. Monosomy 3 is found exclusively in the class 2 tumours is suggested to lead to increased genomic stability [38]. However, there was no statistical difference in the ploidy estimates between these two groups based on a limited sample size (N=45; where we had information on class status). The percentage of non-aberrant cell fraction was higher in the class 2 compared to class 1 tumours and could be due to the presence of immune related cells (macrophages or lymphocytes) surrounding the

tumour tissue previously shown to be associated with poor prognosis in UMs [295,361]. The potential role of these immune cells in the tumorigenic processes could help in developing new treatments and requires exploration.

Unsupervised clustering of broad copy number alterations revealed four distinct groups, each characterized by a combination of gain and losses involving 1p loss, 1q gain, monosomy 3, 6p gain, 6q loss, 8p gain/loss 8q gain and 16q loss. Previously Harbour et al. [38] had established a prognostically relevant molecular classification based on the copy number status of chromosomes 3 and 6p gain using statistical pattern recognition analysis, where they identified three main subgroups: Tumors with normal copies of chromosome 3 and 6p; tumors with gain of 6p and normal copy of chromosome 3; and tumors with monosomy 3 and normal copy of 6p. The latter has the worse prognosis. Based on their study they concluded that monosomy 3 and 6p gain are mutually exclusive and represented a bifurcation in tumour progression [129]. In the current larger study, monosomy 3 and 6p gain were mutually exclusive most of the time, however a subset of tumours (11%) harbored both alterations. This could be due to inter-tumoural heterogeneity and suggests potential ongoing tumor evolution. The genesis and the sequence of these events arising during tumour evolution will require further exploration.

The gene expression based classification of tumours into low metastatic competent class 1 tumours and the highly aggressive class 2 tumors is established and this classification perform far superior in predicting metastatic risk compared to monosomy 3 and other clinicopathological prognostic factors, in addition to revealing the pathobiology of these tumours [42,114]. Incorporating this information can thus shed more light in understanding the underlying biological mechanisms. Recently, PRAME expression was identified as an independent marker of metastasis and found to be associated with 1q gain (specific to class 1 tumours), 6p gain and 16q loss, in addition to 6q loss, 8p loss (specific to class 2 tumours) and 8q gain [46]. However, these alterations (1q gain, 6p gain and 16q loss) did not reach statistical significance for association with metastasis independent of PRAME status in the current study. Taken together these observations, identification of candidate events on these chromosomes would be of high relevance to uncover the mechanism of tumour progression.

The underlying hypothesis explored here was that some copy number and structural alterations which harbor genes that are highly selected for and are instrumental in causing defects in pathways that lead to UM pathogenesis. To uncover the candidate genes that lie in the recurrent region of chromosomal gains and losses, segmented copy number data were subject to analysis with the GISTIC algorithm which identifies focal chromosomal alterations

that are significantly higher than the background rate [301]. Systematic evaluation of each GISTIC peak revealed a total of 203 independent focal regions of gains and losses spanning chromosomes 1, 2, 3, 4, 6, 7, 8, 9, 11, 15, 16, 17, identified across three independent GISTIC analysis performed on data from three different cohorts. To further increase the confidence for identifying candidate events, overlapping peaks boundaries and adjacent peaks within a one megabase distance, across three independent analysis were selected. Ten overlapping peaks were found. These included four deletion peaks at chromosomes 1p35, 2q37.2, 6q27 and 11q24.3 and six amplification peaks at chromosomes 6p25.2, 6p22.1, 6q26, 8q24.3, 16p13.3 and 19p13.3. Alterations in copy number of some of the chromosomes harboring these CNAs has been previously described (e.g. chromosomes 1, 6, 11, 16 and 19), and this study has identified candidate regions and sometimes genes within these chromosomes that could be responsible for the copy number changes. Alterations in chromosome 2q have not previously been described, but deletion of 2q37.2 implicates a novel noncoding RNA, LOC200772. While some of the peaks identified in the current analysis agreed with previous studies performed with lower resolution technology [156,157], a few regions were not replicated. Such as a small region of loss of heterozygosity (LOH) on 1p31, previously refined with microsatellite analysis [155]. The closest peak to 1p31 identified in the current analysis was on 1p32.2 (chr1:57450404-58743937). The SNP markers interrogated here are denser compared to microsatellites and the refinement with the current study is expected to be greater.

RNA-sequencing data generated on a subset of tumours from our cohort was used to explore potential candidate gene fusions resulting from structural events in UM. Although, 191 fusion events (both intra-chromosomal and inter-chromosomal) were identified, after applying stringer criteria, none of the events filtered out **(Appendix 5)** were reasoned to be promising for further follow-up. However, screening through the filtered low fusion spanning read bin (fusions candidates supported by 1 read), two candidate fusions events with ARID1B as the 5' partner was rescued. These included candidates, *ARID1B-RNF149* and *ARID1B-SASH1* in two metastasizing tumours.

To understand how these identified copy number alterations compared with other cancers and to identify those critical regions that are unique to UMs and those that are shared across different cancer types, all the focal CNAs were compared to the pan cancer findings [338]. This analysis revealed three important candidate cancer genes, including *PLEC1* (8q24.3 amplification peak), *NCAPD3* (11q25 deletion peak) and *ARID1A* (1p35.3 deletion peak) which lie on chromosomes 8q and chromosome 1p respectively. To further prioritize genes from the CNAs that are most likely to have a functional effect, those exhibiting a change in

gene expression in UMs were identified. This approach revealed a large complementary set of genes, that lie within the refined focal regions of copy number change and affect many overlapping pathways, including NF-kappa B and MAPK signaling, WNT signaling and mRNA splicing/RNA processing. Exome and transcriptomic fusions were the mined to add more value to these novel individual candidates and those that were present in pathways, previously not described in UMs. These revealed mutations in 19 genes in regions of deletion and only two genes with recurrent mutations in regions of amplification (*AXIN1* and *RAC3*) these mutations could render effect due to haploinsufficiency or over-expression.

Dominant drivers in uveal melanoma

The driver landscape in UMs is dominated by previously identified genes, including *GNAQ, GNA11, CYSLTR2, SF3B1, BAP1* and *EIF1AX*. The current study demonstrated the use of a complementary approach employing different bioinformatics tools and datatypes to help identify mutations that were missed by either by the inherent drawback of using a single tool or due to poor quality data. This helped to get an accurate estimate the frequency of all the known drivers using a cohort of 131 samples: GNAQ (48.09%), *GNA11* (46.56%), *CYSLTR2* (4.58%), *SF3B1* (24.42%), *BAP1* (44.27%) and *EIF1AX* (16.79%).

Additionally, novel mutations in these known UM driver genes were identified and improved to get a clearer picture of the driver landscape in UMs. After accounting for all the mutations identified in the current analysis, ~98% of all samples harbor mutations in at least one of these genes. In addition, five mutations in *PLCB4* gene were identified, including three novel and two previously reported D630 hotspot mutations, which tend to co-occur with the *GNAQ* and *GNA11* genes. One important theme that emerged was the identification of additional *BAP1* mutations in three samples, which had been classified within the class 2 subgroup based on their GEP. After accounting for these secondary driver events, nearly 20% of samples did not harbor a mutation in at least one of these genes, suggesting that either the known driver was missed due to experimental error, because mutations lay in regions that had not been assayed (e.g. introns in the case of *BAP1*) or because there are additional secondary drivers to identify.

In summary, this study has confirmed the occurrence of mutations in all the previously reported known driver genes in UM and report robust estimates of the mutation frequencies for these genes, using a large sample size and complementary bioinformatics approach.

Novel candidates identified in uveal melanoma

Gain of 8q has been long implicated in UM and is associated with metastasis. Previous studies have reported a common region of amplification at 8q24.1 to 8q24.3 [134,388]. The CNA with the highest frequency identified in this study was amplification of a region of chromosome 8q24.3 that contained the Plectin 1 (*PLEC1*) gene (74%). Based on the current data, no important role for c-MYC amplification was found in UM, which agrees with earlier studies. There were no mutations found in PLEC or plectin gene, however it was associated with a significant increase in gene expression in tumors with amplification compared to the non-amplified group. Overexpression of Plectin has been identified as biomarker of pancreatic ductal adenocarcinoma (PDAC) and abnormal detection on the surface of PDAC cells, whereas in normal cell it is restricted to cytoplasm [389]. Plectin is an intermediate filament (IF)-associated protein that belongs to the family of plakins or cytolinkers which are crucial for cross-linking and act as a signaling scaffold for cytoskeletal proteins. It is found in widely different cell types including epithelial cells[389]. It is implicated in regulating the MAPK and the PKC signaling pathways and leads to Erk1/2 activation and cell migration [389]. Interestingly, one study has shown that aberrant expression and mislocalization of plectins in PDAC cells promotes proliferation, migration and invasion in these cells [389]. This could be relevant in the context of UMs, where plectins could be implicated in promoting the metastatic phenotype. Immunostaining of plectin in uveal melanoma has previously not been described and determining its localization and function in UMs will be important in the future.

Exome analysis of 120 primary tumours revealed another novel candidate gene, *HAUS6* that was significantly mutated in ~11% of samples. Interestingly, *HAUS6* was enriched in the subset of tumours that did not harbor any secondary driver event (*SF3B1, BAP1* and *EIF1AX*). This group was comprised of a higher percentage of class 1 tumours. *HAUS6* encodes a protein which is a subunit of the augmin complex and is involved in microtubule attachment to the kinetochore and spindle formation. *HAUS6* binds to gamma tubulin ring factor, a major microtubule nucleator via NEDD1 and promotes nucleation of microtubules [390]. The functional impact of this gene requires further assessment for its role in mediating chromosomal instability in UMs.

Alternative mechanistic pathway

As an alternate to the candidate gene approach, where a single altered gene carries a high burden with respect to cellular pathology, several genes affecting a common pathway could be also be at play in mediating the tumorigenic process. To this effect examine the mutations in enriched genes that lie in copy number regions associated with significantly altered expression revealed a subset of samples (10%) harboring mutations in genes encoding components of the eukaryotic splicing and translational machinery: *SNRNP40*, *SRRM1*, *DNAJC8*, *EIF4G3*, *HNRNPR*, *SRSF4*, *SRSF5*, *SYF2*, *SF3A1*, *SRSF7* and *RBM10.* These mutations were present in samples with known driver mutation, including *SF3B1*. However, these genes did not lie in regions with copy number change. The role of splicing defects in UMs require further investigation to determine how these alterations leading to tumorigenesis

Emerging role of epigenetic modifiers in uveal melanoma

Another theme that emerged by following up on the candidate gene *ARID1A* identified from the pan cancer analysis, was the identification of genetic alterations from additional genes that form a subunit of the human SWI/SNF complexes, Brahma-associated proteins (BAF complex) and polybromo-associated BAP (PBAF) complex [342]. The BAF and the PBAF complexes subunits share homology and exhibit differences with respect to a few subunits, however they are frequently found to be disrupted in cancer [342]. Integrated genomic profiling using mutational data, gene fusion event and copy number alterations, revealed the involvement of the BAF and PBAF chromatin remodelling complex in ~17% of UMs. These genes include: *ARID1A, ARID1B, SMARCA2, SMARCA4, ACTL6A, PHF10, ARID2, BCL7A, BCL7C, SMARCC1, SMARCC2, SMARCD1* and *SMARCD3*, with *ARID1B* showing highest burden of genetic alterations (~7%). ARID1B has been found to be frequently mutated in several cancers [377,385,386] and is thought to regulate chromatin structure [341]. Previous cancer sequencing studies have shown that genes encoding the SWI/SNF chromatin remodelling complex are common targets of mutations. Although when considered individually, these mutations represent “hills”, when all the components are taken together these alterations represent “mountains” and provide an additional pathway to the development of UM.

Technical challenges in the current project

There were a few technical challenges while analyzing these high-throughput data. Due to the heterogeneous nature of SNP data generated across different platforms across three different cohorts, not processing them uniformly increases the inherent noise when combined across different platforms. Further, classification of focal and broad events by GISTIC requires a threshold to select the focal events that exceed the specified threshold to calculate the G-score statistics. This threshold was picked up by examining the histograms of the copy number from the samples. Due to noise in the data, determining accurate threshold level was a challenge and thus a judgment call rather than any empirical method was used to estimate the threshold. The results from GISTIC results cannot be easily integrated across different platforms and this posed an analytical challenge. This was overcome by combining the platform independent segmentation profiles across all the samples and querying the boundaries of the focal copy number peaks identified by GISTIC analysis in each independent analysis. This list was further pruned to select high confident peaks less affected by technical variation based on high frequency across multiple cohorts. The other larger theme that came from the analysis is that the choice of tool and the parameters vary from study to study and require careful curation of the data to infer biologically meaningful findings. Another challenge is that different tools are written in different computing environments and come with their own set of dependencies which can be a problem to perform and compare analysis with different tools, although this scenario is changing with more help from the online bioinformatics community.

Conclusion and future directions

UMs present as unique type of solid tumours with a low mutation burden and a very specific set of recurrent genetic events in tumour subtypes. This study performed a comprehensive investigation into the landscape of genome-wide genetic alterations that encompasses majority of UMs. The novel candidates and pathways require further exploration. The current disease model, involves oncogenic inactivating mutations in genes *GNAQ, GNA11, CYSLTR2* and potentially *PLCB4* which lead to inappropriate re-entry to cell cycle pathways. As the tumour progresses additional genetic alterations are accumulated, although the exact evolutionary process of how and when these sequences of events arise is unknown. However, copy number and mutational changes help define the molecular classes of these

tumours. Based on the gene expression profile, copy number and mutation in secondary drivers (*BAP1, SF3B1* and *EIF1AX*), four subgroups can be identified. These are described as: Group 1: Class 1 expression profile, *EIF1AX* mutation and gain of 6p; Group 2: Class 1 expression profile with *PRAME* over expression, *SF3B1* mutation, gain of 6p, gain of 8q and loss of 6q; Group 3 & 4: have similar profiles that include class 2 expression profiles, *PRAME* over expression, monosomy 3, *BAP1* mutation, 8q gain, 8p loss and 16q loss. PLECTIN could potentially be the candidate gene on 8q helping tumour with 8q gain to metastasize along with BAP1 loss. Mutations in *HAUS6*, a potential novel driver gene identified in the current study was largely be found in group 1 tumours.

In summary, the current project has helped to refine previously described chromosomal alterations and to identify novel candidate gene and/or genes in candidate pathways, including chromatin remodelling and RNA processing. The genes targeting the human SWI/SNF and spliceosome complexes could not have been identified if only the mutation or copy number data had been examined separately because of low power. Here, by using an integrated approach that included high resolution SNP arrays, and transcriptomic and exome sequencing, structural alterations and single nucleotide mutations were identified in the ATPase and putative DNA binding domains of components of the SWI/SNF complex.

In addition to reporting robust estimates of mutation frequency of known UM driver genes, mutations in a potential novel driver gene, *HAUS6* were described. *HAUS6* is part of the HAUS augment complex which is involved in mitotic spindle assembly and maintenance of centrosome integrity. Centrosome dysfunction is linked to aneuploidy and chromosome instability and the role of *HAUS6* in the context of genomic instability in UM could be relevant. Although, one drawback of the current analysis was the lack of orthogonal validation such as Sanger sequencing of the identified mutations, the identification of *HAUS6* variants across multiple samples spanning both the TCGA and WASH-U cohort indicates these are less likely to be due to sequencing artefact. These SNVs are present at low allele frequencies and only could be picked up by a sensitive mutational caller such as Mutect2. The low frequency could reflect genetic heterogeneity due to sub clonal events and can be validated experimentally with approaches such as digital PCR. Once validated, the *HAUS6* gene could be further functionally explored to test the link between mutation and chromosomal defects that are wide spread in UM.

While the current knowledge dictates that the dominant driver mutations are the key players in UM pathogenesis, other new events such as mutations in novel genes or altered pathways contributed by the accompanying copy number alteration identified here, could

play a role in tumourigenesis or metastasis. The roles of the newly identified amplification of PLECTIN and mutations in the *HAUS6* gene warrant further investigation. With respect to PLECTIN, that was found to be overexpressed in UMs, the localization of this protein could be examined with immunostaining to explore its role as a potential biomarker as previously describe in PDACs. PLECTIN are found on chromosome 8q which are associated with metastatic outcome, thus the role of PLECTIN in the context of UM progression can also be investigated. Functional work elucidating the expression of the protein with Western blots followed up with invasion and migration assays will help to demonstrate if there is any link with tumour progression.

The role of chromatin modifier in cancer is beginning to emerge and the role of the SWI/SNF complex genes will help uncover new insights into the dysregulation of gene and pathways in UMs. SWI/SNF subunits utilize energy from ATP to change the configuration of the chromatin by shuffling nucleosomes along DNA which regulates the accessibility of DNA to other proteins [391]. Mutations affecting the components of the complex can thus impair a wide range of effects. In addition, the relationship between BAP1, a component of the PRC1 complex and the chromatin remodelling complex could present an interesting area of research. Functional work to understand the effect of these mutations can be performed with CRISPR gene editing technology to observe the effects of these mutations. Finally, studying the role of small non-coding RNA and methylation in UM could further add knowledge in understanding the mechanistic insights of UM pathogenesis. Understanding the mechanistic basis of these genetic factors will help develop new range of therapeutics and help improve patient outcomes.

References

1. Andreoli, M. T., Mieler, W. F. & Leiderman, Y. I. Epidemiological trends in uveal melanoma. *Br. J. Ophthalmol.* **99,** 1550–1553 (2015).
2. Damato, E. M. & Damato, B. E. Detection and Time to Treatment of Uveal Melanoma in the United Kingdom: An Evaluation of 2384 Patients. *Ophthalmology* **119,** 1582–1589 (2012).
3. Shields, C. L. *et al.* Metastasis of Uveal Melanoma Millimeter-by-Millimeter in 8033 Consecutive Eyes. *Arch. Ophthalmol.* **127,** 989 (2009).
4. McLaughlin, C. C. *et al.* Incidence of noncutaneous melanomas in the U.S. *Cancer* **103,** 1000–1007 (2005).
5. Shields, C. L., Kaliki, S., Furuta, M., Mashayekhi, A. & Shields, J. A. Clinical spectrum and prognosis of uveal melanoma based on age at presentation in 8,033 cases. *Retina* **32,** 1363–1372 (2012).
6. Emmett T. Cunningham, Jr., J. B. C. in *General Ophthalmology* at <http://www.oculist.net/others/ebook/generalophthal/server-java/arknoid/amed/vaughan/co_chapters/ch007/ch007_print_chapter.html>
7. Eskelin, S. & Kivelä, T. Mode of presentation and time to treatment of uveal melanoma in Finland. *Br. J. Ophthalmol.* **86,** 333–8 (2002).
8. Garcia, M. D., Salomao, D. R., Marmorstein, A. D. & Pulido, J. S. Histopathologic Findings in the Areas of Orange Pigment Overlying Choroidal Melanomas. *Transl. Vis. Sci. Technol.* **5,** 4 (2016).
9. Hashmi, F., Rojanaporn, D., Kaliki, S. & Shields, C. L. Orange Pigment Sediment Overlying Small Choroidal Melanoma. *Arch. Ophthalmol.* **130,** 937 (2012).
10. Shields, J. A., Rodrigues, M. M., Sarin, L. K., Tasman, W. S. & Annesley, W. H. Lipofuscin pigment over benign and malignant choroidal tumors. *Trans. Sect. Ophthalmol. Am. Acad. Ophthalmol. Otolaryngol.* **81,** 871–81
11. Fraser, S. & Steel, D. Retinal detachment. *BMJ Clin. Evid.* **2010,** (2010).
12. Shields, C. L., Kaliki, S., Furuta, M. & Shields, J. A. Diffuse versus nondiffuse small (≤ 3 MM thickness) choroidal melanoma: comparative analysis in 1,751 cases. The 2012 F. Phinizy Calhoun lecture. *Retina* **33,** 1763–76 (2013).
13. Finger, P. T. & 7th Edition, AJCC-UICC Ophthalmic Oncology Task Force. The 7th edition AJCC staging system for eye cancer: an international language for ophthalmic oncology. *Arch. Pathol. Lab. Med.* **133,** 1197–8 (2009).
14. Khan, S. *et al.* Clinical and Pathologic Characteristics of Biopsy-Proven Iris Melanoma. *Arch. Ophthalmol.* **130,** 57 (2012).
15. Shields, C. L. *et al.* American Joint Committee on Cancer Classification of Posterior Uveal Melanoma (Tumor Size Category) Predicts Prognosis in 7731 Patients. *Ophthalmology* **120,** 2066–2071 (2013).
16. Accuracy of diagnosis of choroidal melanomas in the Collaborative Ocular Melanoma Study. COMS report no. 1. *Arch. Ophthalmol. (Chicago, Ill. 1960)* **108,** 1268–73 (1990).
17. Shields, J. A., Augsburger, J. J., Brown, G. C. & Stephens, R. F. The differential diagnosis of posterior uveal melanoma. *Ophthalmology* **87,** 518–22 (1980).
18. Shields, J. A., Mashayekhi, A., Ra, S. & Shields, C. L. Pseudomelanomas of the

posterior uveal tract: the 2006 Taylor R. Smith Lecture. *Retina* **25,** 767–71 (2005).
19. van, J. G. M. *et al.* in *Melanoma - From Early Detection to Treatment* (InTech, 2013). doi:10.5772/53631
20. Campbell RJ, Sobin LH, Z. L. & W. H. O. in (Springer: Berlin ; London, 1998). at <https://link.springer.com/book/10.1007%2F978-3-642-72163-2>
21. Burnier, M. N., McLean, I. W. & Gamel, J. W. Immunohistochemical evaluation of uveal melanocytic tumors. Expression of HMB-45, S-100 protein, and neuron-specific enolase. *Cancer* **68,** 809–14 (1991).
22. Barak, A. *et al.* Decreased prevalence of asymptomatic choroidal metastasis in disseminated breast and lung cancer: argument against screening. *Br. J. Ophthalmol.* **91,** 74–5 (2007).
23. Fenton, S., Kemp, E. G. & Harnett, A. N. Screening for ophthalmic involvement in asymptomatic patients with metastatic breast carcinoma. *Eye* **18,** 38–40 (2004).
24. Collaborative Ocular Melanoma Study Group. Assessment of metastatic disease status at death in 435 patients with large choroidal melanoma in the Collaborative Ocular Melanoma Study (COMS): COMS report no. 15. *Arch. Ophthalmol. (Chicago, Ill. 1960)* **119,** 670–6 (2001).
25. Dithmar, S., Diaz, C. E. & Grossniklaus, H. E. Intraocular melanoma spread to regional lymph nodes: report of two cases. *Retina* **20,** 76–9 (2000).
26. Kath, R. *et al.* Prognosis and treatment of disseminated uveal melanoma. *Cancer* **72,** 2219–23 (1993).
27. Mouriaux, F., Diorio, C., Bergeron, D., Berchi, C. & Rousseau, A. Liver Function Testing Is Not Helpful for Early Diagnosis of Metastatic Uveal Melanoma. *Ophthalmology* **119,** 1590–1595 (2012).
28. de Waard-Siebinga, I., Hilders, C. G., Hansen, B. E., van Delft, J. L. & Jager, M. J. HLA expression and tumor-infiltrating immune cells in uveal melanoma. *Graefes Arch. Clin. Exp. Ophthalmol.* **234,** 34–42 (1996).
29. Mäkitie, T., Summanen, P., Tarkkanen, A. & Kivelä, T. Tumor-infiltrating macrophages (CD68(+) cells) and prognosis in malignant uveal melanoma. *Invest. Ophthalmol. Vis. Sci.* **42,** 1414–21 (2001).
30. Mäkitie, T., Summanen, P., Tarkkanen, A. & Kivelä, T. Microvascular loops and networks as prognostic indicators in choroidal and ciliary body melanomas. *J. Natl. Cancer Inst.* **91,** 359–67 (1999).
31. Bronkhorst, I. H. G. *et al.* Detection of M2-Macrophages in Uveal Melanoma and Relation with Survival. *Investig. Opthalmology Vis. Sci.* **52,** 643 (2011).
32. Aalto, Y., Eriksson, L., Seregard, S., Larsson, O. & Knuutila, S. Concomitant loss of chromosome 3 and whole arm losses and gains of chromosome 1, 6, or 8 in metastasizing primary uveal melanoma. *Invest. Ophthalmol. Vis. Sci.* **42,** 313–7 (2001).
33. Sisley, K. *et al.* Abnormalities of chromosomes 3 and 8 in posterior uveal melanoma correlate with prognosis. *Genes. Chromosomes Cancer* **19,** 22–8 (1997).
34. Prescher, G. *et al.* Prognostic implications of monosomy 3 in uveal melanoma. *Lancet (London, England)* **347,** 1222–5 (1996).
35. White, V. A., Chambers, J. D., Courtright, P. D., Chang, W. Y. & Horsman, D. E. Correlation of cytogenetic abnormalities with the outcome of patients with uveal melanoma. *Cancer* **83,** 354–9 (1998).
36. Scholes, A. G. M. *et al.* Monosomy 3 in uveal melanoma: correlation with clinical and

histologic predictors of survival. *Invest. Ophthalmol. Vis. Sci.* **44,** 1008–11 (2003).
37. Kilic, E. *et al.* Concurrent Loss of Chromosome Arm 1p and Chromosome 3 Predicts a Decreased Disease-Free Survival in Uveal Melanoma Patients. *Investig. Opthalmology Vis. Sci.* **46,** 2253 (2005).
38. Ehlers, J. P., Worley, L., Onken, M. D. & Harbour, J. W. Integrative Genomic Analysis of Aneuploidy in Uveal Melanoma. *Clin. Cancer Res.* **14,** 115–122 (2008).
39. Parrella, P., Sidransky, D. & Merbs, S. L. Allelotype of posterior uveal melanoma: implications for a bifurcated tumor progression pathway. *Cancer Res.* **59,** 3032–7 (1999).
40. Kilic, E. *et al.* Clinical and Cytogenetic Analyses in Uveal Melanoma. *Investig. Opthalmology Vis. Sci.* **47,** 3703 (2006).
41. Trolet, J. *et al.* Genomic Profiling and Identification of High-Risk Uveal Melanoma by Array CGH Analysis of Primary Tumors and Liver Metastases. *Investig. Opthalmology Vis. Sci.* **50,** 2572 (2009).
42. Onken, M. D., Worley, L. A., Ehlers, J. P. & Harbour, J. W. Gene Expression Profiling in Uveal Melanoma Reveals Two Molecular Classes and Predicts Metastatic Death. *CANCER Res.* **64,** 7205–7209 (2004).
43. Onken, M. D. *et al.* Collaborative Ocular Oncology Group Report Number 1: Prospective Validation of a Multi-Gene Prognostic Assay in Uveal Melanoma. *Ophthalmology* **119,** 1596–1603 (2012).
44. Worley, L. A. *et al.* Transcriptomic versus Chromosomal Prognostic Markers and Clinical Outcome in Uveal Melanoma. *Clin. Cancer Res.* **13,** 1466–1471 (2007).
45. Onken, M. D., Worley, L. A., Dávila, R. M., Char, D. H. & Harbour, J. W. Prognostic Testing in Uveal Melanoma by Transcriptomic Profiling of Fine Needle Biopsy Specimens. *J. Mol. Diagnostics* **8,** 567–573 (2006).
46. Field, M. G. *et al.* Epigenetic reprogramming and aberrant expression of PRAME are associated with increased metastatic risk in Class 1 and Class 2 uveal melanomas. *Oncotarget* **7,** 59209–59219 (2016).
47. Bechrakis, N. E., Bornfeld, N., Zöller, I. & Foerster, M. H. Iodine 125 plaque brachytherapy versus transscleral tumor resection in the treatment of large uveal melanomas. *Ophthalmology* **109,** 1855–61 (2002).
48. Caminal, J. M. *et al.* Transscleral resection without hypotensive anaesthesia vs iodine-125 plaque brachytherapy in the treatment of choroidal melanoma. *Eye* **30,** 833–842 (2016).
49. Seibel, I. *et al.* Local Recurrence After Primary Proton Beam Therapy in Uveal Melanoma: Risk Factors, Retreatment Approaches, and Outcome. *Am. J. Ophthalmol.* **160,** 628–636 (2015).
50. Mishra, K. K. *et al.* Long-term Results of the UCSF-LBNL Randomized Trial: Charged Particle With Helium Ion Versus Iodine-125 Plaque Therapy for Choroidal and Ciliary Body Melanoma. *Int. J. Radiat. Oncol.* **92,** 376–383 (2015).
51. Turcotte, S., Bergeron, D., Rousseau, A. P. & Mouriaux, F. Primary transpupillary thermotherapy for choroidal indeterminate melanocytic lesions. *Can. J. Ophthalmol. / J. Can. d'Ophtalmologie* **49,** 464–467 (2014).
52. Mashayekhi, A. *et al.* Primary Transpupillary Thermotherapy for Choroidal Melanoma in 391 Cases: Importance of Risk Factors in Tumor Control. *Ophthalmology* **122,** 600–609 (2015).
53. Harbour, J. W., Meredith, T. A., Thompson, P. A. & Gordon, M. E. Transpupillary

thermotherapy versus plaque radiotherapy for suspected choroidal melanomas. *Ophthalmology* **110,** 2207–2214 (2003).
54. Puusaari, I., Heikkonen, J., Summanen, P., Tarkkanen, A. & Kivelä, T. Iodine brachytherapy as an alternative to enucleation for large uveal melanomas. *Ophthalmology* **110,** 2223–2234 (2003).
55. Diener-West, M. *et al.* The COMS randomized trial of iodine 125 brachytherapy for choroidal melanoma, III: initial mortality findings. COMS Report No. 18. *Arch. Ophthalmol. (Chicago, Ill. 1960)* **119,** 969–82 (2001).
56. Desjardins L, Dorval T, Levy C, et al. Etude randomisée de chimiothérapie adjuvante par le deticène dans le mélanome choroïdien [Randomized study of adjuvant chemotherapy with dacarbazine in choroidal melanoma]. *Ophtalmologie* **12,** 168–173 (1998).
57. Aaberg, T. *et al.* Current clinical practice: differential management of uveal melanoma in the era of molecular tumor analyses. *Clin. Ophthalmol.* **8,** 2449 (2014).
58. Landreville, S. *et al.* Histone deacetylase inhibitors induce growth arrest and differentiation in uveal melanoma. *Clin. Cancer Res.* **18,** 408–16 (2012).
59. Swerdlow, A. J. Epidemiology of eye cancer in adults in England and Wales, 1962-1977. *Am. J. Epidemiol.* **118,** 294–300 (1983).
60. Osterlind, A. Trends in incidence of ocular malignant melanoma in Denmark 1943-1982. *Int. J. cancer* **40,** 161–4 (1987).
61. Hakulinen, T., Teppo, L. & Saxén, E. Cancer of the eye, a review of trends and differentials. *World Health Stat. Q.* **31,** 143–58 (1978).
62. Yousef, Y. A. & Alkilany, M. Characterization, treatment, and outcome of uveal melanoma in the first two years of life. *Hematol. Oncol. Stem Cell Ther.* **8,** 1–5 (2015).
63. Stang, A., Parkin, D. M., Ferlay, J. & Jöckel, K.-H. International uveal melanoma incidence trends in view of a decreasing proportion of morphological verification. *Int. J. Cancer* **114,** 114–123 (2005).
64. Bergman, L. *et al.* Incidence of uveal melanoma in Sweden from 1960 to 1998. *Invest. Ophthalmol. Vis. Sci.* **43,** 2579–83 (2002).
65. Vidal, J. L. *et al.* [Epidemiological study of uveal melanoma in France]. *J. Fr. Ophtalmol.* **18,** 520–8 (1995).
66. Teikari, J. M. & Raivio, I. Incidence of choroidal malignant melanoma in Finland in the years 1973-1980. *Acta Ophthalmol.* **63,** 661–5 (1985).
67. Stang, A. *et al.* Population-based incidence estimates of uveal melanoma in Germany. Supplementing cancer registry data by case–control data. *Eur. J. Cancer Prev.* **15,** 165–170 (2006).
68. Koomen, E. R. *et al.* Epidemiology of Extracutaneous Melanoma in the Netherlands. *Cancer Epidemiol. Biomarkers Prev.* **19,** 1453–1459 (2010).
69. Virgili, G. *et al.* Incidence of Uveal Melanoma in Europe. *Ophthalmology* **114,** 2309–2315.e2 (2007).
70. Scotto, J., Fraumeni, J. F. & Lee, J. A. Melanomas of the eye and other noncutaneous sites: epidemiologic aspects. *J. Natl. Cancer Inst.* **56,** 489–91 (1976).
71. Keenan, T. D. L., Yeates, D. & Goldacre, M. J. Uveal melanoma in England: trends over time and geographical variation. *Br. J. Ophthalmol.* **96,** 1415–1419 (2012).
72. Seddon, J. M. *et al.* Host factors, UV radiation, and risk of uveal melanoma. A case-control study. *Arch. Ophthalmol. (Chicago, Ill. 1960)* **108,** 1274–80 (1990).
73. Margo, C. E., Mulla, Z. & Billiris, K. Incidence of surgically treated uveal melanoma by

race and ethnicity. *Ophthalmology* **105,** 1087–1090 (1998).
74. Vajdic, C. M. *et al.* Incidence of ocular melanoma in Australia from 1990 to 1998. *Int. J. Cancer* **105,** 117–122 (2003).
75. Kuo, P. K., Puliafito, C. A., Wang, K. M., Liu, H. S. & Wu, B. F. Uveal melanoma in China. *Int. Ophthalmol. Clin.* **22,** 57–71 (1982).
76. Cheng, C.-Y. & Hsu, W.-M. Incidence of eye cancer in Taiwan: an 18-year review. *Eye* **18,** 152–158 (2004).
77. Lock-Andersen, J., Wulf, H. C. & Knudstorp, N. D. Interdependence of eye and hair colour, skin type and skin pigmentation in a Caucasian population. *Acta Derm. Venereol.* **78,** 214–9 (1998).
78. Vajdic, C. M. *et al.* Eye color and cutaneous nevi predict risk of ocular melanoma in Australia. *Int. J. Cancer* **92,** 906–912 (2001).
79. Weis, E., Shah, C. P., Lajous, M., Shields, J. A. & Shields, C. L. The Association Between Host Susceptibility Factors and Uveal Melanoma. *Arch. Ophthalmol.* **124,** 54 (2006).
80. Guénel, P. *et al.* Occupational risk factors, ultraviolet radiation, and ocular melanoma: a case-control study in France. *Cancer Causes Control* **12,** 451–9 (2001).
81. Gallagher, R. P. *et al.* Risk factors for ocular melanoma: Western Canada Melanoma Study. *J. Natl. Cancer Inst.* **74,** 775–8 (1985).
82. Rootman, J. & Gallagher, R. P. Color as a Risk Factor in Iris Melanoma. *Am. J. Ophthalmol.* **98,** 558–561 (1984).
83. Tucker, M. A. *et al.* Sunlight Exposure as Risk Factor for Intraocular Malignant Melanoma. *N. Engl. J. Med.* **313,** 789–792 (1985).
84. Wakamatsu, K., Hu, D.-N., McCormick, S. A. & Ito, S. Characterization of melanin in human iridal and choroidal melanocytes from eyes with various colored irides. *Pigment Cell Melanoma Res.* **21,** 97–105 (2007).
85. De Leeuw, S. M. *et al.* Melanin content of cultured human melanocytes and UV-induced cytotoxicity. *J. Photochem. Photobiol. B Biol.* **61,** 106–113 (2001).
86. Lerman, S. in *Clinical Light Damage to the Eye* 183–215 (Springer New York, 1987). doi:10.1007/978-1-4612-4704-3_10
87. Li, W., Judge, H., Gragoudas, E. S., Seddon, J. M. & Egan, K. M. Patterns of Tumor Initiation in Choroidal Melanoma. *Cancer Res.* **60,** (2000).
88. Moan, J. *et al.* Time trends and latitude dependence of uveal and cutaneous malignant melanoma induced by solar radiation. *Dermatoendocrinol.* **2,** 3–8 (2010).
89. Singh, A. D., Turell, M. E. & Topham, A. K. Uveal Melanoma: Trends in Incidence, Treatment, and Survival. *Ophthalmology* **118,** 1881–1885 (2011).
90. Dolin, P. J. & Johnson, G. J. Solar ultraviolet radiation and ocular disease: a review of the epidemiological and experimental evidence. *Ophthalmic Epidemiol.* **1,** 155–64 (1994).
91. Holly, E. A., Aston, D. A., Char, D. H., Kristiansen, J. J. & Ahn, D. K. Uveal melanoma in relation to ultraviolet light exposure and host factors. *Cancer Res.* **50,** 5773–7 (1990).
92. Pane, A. R. & Hirst, L. W. Ultraviolet light exposure as a risk factor for ocular melanoma in Queensland, Australia. *Ophthalmic Epidemiol.* **7,** 159–67 (2000).
93. Vajdic, C. M. *et al.* Sun exposure predicts risk of ocular melanoma in Australia. *Int. J. Cancer* **101,** 175–182 (2002).
94. Holly, E. A., Aston, D. A., Ahn, D. K., Kristiansen, J. J. & Char, D. H. Uveal melanoma, hormonal and reproductive factors in women. *Cancer Res.* **51,** 1370–2 (1991).
95. Hartge, P. *et al.* Case-control study of female hormones and eye melanoma. *Cancer*

Res. **49,** 4622–5 (1989).
96. Egan, K. M., Walsh, S. M., Seddon, J. M. & Gragoudas, E. S. An evaluation of the influence of reproductive factors on the risk of metastases from uveal melanoma. *Ophthalmology* **100,** 1160–5; discussion 1166 (1993).
97. Egan, K. M., Gragoudas, E. S., Seddon, J. M. & Walsh, S. M. Smoking and the risk of early metastases from uveal melanoma. *Ophthalmology* **99,** 537–41 (1992).
98. Albert, D. M. The association of viruses with urveal melanoma. *Trans. Am. Ophthalmol. Soc.* **77,** 367–421 (1979).
99. Albert, D. M., Lahav, M., Colby, E. & al., et. The role of viruses in the development of ocular tumors. *Ophthalmology* **87,** 1219–25 (1980).
100. Albert, D. M. *et al.* Increased incidence of choroidal malignant melanoma occurring in a single population of chemical workers. *Am. J. Ophthalmol.* **89,** 323–37 (1980).
101. Shildkrot, Y. *et al.* Socioeconomic Factors and Diagnosis of Uveal Melanoma in the Mid-Southern United States. *Curr. Eye Res.* **36,** 824–830 (2011).
102. Weis, E., Shah, C. P., Lajous, M., Shields, J. A. & Shields, C. L. The Association of Cutaneous and Iris Nevi with Uveal Melanoma: A Meta-analysis. *Ophthalmology* **116,** 536–543.e2 (2009).
103. Lynch, H. T. & Krush, A. J. Heredity and malignant melanoma: implications for early cancer detection. *Can. Med. Assoc. J.* **99,** 17–21 (1968).
104. Walker, J. P., Weiter, J. J., Albert, D. M., Osborn, E. L. & Weichselbaum, R. R. Uveal malignant melanoma in three generations of the same family. *Am. J. Ophthalmol.* **88,** 723–6 (1979).
105. GREER, C. H. Congenital Melanoma of the Anterior Uvea. *Arch. Ophthalmol.* **76,** 77–78 (1966).
106. Greene, M. H. *et al.* The Familial Occurrence of Cutaneous Melanoma, Intraocular Melanoma, and the Dysplastic Nevus Syndrome. *Am. J. Ophthalmol.* **96,** 238–245 (1983).
107. Canning, C. R. & Hungerford, J. Familial uveal melanoma. *Br. J. Ophthalmol.* **72,** 241–3 (1988).
108. Lynch, H. T., Anderson, D. E. & Krush, A. J. Heredity and intraocular malignant melanoma:Study of two families and review of forty-five cases. *Cancer* **21,** 119–125 (1968).
109. Singh, A. D. *et al.* Familial Uveal Melanoma, III. *Arch. Ophthalmol.* **114,** 1101 (1996).
110. Wiesner, T. *et al.* Germline mutations in BAP1 predispose to melanocytic tumors. *Nat. Genet.* **43,** 1018–21 (2011).
111. Testa, J. R. *et al.* Germline BAP1 mutations predispose to malignant mesothelioma. *Nat. Genet.* **43,** 1022–5 (2011).
112. Mort, R. L., Jackson, I. J. & Patton, E. E. The melanocyte lineage in development and disease. *Development* **142,** 620–32 (2015).
113. Praetorius, C. *et al.* A Polymorphism in IRF4 Affects Human Pigmentation through a Tyrosinase-Dependent MITF/TFAP2A Pathway. *Cell* **155,** 1022–1033 (2013).
114. Onken, M. D. *et al.* Functional gene expression analysis uncovers phenotypic switch in aggressive uveal melanomas. *Cancer Res.* **66,** 4602–9 (2006).
115. Harbour, J. W. *et al.* Frequent Mutation of BAP1 in Metastasizing Uveal Melanomas. *Science (80-.).* **330,** 1410–1413 (2010).
116. Matatall, K. A. *et al.* BAP1 deficiency causes loss of melanocytic cell identity in uveal melanoma. *BMC Cancer* **13,** 371 (2013).

117. Ferguson, R. *et al.* Genetic markers of pigmentation are novel risk loci for uveal melanoma. *Sci. Rep.* **6,** 31191 (2016).
118. Mobuchon, L. et al. A GWAS in uveal melanoma identifies risk polymorphisms in the CLPTM1L locus. *npj Genomic Med.* **doi:10.103,** (2017).
119. Brodeur, G. M. & Hogarty, M. D. *The Genetic Basis of Human Cancer*. (McGraw-Hill, New York, 1998).
120. Mitelman, F., Johansson, B. & Mertens, F. *Catalog of Chromosome Aberrations in Cancer Vol 2*. (WileyLiss, New York, 1994).
121. ROWLEY, J. D. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature* **243,** 290–293 (1973).
122. Stranger, B. E. *et al.* Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. *Science (80-.).* **315,** 848–853 (2007).
123. Little, C. D., Nau, M. M., Carney, D. N., Gazdar, A. F. & Minna, J. D. Amplification and expression of the c-myc oncogene in human lung cancer cell lines. *Nature* **306,** 194–196 (1983).
124. Kallioniemi, O. P. *et al.* ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 5321–5 (1992).
125. Di Fiore, P. P. *et al.* erbB-2 is a potent oncogene when overexpressed in NIH/3T3 cells. *Science* **237,** 178–82 (1987).
126. Friend, S. H. *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323,** 643–646 (1986).
127. Li, J. *et al.* PTEN, a Putative Protein Tyrosine Phosphatase Gene Mutated in Human Brain, Breast, and Prostate Cancer. *Science (80-.).* **275,** (1997).
128. Harbour, J. W. in *Ocular Oncology* (ed. Albert DM, P. A.) 1–18 (Marcel Dekker; NY, USA, 2003).
129. Landreville, S., Agapova, O. A. & Harbour, J. W. Emerging insights into the molecular pathogenesis of uveal melanoma. *Future Oncol.* **4,** 629–36 (2008).
130. Cross, N. A., Murray, A. K., Rennie, I. G., Ganesh, A. & Sisley, K. Instability of microsatellites is an infrequent event in uveal melanoma. *Melanoma Res.* **13,** 435–40 (2003).
131. Papadopoulos, S. *et al.* Assessment of genomic instability in breast cancer and uveal melanoma by random amplified polymorphic DNA analysis. *Int. J. Cancer* **99,** 193–200 (2002).
132. Sisley, K. *et al.* Cytogenetic findings in six posterior uveal melanomas: involvement of chromosomes 3, 6, and 8. *Genes. Chromosomes Cancer* **2,** 205–9 (1990).
133. Tschentscher, F., Prescher, G., Zeschnigk, M., Horsthemke, B. & Lohmann, D. R. Identification of chromosomes 3, 6, and 8 aberrations in uveal melanoma by microsatellite analysis in comparison to comparative genomic hybridization. *Cancer Genet. Cytogenet.* **122,** 13–7 (2000).
134. Hughes, S. *et al.* Microarray comparative genomic hybridisation analysis of intraocular uveal melanomas identifies distinctive imbalances associated with loss of chromosome 3. *Br. J. Cancer* **93,** 1191–6 (2005).
135. Prescher, G., Bornfeld, N., Friedrichs, W., Seeber, S. & Becher, R. Cytogenetics of twelve cases of uveal melanoma and patterns of nonrandom anomalies and isochromosome formation. *Cancer Genet. Cytogenet.* **80,** 40–6 (1995).
136. Horsman, D. E., Sroka, H., Rootman, J. & White, V. A. Monosomy 3 and

isochromosome 8q in a uveal melanoma. *Cancer Genet. Cytogenet.* **45,** 249–53 (1990).
137. Association for Research in Vision and Ophthalmology., Y., Eriksson, L., Seregard, S., Larsson, O. & Knuutila, S. *Investigative ophthalmology & visual science. Br J Cancer* **82,** (C.V. Mosby Co, 1977).
138. Shields, C. L. *et al.* Prognosis of Uveal Melanoma in 500 Cases Using Genetic Testing of Fine-Needle Aspiration Biopsy Specimens. *Ophthalmology* **118,** 396–401 (2011).
139. Abdel-Rahman, M. H. *et al.* Frequency, molecular pathology and potential clinical significance of partial chromosome 3 aberrations in uveal melanoma. *Mod. Pathol.* **24,** 954–962 (2011).
140. Mensink, H. W. *et al.* Chromosome 3 Intratumor Heterogeneity in Uveal Melanoma. *Investig. Opthalmology Vis. Sci.* **50,** 500 (2009).
141. Lynn Schoenfield, James Pettay, Raymond R. Tubbs, and A. D. S. Variation of Monosomy 3 Status Within Uveal Melanoma. *Arch. Pathol. Lab. Med.* **133,** 1219–1222 (2009).
142. Sisley, K. *et al.* Association of specific chromosome alterations with tumour phenotype in posterior uveal melanoma. *Br. J. Cancer* **82,** 330–8 (2000).
143. Kilic, E. *et al.* Increased expression of p73Δex2 transcript in uveal melanoma with loss of chromosome 1p. *Melanoma Res.* **18,** 208–213 (2008).
144. Damato, B. *et al.* Cytogenetics of Uveal Melanoma. *Ophthalmology* **114,** 1925–1931.e1 (2007).
145. Onken, M. D., Worley, L. A. & Harbour, J. W. A Metastasis Modifier Locus on Human Chromosome 8p in Uveal Melanoma Identified by Integrative Genomic Analysis. *Clin. Cancer Res.* **14,** 3737–3745 (2008).
146. Prescher, G., Bornfeld, N. & Becher, R. Two subclones in a case of uveal melanoma. Relevance of monosomy 3 and multiplication of chromosome 8q. *Cancer Genet. Cytogenet.* **77,** 144–6 (1994).
147. Horsman, D. E. & White, V. A. Cytogenetic analysis of uveal melanoma. Consistent occurrence of monosomy 3 and trisomy 8q. *Cancer* **71,** 811–9 (1993).
148. van den Bosch, T. *et al.* Higher Percentage of FISH-Determined Monosomy 3 and 8q Amplification in Uveal Melanoma Cells relate to Poor Patient Prognosis. *Investig. Opthalmology Vis. Sci.* **53,** 2668 (2012).
149. Harbour, J. W. The genetics of uveal melanoma: an emerging framework for targeted therapy. *Pigment Cell Melanoma Res.* **25,** 171–81 (2012).
150. Prescher, G., Bornfeld, N. & Becher, R. Nonrandom chromosomal abnormalities in primary uveal melanoma. *J. Natl. Cancer Inst.* **82,** 1765–9 (1990).
151. Scholes, A. G. *et al.* Loss of heterozygosity on chromosomes 3, 9, 13, and 17, including the retinoblastoma locus, in uveal melanoma. *Invest. Ophthalmol. Vis. Sci.* **42,** 2472–7 (2001).
152. White, J. S., McLean, I. W., Becker, R. L., Director-Myska, A. E. & Nath, J. Correlation of comparative genomic hybridization results of 100 archival uveal melanomas with patient survival. *Cancer Genet. Cytogenet.* **170,** 29–39 (2006).
153. Mensink, H. W. *et al.* Molecular cytogenetic analysis of archival uveal melanoma with known clinical outcome. *Cancer Genet. Cytogenet.* **181,** 108–11 (2008).
154. Vajdic, C. M. *et al.* Chromosomal gains and losses in ocular melanoma detected by comparative genomic hybridization in an Australian population-based study. *Cancer Genet. Cytogenet.* **144,** 12–7 (2003).

155. Häusler, T. *et al.* Loss of heterozygosity of 1p in uveal melanomas with monosomy 3. *Int. J. Cancer* **116,** 909–913 (2005).
156. van Gils, W. *et al.* Regional deletion and amplification on chromosome 6 in a uveal melanoma case without abnormalities on chromosomes 1p, 3 and 8. *Melanoma Res.* **18,** 10–15 (2008).
157. Speicher, M. R. *et al.* Chromosomal gains and losses in uveal melanomas detected by comparative genomic hybridization. *Cancer Res.* **54,** 3817–23 (1994).
158. Parrella, P., Caballero, O. L., Sidransky, D. & Merbs, S. L. Detection of c-myc amplification in uveal melanoma by fluorescent in situ hybridization. *Invest. Ophthalmol. Vis. Sci.* **42,** 1679–84 (2001).
159. Ohta, M. *et al.* Deletion mapping of chromosome region 9p21-p22 surrounding theCDKN2 locus in melanoma. *Int. J. Cancer* **65,** 762–767 (1996).
160. Merbs, S. L. & Sidransky, D. Analysis of p16 (CDKN2/MTS-1/INK4A) alterations in primary sporadic uveal melanoma. *Investig. Ophthalmol. Vis. Sci.* **40,** 779–783 (1999).
161. van der Velden, P. A. *et al.* Promoter hypermethylation: a common cause of reduced p16(INK4a) expression in uveal melanoma. *Cancer Res.* **61,** 5303–6 (2001).
162. Höglund, M. *et al.* Dissecting karyotypic patterns in malignant melanomas: Temporal clustering of losses and gains in melanoma karyotypic evolution. *Int. J. Cancer* **108,** 57–65 (2004).
163. Horsthemke, B., Prescher, G., Bornfeld, N. & Becher, R. Loss of chromosome 3 alleles and multiplication of chromosome 8 alleles in uveal melanoma. *Genes. Chromosomes Cancer* **4,** 217–21 (1992).
164. Blasi, M. A. *et al.* 3p13 region: a possible location of a tumor suppressor gene involved in uveal melanoma. *Cancer Genet. Cytogenet.* **108,** 81–3 (1999).
165. Cross, N. A. *et al.* Multiple locations on chromosome 3 are the targets of specific deletions in uveal melanoma. *Eye* **20,** 476–481 (2006).
166. Tschentscher, F. *et al.* Partial deletions of the long and short arm of chromosome 3 point to two tumor suppressor genes in uveal melanoma. *Cancer Res.* **61,** 3439–42 (2001).
167. Parrella, P., Fazio, V. M., Gallo, A. P., Sidransky, D. & Merbs, S. L. Fine mapping of chromosome 3 in uveal melanoma: identification of a minimal region of deletion on chromosomal arm 3p25.1-p25.2. *Cancer Res.* **63,** 8507–10 (2003).
168. Myatt, N. *et al.* Abnormalities of the transforming growth factor-beta pathway in ocular melanoma. *J. Pathol.* **192,** 511–518 (2000).
169. Sisley, K., Curtis, D., Rennie, I. G. & Rees, R. C. Loss of heterozygosity of the thyroid hormone receptor B in posterior uveal melanoma. *Melanoma Res.* **3,** 457–61 (1993).
170. Zeschnigk, M. *et al.* Methylation Analysis of Several Tumour Suppressor Genes Shows a Low Frequency of Methylation of *CDKN2A* and *RARB* in Uveal Melanomas. *Comp. Funct. Genomics* **4,** 329–336 (2003).
171. van Essen, T. H. *et al.* Prognostic parameters in uveal melanoma and their association with BAP1 expression. *Br. J. Ophthalmol.* **98,** 1738–43 (2014).
172. Shaw, A. T. *et al.* Crizotinib versus Chemotherapy in Advanced ALK -Positive Lung Cancer. *N. Engl. J. Med.* **368,** 2385–2394 (2013).
173. Onida, F. *et al.* Characteristics and outcome of patients with Philadelphia chromosome negative,bcr/abl negative chronic myelogenous leukemia. *Cancer* **95,** 1673–1684 (2002).
174. Tomlins, S. A. *et al.* Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes

in Prostate Cancer. *Science (80-.).* **310,** 644–648 (2005).
175. Tomlins, S. A. *et al.* TMPRSS2:ETV4 Gene Fusions Define a Third Molecular Subtype of Prostate Cancer. *Cancer Res.* **66,** 3396–3400 (2006).
176. Mitelman F, J. B. and M. F. (Eds. . Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. at <https://cgap.nci.nih.gov/Chromosomes/Mitelman>
177. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **7,** 233–245 (2007).
178. Chizuka, A. *et al.* Difference between hematological malignancy and Solid tumor research articles published in four major medical journals. *Leukemia* **20,** 1655–1657 (2006).
179. Jividen, K. & Li, H. Chimeric RNAs generated by intergenic splicing in normal and cancer cells. *Genes, Chromosom. Cancer* **53,** 963–971 (2014).
180. Royer-Bertrand, B. *et al.* Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing. *Am. J. Hum. Genet.* **99,** 1190–1198 (2016).
181. Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* **545,** 175–180 (2017).
182. Yoshihara, K. *et al.* The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **34,** 4845–4854 (2015).
183. Kumar-Sinha, C., Kalyana-Sundaram, S. & Chinnaiyan, A. M. Landscape of gene fusions in epithelial cancers: seq and ye shall find. *Genome Med.* **7,** 129 (2015).
184. Maat, W. *et al.* The Heterogeneous Distribution of Monosomy 3 in Uveal Melanomas Implications for Prognostication Based on Fine-Needle Aspiration Biopsies. *Arch Pathol Lab Med—Vol* **131,** (2007).
185. Zuidervaart, W. *et al.* Gene expression profiling identifies tumour markers potentially playing a role in uveal melanoma development. *Br. J. Cancer* **89,** 1914–1919 (2003).
186. Thies, A., Mangold, U., Moll, I. & Schumacher, U. PAS-positive loops and networks as a prognostic indicator in cutaneous malignant melanoma. *J. Pathol.* **195,** 537–542 (2001).
187. Tschentscher, F. *et al.* Tumor classification based on gene expression profiling shows that uveal melanomas with and without monosomy 3 represent two distinct entities. *Cancer Res.* **63,** 2578–84 (2003).
188. van Gils, W. *et al.* Gene Expression Profiling in Uveal Melanoma: Two Regions on 3p Related to Prognosis. *Investig. Opthalmology Vis. Sci.* **49,** 4254 (2008).
189. Petrausch, U. *et al.* Significance of gene expression analysis in uveal melanoma in comparison to standard risk factors for risk assessment of subsequent metastases. *Eye* **22,** 997–1007 (2008).
190. Onken, M. D., Worley, L. A., Tuscan, M. D. & Harbour, J. W. An Accurate, Clinically Feasible Multi-Gene Expression Assay for Predicting Metastasis in Uveal Melanoma. *J. Mol. Diagnostics* **12,** 461–468 (2010).
191. Field, M. G. & Harbour, J. W. Recent developments in prognostic and predictive testing in uveal melanoma. *Curr. Opin. Ophthalmol.* **25,** 234–9 (2014).
192. Chang, S.-H., Worley, L. A., Onken, M. D. & Harbour, J. W. Prognostic biomarkers in uveal melanoma: evidence for a stem cell-like phenotype associated with metastasis. *Melanoma Res.* **18,** 191–200 (2008).
193. Worley, L. A., Long, M. D., Onken, M. D. & Harbour, J. W. Micro-RNAs associated with metastasis in uveal melanoma identified by multiplexed microarray profiling. *Melanoma Res.* **18,** 184–190 (2008).

194. Yan, D. *et al.* MicroRNA-34a Inhibits Uveal Melanoma Cell Proliferation and Migration through Downregulation of c-Met. *Investig. Opthalmology Vis. Sci.* **50,** 1559 (2009).
195. Radhakrishnan, A., Badhrinarayanan, N., Biswas, J. & Krishnakumar, S. Analysis of chromosomal aberration (1, 3, and 8) and association of microRNAs in uveal melanoma. *Mol. Vis.* **15,** 2146–54 (2009).
196. Chen, X. *et al.* Epigenetics, MicroRNAs, and Carcinogenesis: Functional Role of MicroRNA-137 in Uveal Melanoma. *Investig. Opthalmology Vis. Sci.* **52,** 1193 (2011).
197. Dong, F. & Lou, D. MicroRNA-34b/c suppresses uveal melanoma cell proliferation and migration through multiple targets. *Mol. Vis.* **18,** 537–46 (2012).
198. Johansson, P. *et al.* Deep sequencing of uveal melanoma identifies a recurrent mutation in PLCB4. *Oncotarget* **7,** 4624–31 (2016).
199. Furney, S. J. *et al.* SF3B1 Mutations Are Associated with Alternative Splicing in Uveal Melanoma. *Cancer Discov.* **3,** 1122–1129 (2013).
200. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–339 (2013).
201. Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat. Genet.* **44,** 1006 (2012).
202. Ikehata, H. & Ono, T. The mechanisms of UV mutagenesis. *J. Radiat. Res.* **52,** 115–25 (2011).
203. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
204. Yu, H. *et al.* Tumor suppressor and deubiquitinase BAP1 promotes DNA double-strand break repair. *Proc. Natl. Acad. Sci.* **111,** 285–290 (2014).
205. Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* **545,** 175–180 (2017).
206. Cruz, F. *et al.* Absence of BRAF and NRAS mutations in uveal melanoma. *Cancer Res.* **63,** 5761–6 (2003).
207. Polsky, D. & Cordon-Cardo, C. Oncogenes in melanoma. *Oncogene* **22,** 3087–3091 (2003).
208. Wu, H., Goel, V. & Haluska, F. G. PTEN signaling pathways in melanoma. *Oncogene* **22,** 3113–3122 (2003).
209. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science (80-.).* **339,** 1546–1558 (2013).
210. Zuidervaart, W. *et al.* Activation of the MAPK pathway is a common event in uveal melanomas although it rarely occurs through mutation of BRAF or RAS. *Br. J. Cancer* **92,** 2032–2038 (2005).
211. Calipel, A. *et al.* Extracellular Signal-regulated Kinase-dependent Proliferation Is Mediated through the Protein Kinase A/B-Raf Pathway in Human Uveal Melanoma Cells. *J. Biol. Chem.* **281,** 9238–9250 (2006).
212. Van Raamsdonk, C. D. *et al.* Mutations in GNA11 in uveal melanoma. *N. Engl. J. Med.* **363,** 2191–9 (2010).
213. Van Raamsdonk, C. D. *et al.* Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature* **457,** 599–602 (2009).
214. Van Raamsdonk, C. D., Fitch, K. R., Fuchs, H., de Angelis, M. H. & Barsh, G. S. Effects of G-protein mutations on skin color. *Nat. Genet.* **36,** 961–968 (2004).
215. Fitch, K. R. *et al.* Genetics of dark skin in mice. *Genes Dev.* **17,** 214–228 (2003).
216. Dono, M. *et al.* Mutation frequencies of GNAQ, GNA11, BAP1, SF3B1, EIF1AX and

TERT in uveal melanoma: detection of an activating mutation in the TERT gene promoter in a single case of uveal melanoma. *Br. J. Cancer* **110,** 1058–65 (2014).
217. Xu, X. *et al.* Oncogenic GNAQ and GNA11 Mutations in Uveal Melanoma in Chinese. *PLoS One* **9,** e109699 (2014).
218. Decatur, C. L. *et al.* Driver Mutations in Uveal Melanoma: Associations With Gene Expression Profile and Patient Outcomes. *JAMA Ophthalmol.* **134,** 728–33 (2016).
219. Martin, M. *et al.* Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nat. Genet.* **45,** 933–936 (2013).
220. Bauer, J. *et al.* Oncogenic GNAQ mutations are not correlated with disease-free survival in uveal melanoma. *Br. J. Cancer* **101,** 813–815 (2009).
221. Onken, M. D. *et al.* Oncogenic Mutations in GNAQ Occur Early in Uveal Melanoma. *Investig. Opthalmology Vis. Sci.* **49,** 5230 (2008).
222. Moore, A. R. *et al.* Recurrent activating mutations of G-protein-coupled receptor CYSLTR2 in uveal melanoma. *Nat. Genet.* **48,** 675–680 (2016).
223. Markby, D., Onrust, R. & Bourne, H. Separate GTP binding and GTPase activating domains of a G alpha subunit. *Science (80-.).* **262,** (1993).
224. Weber, A. *et al.* Absence of mutations of the BRAF gene and constitutive activation of extracellular-regulated kinase in malignant melanomas of the uvea. *Lab. Invest.* **83,** 1771–6 (2003).
225. Lutz, S. *et al.* The guanine nucleotide exchange factor p63RhoGEF, a specific link between Gq/11-coupled receptor signaling and RhoA. *J. Biol. Chem.* **280,** 11134–9 (2005).
226. Chua, V. *et al.* Dysregulated GPCR Signaling and Therapeutic Options in Uveal Melanoma. *Mol. Cancer Res.* **15,** (2017).
227. Onken, M. D., Worley, L. A., Tuscan, M. D. & Harbour, J. W. An accurate, clinically feasible multi-gene expression assay for predicting metastasis in uveal melanoma. *J. Mol. Diagn.* **12,** 461–8 (2010).
228. Harbour, J. W. *et al.* Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330,** 1410–3 (2010).
229. Gupta, M. P. *et al.* Clinical Characteristics of Uveal Melanoma in Patients With Germline BAP1 Mutations. *JAMA Ophthalmol.* **133,** 881 (2015).
230. Murali, R., Wiesner, T. & Scolyer, R. A. Tumours associated with BAP1 mutations. *Pathology* **45,** 116–126 (2013).
231. Koopmans, A. E. *et al.* Clinical significance of immunohistochemistry for detection of BAP1 mutations in uveal melanoma. *Mod. Pathol.* **27,** 1321–1330 (2014).
232. Carbone, M. *et al.* BAP1 cancer syndrome: malignant mesothelioma, uveal and cutaneous melanoma, and MBAITs. *J. Transl. Med.* **10,** 179 (2012).
233. Turunen, J. A. *et al.* BAP1 Germline Mutations in Finnish Patients with Uveal Melanoma. *Ophthalmology* **123,** 1112–7 (2016).
234. Jensen, D. E. *et al.* BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* **16,** 1097–112 (1998).
235. Eletr, Z. M. & Wilkinson, K. D. An Emerging Model for BAP1's Role in Regulating Cell Cycle Progression. *Cell Biochem. Biophys.* **60,** 3–11 (2011).
236. Tyagi, S., Chabes, A. L., Wysocka, J. & Herr, W. E2F Activation of S Phase Promoters via Association with HCF-1 and the MLL Family of Histone H3K4 Methyltransferases. *Mol. Cell* **27,** 107–119 (2007).

237. Yu, H. *et al.* The Ubiquitin Carboxyl Hydrolase BAP1 Forms a Ternary Complex with YY1 and HCF-1 and Is a Critical Regulator of Gene Expression. *Mol. Cell. Biol.* **30,** 5071–5085 (2010).
238. Okino, Y., Machida, Y., Frankland-Searby, S. & Machida, Y. J. BRCA1-associated Protein 1 (BAP1) Deubiquitinase Antagonizes the Ubiquitin-mediated Activation of FoxK2 Target Genes. *J. Biol. Chem.* **290,** 1580–1591 (2015).
239. Dey, A. *et al.* Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science* **337,** 1541–1546 (2012).
240. Scheuermann, J. C. *et al.* Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465,** 243–247 (2010).
241. Carbone, M. *et al.* BAP1 and cancer. *Nat. Rev. Cancer* **13,** 153–9 (2013).
242. Harbour, J. W. *et al.* Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat. Genet.* **45,** 133–135 (2013).
243. Hahn, C. N. & Scott, H. S. Spliceosome mutations in hematopoietic malignancies. *Nat. Genet.* **44,** 9–10 (2011).
244. Wang, L. *et al.* SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **365,** 2497–2506 (2011).
245. Papaemmanuil, E. *et al.* Somatic SF3B1 Mutation in Myelodysplasia with Ring Sideroblasts. *N. Engl. J. Med.* **365,** 1384–1395 (2011).
246. Yavuzyigitoglu, S. *et al. Correlation of Gene Mutation Status with Copy Number Profile in Uveal Melanoma*. *Ophthalmology* **124,** (2017).
247. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478,** 64–69 (2011).
248. Rossi, D. *et al.* Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood* **118,** 6904–6908 (2011).
249. DeBoever, C. *et al.* Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. *PLoS Comput. Biol.* **11,** e1004105 (2015).
250. Alsafadi, S. *et al.* Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.* **7,** 10615 (2016).
251. Darman, R. B. *et al.* Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Rep.* **13,** 1033–1045 (2015).
252. Martin, M. *et al.* Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nat. Genet.* **45,** 933–6 (2013).
253. Chaudhuri, J., Si, K. & Maitra, U. Function of eukaryotic translation initiation factor 1A (eIF1A) (formerly called eIF-4C) in initiation of protein synthesis. *J. Biol. Chem.* **272,** 7883–91 (1997).
254. Hunter, S. M. *et al.* Molecular profiling of low grade serous ovarian tumours identifies novel candidate driver genes. *Oncotarget* **6,** 37663–77 (2015).
255. Ewens, K. G. *et al.* Chromosome 3 Status Combined With BAP1 and EIF1AX Mutation Profiles Are Associated With Metastasis in Uveal Melanoma. *Investig. Opthalmology Vis. Sci.* **55,** 5160 (2014).
256. Kunstman, J. W. *et al.* Characterization of the mutational landscape of anaplastic thyroid cancer via whole-exome sequencing. *Hum. Mol. Genet.* **24,** 2318–2329 (2015).
257. Fekete, C. A. *et al.* The eIF1A C-terminal domain promotes initiation complex

assembly, scanning and AUG selection in vivo. *EMBO J.* **24,** 3588–601 (2005).
258. Maag, D. & Lorsch, J. R. Communication between eukaryotic translation initiation factors 1 and 1A on the yeast small ribosomal subunit. *J. Mol. Biol.* **330,** 917–24 (2003).
259. Yu, C. *et al.* Molecular network including eIF1AX, RPS7, and 14-3-3γ regulates protein translation and cell proliferation in bovine mammary epithelial cells. *Arch. Biochem. Biophys.* **564,** 142–155 (2014).
260. Griewank, K. G. *et al.* TERT promoter mutations in ocular melanoma distinguish between conjunctival and uveal tumours. *Br. J. Cancer* **109,** 497–501 (2013).
261. Mobuchon, L. *et al.* A GWAS in uveal melanoma identifies risk polymorphisms in the CLPTM1L locus. *npj Genomic Med.* **2,** 5 (2017).
262. Field, M. G. *et al.* PRAME as an Independent Biomarker for Metastasis in Uveal Melanoma. *Clin. Cancer Res.* **22,** 1234–1242 (2016).
263. Costessi, A. *et al.* The Human EKC/KEOPS Complex Is Recruited to Cullin2 Ubiquitin Ligases by the Human Tumour Antigen PRAME. *PLoS One* **7,** e42822 (2012).
264. Costessi, A. *et al.* The tumour antigen PRAME is a subunit of a Cul2 ubiquitin ligase and associates with active NFY promoters. *EMBO J.* **30,** 3786–3798 (2011).
265. Epping, M. T. *et al.* The Human Tumor Antigen PRAME Is a Dominant Repressor of Retinoic Acid Receptor Signaling. *Cell* **122,** 835–847 (2005).
266. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).
267. Sun, Y., Tran, B. N., Worley, L. A., Delston, R. B. & Harbour, J. W. Functional Analysis of the p53 Pathway in Response to Ionizing Radiation in Uveal Melanoma. *Investig. Opthalmology Vis. Sci.* **46,** 1561 (2005).
268. Jr., M. A. B. & Harbour, J. W. Deregulation of the Rb and p53 Pathways in Uveal Melanoma. *Am. J. Pathol.* **157,** 1795 (2000).
269. Brantley, M. A. & Harbour, J. W. Inactivation of retinoblastoma protein in uveal melanoma by phosphorylation of sites in the COOH-terminal region. *Cancer Res.* **60,** 4320–3 (2000).
270. An, J. *et al.* A comparative transcriptomic analysis of uveal melanoma and normal uveal melanocyte. *PLoS One* **6,** e16516 (2011).
271. van der Velden, P. A. *et al.* Promoter hypermethylation: a common cause of reduced p16(INK4a) expression in uveal melanoma. *Cancer Res.* **61,** 5303–6 (2001).
272. Coupland, S. E. *et al.* Expression patterns of cyclin D1 and related proteins regulating G1-S phase transition in uveal melanoma and retinoblastoma. *Br. J. Ophthalmol.* **82,** 961–70 (1998).
273. Moulin, A. P., Clement, G., Bosman, F. T., Zografos, L. & Benhattar, J. Methylation of CpG island promoters in uveal melanoma. *Br. J. Ophthalmol.* **92,** 281–285 (2008).
274. Zeschnigk, M. *et al.* Methylation Analysis of Several Tumour Suppressor Genes Shows a Low Frequency of Methylation of CDKN2A and RARB in Uveal Melanomas. *Comp. Funct. Genomics* **4,** 329–336 (2003).
275. Coupland, S. E. *et al.* The prognostic value of cyclin D1, p53, and MDM2 protein expression in uveal melanoma. *J. Pathol.* **191,** 120–126 (2000).
276. Saraiva, V. S., Caissie, A. L., Segal, L., Edelstein, C. & Burnier, M. N. Immunohistochemical expression of phospho-Akt in uveal melanoma. *Melanoma Res.* **15,** 245–50 (2005).
277. Pópulo, H., Soares, P., Rocha, A. S., Silva, P. & Lopes, J. M. Evaluation of the mTOR

pathway in ocular (uvea and conjunctiva) melanoma. *Melanoma Res.* **20,** 107–117 (2010).
278. Wu, X. *et al.* c-Met, epidermal growth factor receptor, and insulin-like growth factor-1 receptor are important for growth in uveal melanoma and independently contribute to migration and metastatic potential. *Melanoma Res.* **22,** 123–132 (2012).
279. All-Ericsson, C. *et al.* Insulin-like growth factor-1 receptor in uveal melanoma: a predictor for metastatic disease and a potential therapeutic target. *Invest. Ophthalmol. Vis. Sci.* **43,** 1–8 (2002).
280. Abdel-Rahman, M. H. *et al.* High Frequency of Submicroscopic Hemizygous Deletion Is a Major Mechanism of Loss of Expression of *PTEN* in Uveal Melanoma. *J. Clin. Oncol.* **24,** 288–295 (2006).
281. Naus, N. C. *et al.* Mutation analysis of the PTEN gene in uveal melanoma cell lines. *Int. J. cancer* **87,** 151–3 (2000).
282. Pache, M. *et al.* Sequence analysis and high-throughput immunhistochemical profiling of KIT (CD 117) expression in uveal melanoma using tissue microarrays. *Virchows Arch.* **443,** 741–744 (2003).
283. Weber, A. *et al.* Absence of mutations of the BRAF gene and constitutive activation of extracellular-regulated kinase in malignant melanomas of the uvea. *Lab. Invest.* **83,** 1771–6 (2003).
284. Chen, X. *et al.* Combined PKC and MEK inhibition in uveal melanoma with GNAQ and GNA11 mutations. *Oncogene* **33,** 4724–34 (2014).
285. Abdel-Rahman, M. H., Craig, E. L., Davidorf, F. H. & Eng, C. Expression of vascular endothelial growth factor in uveal melanoma is independent of 6p21-region copy number. *Clin. Cancer Res.* **11,** 73–8 (2005).
286. Cools-Lartigue, J., Marshall, J. C., Caissie, A. L., Saraiva, V. S. & Burnier, M. N. Secretion of hepatocyte growth factor and vascular endothelial growth factor during uveal melanoma-monocyte in vitro interactions. *Melanoma Res.* **15,** 141–5 (2005).
287. Sheidow, T. G., Hooper, P. L., Crukley, C., Young, J. & Heathcote, J. G. Expression of vascular endothelial growth factor in uveal melanoma and its correlation with metastasis. *Br. J. Ophthalmol.* **84,** 750–6 (2000).
288. Zuidervaart, W. *et al.* Expression of Wnt5a and its downstream effector β-catenin in uveal melanoma. *Melanoma Res.* **17,** 380–386 (2007).
289. Anastassiou, G., Tschentscher, F., Zeschnigk, M. & Bornfeld, N. Cadherin Expression in Uveal Melanoma. *Exp. Eye Res.* **74,** 423–425 (2002).
290. Coupland, S. E., Lake, S. L., Zeschnigk, M. & Damato, B. E. Molecular pathology of uveal melanoma. *Eye* **27,** 230–242 (2013).
291. Yang, W., Chen, P. W., Li, H., Alizadeh, H. & Niederkorn, J. Y. PD-L1: PD-1 Interaction Contributes to the Functional Suppression of T-Cell Responses to Human Uveal Melanoma Cells In Vitro. *Investig. Opthalmology Vis. Sci.* **49,** 2518 (2008).
292. Whelchel, J. C., Farah, S. E., McLean, I. W. & Burnier, M. N. Immunohistochemistry of infiltrating lymphocytes in uveal malignant melanoma. *Invest. Ophthalmol. Vis. Sci.* **34,** 2603–6 (1993).
293. Mäkitie, T., Summanen, P., Tarkkanen, A. & Kivelä, T. Tumor-infiltrating macrophages (CD68(+) cells) and prognosis in malignant uveal melanoma. *Invest. Ophthalmol. Vis. Sci.* **42,** 1414–21 (2001).
294. Toivonen, P., Mäkitie, T., Kujala, E. & Kivelä, T. Microcirculation and tumor-infiltrating macrophages in choroidal and ciliary body melanoma and corresponding metastases.

Invest. Ophthalmol. Vis. Sci. **45,** 1–6 (2004).
295. Jager, M. J., Ly, L. V., El Filali, M. & Madigan, M. C. Macrophages in uveal melanoma and in experimental ocular tumor models: Friends or foes? *Prog. Retin. Eye Res.* **30,** 129–146 (2011).
296. Singh, A. D., Kalyani, P. & Topham, A. Estimating the Risk of Malignant Transformation of a Choroidal Nevus. *Ophthalmology* **112,** 1784–1789 (2005).
297. Peiffer, D. A. *et al.* High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* **16,** 1136–48 (2006).
298. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16910–5 (2010).
299. Winkler, G. *et al.* Smoothers for Discontinuous Signals. *J. Nonparametr. Stat.* **14,** 203–222 (2002).
300. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29,** 24–6 (2011).
301. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41 (2011).
302. Newman, S. Interactive analysis of large cancer copy number studies with Copy Number Explorer. *Bioinformatics* **31,** 2874–6 (2015).
303. Gentleman, R. C. *et al.* voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **5,** R80 (2004).
304. Storey, J. D. The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Stat.* **31,** 2013–2035 (2003).
305. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–504 (2003).
306. S, A. FastQC: a quality control tool for high throughput sequence data. (2010). at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
307. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
308. Hwang, S., Kim, E., Lee, I. & Marcotte, E. M. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci. Rep.* **5,** 17875 (2015).
309. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
310. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013). at <http://arxiv.org/abs/1303.3997>
311. Picard: A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. at <http://broadinstitute.github.io/picard>
312. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43,** 11.10.1-33 (2013).
313. Liu, X., Han, S., Wang, Z., Gelernter, J. & Yang, B.-Z. Variant Callers for Next-Generation Sequencing Data: A Comparison Study. *PLoS One* **8,** e75619 (2013).
314. D, B. Variant calling comparison CASAVA1.8 and GATK. in *Nature Precedings* (2011).
315. Yi, M. *et al.* Performance comparison of SNP detection tools with illumina exome sequencing data—an assessment using both family pedigree information and sample-matched SNP array data. *Nucleic Acids Res.* **42,** e101–e101 (2014).
316. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and

heterogeneous cancer samples. *Nat. Biotechnol.* **31,** 213–219 (2013).
317. Engström, P. G. *et al.* Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* **10,** 1185–1191 (2013).
318. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26,** 2069–2070 (2010).
319. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164–e164 (2010).
320. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* **6,** 80–92 (2012).
321. McCarthy, D. J. *et al.* Choice of transcripts and software has a large effect on variant annotation. *Genome Med.* **6,** 26 (2014).
322. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
323. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–315 (2014).
324. Whiffin, N. *et al.* Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet. Med.* (2017). doi:10.1038/gim.2017.26
325. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
326. Brunet, J.-P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 4164–9 (2004).
327. Mayakonda, A. & Koeffler, H. P. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. *bioRxiv* (2016). at <http://biorxiv.org/content/early/2016/05/11/052662>
328. Skidmore, Z. L. *et al.* GenVisR: Genomic Visualizations in R. *Bioinformatics* **32,** 3012–3014 (2016).
329. Kumar, S., Vo, A. D., Qin, F., Li, H. & Wold, B. Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Sci. Rep.* **6,** 21597 (2016).
330. Davidson, N. M., Majewski, I. J. & Oshlack, A. JAFFA: High sensitivity transcriptome-focused fusion gene detection. *Genome Med.* **7,** 43 (2015).
331. Carrara, M. *et al.* State-of-the-art fusion-finder algorithms sensitivity and specificity. *Biomed Res. Int.* **2013,** 340620 (2013).
332. Babiceanu, M. *et al.* Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res.* **44,** 2859–72 (2016).
333. Shugay, M., Ortiz de Mendíbil, I., Vizmanos, J. L. & Novo, F. J. Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. *Bioinformatics* **29,** 2539–2546 (2013).
334. Weir, B. A. *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450,** 893–898 (2007).
335. Zender, L. *et al.* Identification and Validation of Oncogenes in Liver Cancer Using an Integrative Oncogenomic Approach. *Cell* **125,** 1253–1267 (2006).
336. McLendon, R. *et al.* Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
337. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute

lymphoblastic leukaemia. *Nature* **446,** 758–764 (2007).
338. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463,** 899–905 (2010).
339. Tamborero, D. *et al.* Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **3,** 1113–1120 (2013).
340. Leiserson, M. D. M. *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* **47,** 106–114 (2014).
341. Wang, X. *et al.* Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. *Biochem. J.* **383,** 319–25 (2004).
342. Hodges, C., Kirkland, J. G. & Crabtree, G. R. The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. *Cold Spring Harb. Perspect. Med.* **6,** a026930 (2016).
343. Cai, Y. *et al.* Loss of Chromosome 8p Governs Tumor Progression and Drug Response by Altering Lipid Metabolism. *Cancer Cell* **29,** 751–766 (2016).
344. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–50 (2005).
345. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
346. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39,** D691–D697 (2011).
347. Windheim, M., Lang, C., Peggie, M., Plater, L. A. & Cohen, P. Molecular mechanisms involved in the regulation of cytokine production by muramyl dipeptide. *Biochem. J.* **404,** 179–190 (2007).
348. McCarthy, J. V, Ni, J. & Dixit, V. M. RIP2 is a novel NF-kappaB-activating and cell death-inducing kinase. *J. Biol. Chem.* **273,** 16968–75 (1998).
349. Karin, M., Cao, Y., Greten, F. R. & Li, Z.-W. NF-κB IN CANCER: FROM INNOCENT BYSTANDER TO MAJOR CULPRIT. *Nat. Rev. Cancer* **2,** 301–310 (2002).
350. Dror, R. *et al.* Characterizing the Involvement of the Nuclear Factor-kappa B (NFκB) Transcription Factor in Uveal Melanoma. *Investig. Opthalmology Vis. Sci.* **51,** 1811 (2010).
351. Bignone, P. A. *et al.* RPS6KA2, a putative tumour suppressor gene at 6q27 in sporadic epithelial ovarian cancer. *Oncogene* **26,** 683–700 (2007).
352. Pardal, R., Clarke, M. F. & Morrison, S. J. Applying the principles of stem-cell biology to cancer. *Nat. Rev. Cancer* **3,** 895–902 (2003).
353. Zuidervaart, W. *et al.* Expression of Wnt5a and its downstream effector β-catenin in uveal melanoma. *Melanoma Res.* **17,** 380–386 (2007).
354. Blencowe, B. J., Issner, R., Nickerson, J. A. & Sharp, P. A. A coactivator of pre-mRNA splicing. *Genes Dev.* **12,** 996–1009 (1998).
355. Eldridge, A. G., Li, Y., Sharp, P. A. & Blencowe, B. J. The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6125–30 (1999).
356. Blencowe, B. J. *et al.* The SRm160/300 splicing coactivator subunits. *RNA* **6,** 111–20 (2000).
357. McCracken, S., Lambermon, M. & Blencowe, B. J. SRm160 splicing coactivator promotes transcript 3'-end cleavage. *Mol. Cell. Biol.* **22,** 148–60 (2002).
358. McCracken, S., Longman, D., Johnstone, I. L., Caceres, J. F. & Blencowe, B. J. An

Evolutionarily Conserved Role for SRm160 in 3'-End Processing That Functions Independently of Exon Junction Complex Formation. *J. Biol. Chem.* **278,** 44153–44160 (2003).
359. Szymczyna, B. R. *et al.* Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. *Genes Dev.* **17,** 461–475 (2003).
360. Fan, Y.-J. *et al.* Multifunctional RNA Processing Protein SRm160 Induces Apoptosis and Regulates Eye and Genital Development in Drosophila. *Genetics* **197,** 1251–1265 (2014).
361. Niederkorn, J. Y. & Wang, S. Immunology of intraocular tumors. *Ocul. Immunol. Inflamm.* **13,** 105–10 (2005).
362. Wei, M. *et al.* A Novel Plant Homeodomain Finger 10-Mediated Antiapoptotic Mechanism Involving Repression of Caspase-3 in Gastric Cancer Cells. *Mol. Cancer Ther.* **9,** 1764–1774 (2010).
363. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39,** D945-50 (2011).
364. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29,** 24–26 (2011).
365. Harbour, J. W. *et al.* Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330,** 1410–1413 (2010).
366. Lawo, S. *et al.* HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity. *Curr. Biol.* **19,** 816–26 (2009).
367. Zhu, H., Coppinger, J. A., Jang, C.-Y., Yates, J. R. & Fang, G. FAM29A promotes microtubule amplification via recruitment of the NEDD1-gamma-tubulin complex to the mitotic spindle. *J. Cell Biol.* **183,** 835–48 (2008).
368. Lyon, A. M. & Tesmer, J. J. G. Structural insights into phospholipase C-β function. *Mol. Pharmacol.* **84,** 488–500 (2013).
369. Shain, A. H. *et al.* Convergent structural alterations define SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler as a central tumor suppressive complex in pancreatic cancer. *Proc. Natl. Acad. Sci.* **109,** E252–E259 (2012).
370. Reisman, D., Glaros, S. & Thompson, E. A. The SWI/SNF complex and cancer. *Oncogene* **28,** 1653–1668 (2009).
371. Chase, A. *et al.* TFG, a target of chromosome translocations in lymphoma and soft tissue tumors, fuses to GPR128 in healthy individuals. *Haematologica* **95,** 20–6 (2010).
372. Koolen, D. A. *et al.* The Koolen-de Vries syndrome: a phenotypic comparison of patients with a 17q21.31 microdeletion versus a KANSL1 sequence variant. *Eur. J. Hum. Genet.* **24,** 652–659 (2016).
373. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37,** 129–137 (2005).
374. Jones, D. T. W. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488,** 100–105 (2012).
375. Jones, S. *et al.* Frequent Mutations of Chromatin Remodeling Gene ARID1A in Ovarian Clear Cell Carcinoma. *Science (80-.).* **330,** 228–231 (2010).
376. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486,** 400–4 (2012).
377. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44,** 760–764 (2012).

378. Sun, D. *et al.* SASH1 inhibits proliferation and invasion of thyroid cancer cells through PI3K/Akt signaling pathway. *Int. J. Clin. Exp. Pathol.* **8,** 12276–83 (2015).
379. Zeller, C. *et al.* SASH1: a candidate tumor suppressor gene on chromosome 6q24.3 is downregulated in breast cancer. *Oncogene* **22,** 2972–2983 (2003).
380. Delattre, O. *et al.* Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359,** 162–165 (1992).
381. Williams, S. V., Hurst, C. D. & Knowles, M. A. Oncogenic FGFR3 gene fusions in bladder cancer. *Hum. Mol. Genet.* **22,** 795–803 (2013).
382. Soda, M. *et al.* Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature* **448,** 561–566 (2007).
383. Kroll, T. G. *et al.* PAX8-PPARgamma1 fusion oncogene in human thyroid carcinoma [corrected]. *Science* **289,** 1357–60 (2000).
384. Zhou, J. X. *et al.* Identification of KANSARL as the first cancer predisposition fusion gene specific to the population of European ancestry origin. *Oncotarget* **5,** (2017).
385. Shago, M., Abla, O., Hitzler, J., Weitzman, S. & Abdelhaleem, M. Frequency and outcome of pediatric acute lymphoblastic leukemia with ZNF384 gene rearrangements including a novel translocation resulting in an ARID1B/ZNF384 gene fusion. *Pediatr. Blood Cancer* **63,** 1915–1921 (2016).
386. Sausen, M. *et al.* Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nat. Genet.* **45,** 12–17 (2012).
387. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45,** 1134–1140 (2013).
388. Sisley, K. *et al.* Multiplex fluorescence in situ hybridization identifies novel rearrangements of chromosomes 6, 15, and 18 in primary uveal melanoma. *Exp. Eye Res.* **83,** 554–559 (2006).
389. Casta??n, M. J., Walko, G., Winter, L. & Wiche, G. Plectin?intermediate filament partnership in skin, skeletal muscle, and peripheral nerve. *Histochem. Cell Biol.* **140,** 33–53 (2013).
390. Wu, J. N. & Roberts, C. W. M. ARID1A mutations in cancer: another epigenetic tumor suppressor? *Cancer Discov.* **3,** 35–43 (2013).
391. Roberts, C. W. M. & Orkin, S. H. The SWI/SNF complex - chromatin and cancer. *Nat. Rev. Cancer* **4,** 133–142 (2004).

Appendices

Appendix 1 – Listed code used to perform analysis

Code 2.1: ASCAT code for generating allele specific copy number calls for paired and unpaired LRR and BAF data. Code adapter from Peter Van Loo et al [298] (https://www.crick.ac.uk/petervanloo/software/ASCAT)

```
#ASCAT steps for paired and unpaired data

#Load LRR and BAF data

#unpaired data
ascat.bc = ascat.loadData("UM-LRR.txt","UM-BAF.txt")

#paired data
ascat.bc = ascat.loadData("UM-Tumor-LRR.txt","UM-Tumor-BAF.txt","UM-Normal.txt", "UM-Normal-BAF.txt")

#GC correction
ascat.bc = ascat.GCcorrect(ascat.bc, "GC_IlluminaOmniexpress.txt")

#infer the necessary germline genotypes from the tumour data for unpaired data
ascat.gg = ascat.predictGermlineGenotypes(ascat.bc, "IlluminaCytoSNP")

#ascat function to call allele specific copy numbers
ascat.bc = ascat.aspcf(ascat.bc,ascat.gg=ascat.gg)
ascat.plotSegmentedData(ascat.bc)
ascat.output = ascat.runAscat(ascat.bc, pdfPlot = T)
```

Code 2.2: Code to run the BWA alignment and the SAM file clean up step.

```
#BWA alignment

# Script Usage:
if [ "$#" != "3" ]; then
   echo -e "Usage of the script: bash nga_step1.sh <read-file1> <read-file2> <output-file>"
   exit 1
fi

File1=${1}
File2=${2}
File3=${3}

REF=/data/volume1/bin/reference/hg19m.fa
PICARD=/usr/local/picard.jar

bwa aln -t 4 $REF $File1 > ${File1}.sai
bwa aln -t 4 $REF $File2 > ${File2}.sai
bwa sampe -a 400 $REF ${File1}.sai ${File2}.sai $File1 $File2 > ${File3}.sam

java -jar $PICARD CleanSam INPUT=${File3}.sam OUTPUT=${File3}_c.sam VALIDATION_STRINGENCY=LENIENT
```

Code 2.3: Post-alignment processing steps on the aligned SAM format file

```
#Post-alignment processing

# Script Usage:
if [ "$#" != ”6" ]; then
   echo -e "Usage of the script: bash nga_step2.sh <aligned BAM file> <RGID> <RGLB> <RGPL> <RGPU> <RGSM> "
   exit 1
fi

File1=${1}
RGID = ${2}
RGLB = ${3}
RGPL = ${4}
RGPU = ${5}
RGSM = ${6}

REF=/data/volume1/bin/reference/hg19m.fa
DB132=/data/volume1/bin/reference/dbsnp_132_reordered.vcf
PICARD=/usr/local/picard.jar
GATK=/home/volume1/bin/GenomeAnalysisTK.jar

java -jar $PICARD SamFormatConverter INPUT=${File1} OUTPUT=${File1}.bam

java -jar $PICARD AddOrReplaceReadGroups INPUT=${File1}.bam OUTPUT=${File1}_nfo.bam RGID=$RGID RGLB=$RGLB RGPL=$RGPL
RGPU=$RGPU RGSM=$RGSM

java -jar $PICARD SortSam INPUT=${File1}_nfo.bam OUTPUT=${File1}_nfo_srt.bam SORT_ORDER=coordinate

java -jar $PICARD FixMateInformation INPUT=${File1}_nfo_srt.bam OUTPUT=${File1}_nfo_srt_fix.bam SORT_ORDER=coordinate

java -jar $PICARD MarkDuplicates INPUT=${File1}_nfo_srt_fix.bam OUTPUT=${File1}_nfo_srt_fix_nodup.bam
METRICS_FILE=${File1}_duplicate.txt REMOVE_DUPLICATES=true

java -jar $PICARD BuildBamIndex INPUT=${File1} OUTPUT=${File1}.bai

java -jar $GATK -T BaseRecalibrator -I ${File1}_nfo_srt_fix_nodup.bam -R $REF -knownSites  $DB132 -o ${File1}_recalibration_report.grp

java -jar $GATK -T PrintReads -R $REF -I ${File1}_nfo_srt_fix_nodup.bam  –BQSR ${File1}_recalibration_report.grp -o
${File1}_realn_recal.bam
```

Code 2.4: Variant calling with MuTect2: (1) creation of panel of normal; (2) somatic variant calling on all tumor samples

```
#MuTect2 variant calling – (1) Creation of panel of normal

REF=/data/volume1/bin/reference/hg19m.fa
DB132=/data/volume1/bin/reference/dbsnp_132_reordered.vcf
GATK=/home/volume1/bin/GenomeAnalysisTK.jar
interval=/data/volume1/bin/SeqCap_ver2_primary_targets_modified.intervals
filename1="/data/volume4/ocular_melanoma/normal-samples.txt"

cat $filename1 | while read -r line
do
   java -jar $GATK -T MuTect2 -R $REF -nct 30 --artifact_detection_mode -L $interval --dbsnp $DB132 -I:tumor
${line}.bam -o ${line}.mutect2.output.vcf
done

ls > temp
grep –v “sample” temp > input_vcf.list
rm temp

java -jar $ GATK -T CombineVariants -R $REF -V input_vcf.list -minN 2 --setKey "null" --filteredAreUncalled --
filteredrecordsmergetype KEEP_IF_ANY_UNFILTERED -L $interval -o MuTect2_PON.vcf
```

```
#MuTect2 variant calling – (2) Somatic variant calling

REF=/data/volume1/bin/reference/hg19m.fa
DB132=/data/volume1/bin/reference/dbsnp_132_reordered.vcf
GATK=/home/volume1/bin/GenomeAnalysisTK.jar
interval=/data/volume1/bin/SeqCap_ver2_primary_targets_modified.intervals
cosmic=/data/volume1/bin/hg19_cosmic.vcf
pon=/data/volume4/ocular_melanoma/MuTect2_PON.vcf

filename="/data/volume4/ocular_melanoma/tumor-samples.txt"

cat $filename | while read -r line
do
java -jar $GATK -T MuTect2 -R $REF -I:tumor ${line}.bam --normal_panel $pon --dbsnp $DB132 --cosmic $cosmic -L
$interval -o ${line}.mutect2.tumor.output.vcf
done
```

Code 2.5: Variant calling with HaplotypeCaller

```
#HaplotypeCaller variant calling

REF=/data/volume1/bin/reference/hg19m.fa
DB132=/data/volume1/bin/reference/dbsnp_132_reordered.vcf
GATK=/home/volume1/bin/GenomeAnalysisTK.jar
interval=/data/volume1/bin/SeqCap_ver2_primary_targets_modified.intervals
filename="/data/volume4/ocular_melanoma/tumor-and-normal-samples.txt"

cat $filename | while read -r line
do
java -jar $GATK -I ${File1}.bam -R $REF -T HaplotypeCaller -L $interval --dbsnp $DB132 --emitRefConfidence GVCF --
variant_index_type LINEAR --variant_index_parameter 128000 -o ${File1}.raw.snps.indels.g.vcf
done

ls > temp
grep –v “sample” temp > input_vcf.list
rm temp

# After generating all the gVCFs
java -jar $GATK -R $REF -T GenotypeGVCFs -V input_gvcf.list --dbsnp $DB132 -o HC_output.vcf
```

Code 2.6: RNA-seq variant calling steps; variant calling performed with HaplotypeCaller (ref code 2.5)

```
#RNA-seq variant calling

GATK=/home/volume1/bin/GenomeAnalysisTK.jar
STAR1=/home/volume1/bin/STAR-STAR_2.4.1d/source/STAR
REF=/data/volume1/bin/reference/hg19m.fa
GENOMEDIR=/data/volume1/UM-RNA-Seq-expression/STAR-reference-index/GRCh37-reference/
REFERENCE=/data/volume1/bin/STAR-reference-index/GRCh37-reference/Homo_sapiens.GRCh37.75.dna.fa
GENECODE_ANNO=/data/volume1/bin/STAR-reference-index/GRCh37-reference/Homo_sapiens.GRCh37.75.gtf
SJDBFILE=/data/volume1/bin/STAR-reference-index/ucsc_intron_coordinate_modified.txt

# Script Usage:
if [ "$#" != "3" ]; then
   echo -e "Usage of the script: bash nga_step5.sh <read-file1> <read-file2> <output-file>"
   exit 1
fi

File1=${1}
File2=${2}
File3=${3}

$STAR1 --runThreadN 4 --genomeDir $GENOMEDIR --readFilesIn $File1 $File2 --sjdbGTFfile $GENECODE_ANNO -sjdbOverhang 75 --
outSAMtype BAM SortedByCoordinate --outFileNamePrefix ${File3}

$STAR1 --runThreadN 4 --runMode genomeGenerate --genomeDir $GENOMEDIR --genomeFastaFiles $REFERENCE --sjdbFileChrStartEnd
$SJDBFILE ${File3}SJ.out.tab --sjdbOverhang 75 --outFileNamePrefix ${File3}_2_pass_step1

$STAR1 --runThreadN 4 --genomeDir $GENOMEDIR --readFilesIn $File1 $File2 --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix ${File3}_2_pass_step2

java -jar $PICARD AddOrReplaceReadGroups INPUT=${File3}_2_pass_step2.bam OUTPUT= ${File3}_2_pass_step2_nfo.bam RGID=$RGID
RGLB=$RGLB RGPL=$RGPL RGPU=$RGPU RGSM=$RGSM

java -jar $PICARD SortSam INPUT=${File1}_2_pass_step2_nfo.bam OUTPUT=${File1}_2_pass_step2_nfo_srt.bam
SORT_ORDER=coordinate

java -jar $PICARD MarkDuplicates INPUT=${File1}_2_pass_step2_nfo_srt.bam OUTPUT=${File1} _2_pass_step2_nfo_srt_fix_nodup.bam
METRICS_FILE=${File1}_duplicate.txt REMOVE_DUPLICATES=true

java -jar $GATK -T SplitNCigarReads -R $REF -I ${File1} _2_pass_step2_nfo_srt_fix_nodup.bam -o ${File1}_split.bam -rf
ReassignOneMappingQuality -RMQF 255 -RMQT 60 -U ALLOW_N_CIGAR_READS

java -jar $GATK -T BaseRecalibrator -I ${File1}_split.bam -R $REF -knownSites $DB132 -o ${File1}_split_recalibration_report.grp

java -jar $GATK -T PrintReads -R $REF -I ${File1}_split.bam –BQSR ${File1}_split_recalibration_report.grp -o ${File1}_split_recal.bam
```

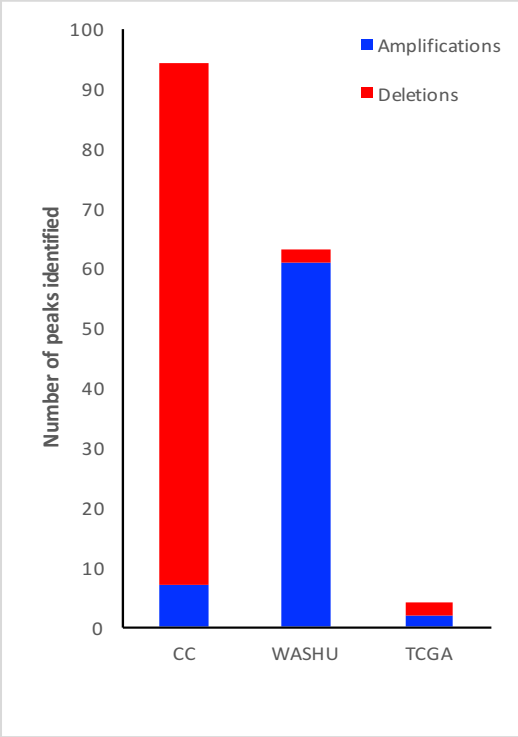
Code 2.7: Variant annotation

```
#Variant annotation

DB=/data/volume1/bin/resources/oncotator_v1_ds_Jan262014
oncotator -v --input_format=VCF --output_format=TCGAMAF --db-dir $DB input.vcf output.maf.annotated hg19

#ANNOVAR
perl  convert2annovar.pl -format vcf4old HC_output.vcf -outfile HC_output.vcf.avinput
perl  annotate_variation.pl -geneanno -buildver hg19 HC_output.vcf.avinput humandb/
perl  annotate_variation.pl -regionanno -dbtype cytoBand -buildver hg19 HC_output.vcf.avinput humandb/
perl  annotate_variation.pl -filter -dbtype exac03 -buildver hg19 HC_output.vcf.avinput humandb/
```

Appendix 2 - GISTIC analysis re-run with q-value threshold of 0.05

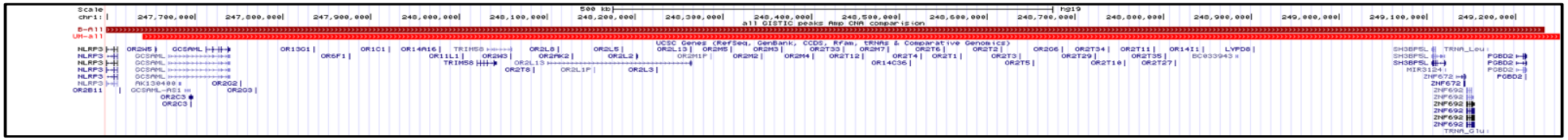


Comparison of peaks between three independent analysis with q-value of 0.05. The number of amplification and deletion peaks were higher for CC and WASHU cohorts compared to TCGA. This difference could be due to platform specific effects since data for each of the cohorts were generated using different array platforms.

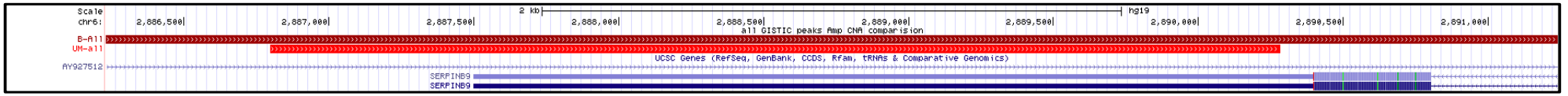
Appendix 3 - GISTIC defined recurrent copy number amplification and deletions shared between UM and other solid cancers shown in the image from UCSC genome browser.

Figure 1A-E showing partial overlapping amplification peaks. The amplification peaks and frequency of alteration in our cohort are as follows: 1q44 (16%), 6q25.2 peak 1 (52%), 6q25.2 peak2 (50%) and 8q24.3 peak 1 & 2 (74%). UM amplification peaks indicated in Red and peaks from other pan cancers indicated in brown. Figure 2A-J showing partial overlapping deletion peaks. The deletion peaks and frequency of alteration in our cohort are as follows: 6q16.2 peak 1 & 2 (37% & 35%), 1p13.1 (20%), 1p35.3 (34%), 4q35.1 (23%), 7q35 (19%), 8p11.2 peak 1, 2 & 3 (27%, 28% & 30%), 8p21.2 (35%), 11q24.3 peaks 1, 2 & 3 (16%, 11% & 10%) and 16q24.3 (24%). Deletion peaks are indicated in blue and peaks from other solid cancers indicated in purple.

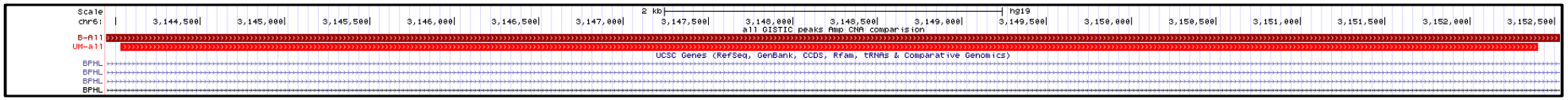
1A) Amplification Peak: Chromosome 1q44



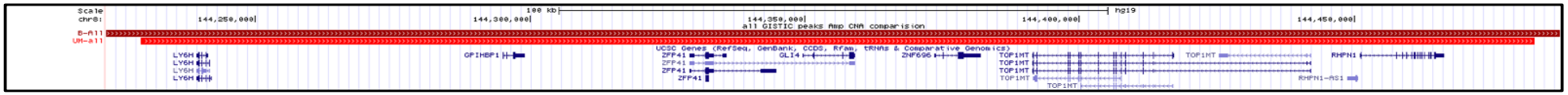
1B) Amplification Peak: 6q25.2-peak1



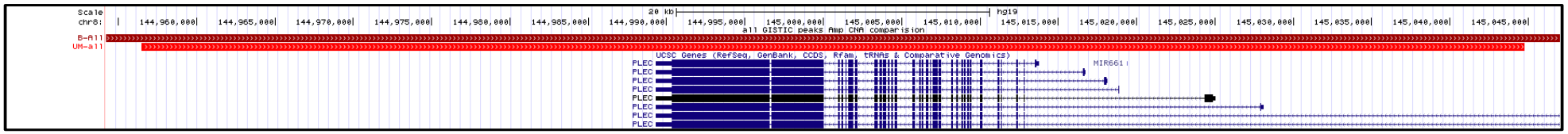
1C) Amplification Peak: 6q25.2 –peak2



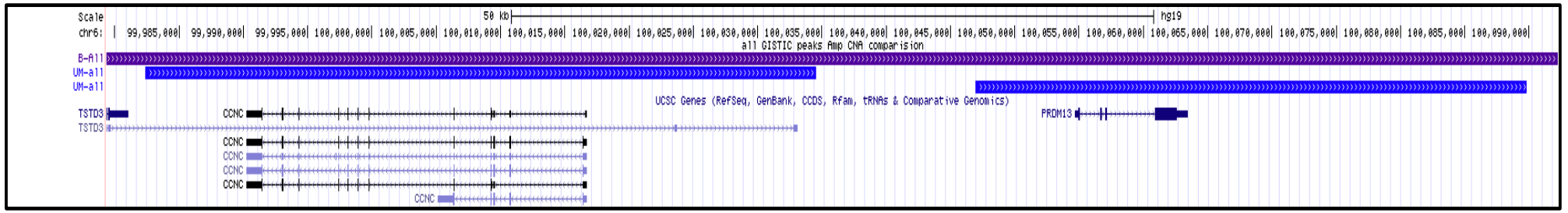
1D) Amplification Peak: 8q24.3 - peak1



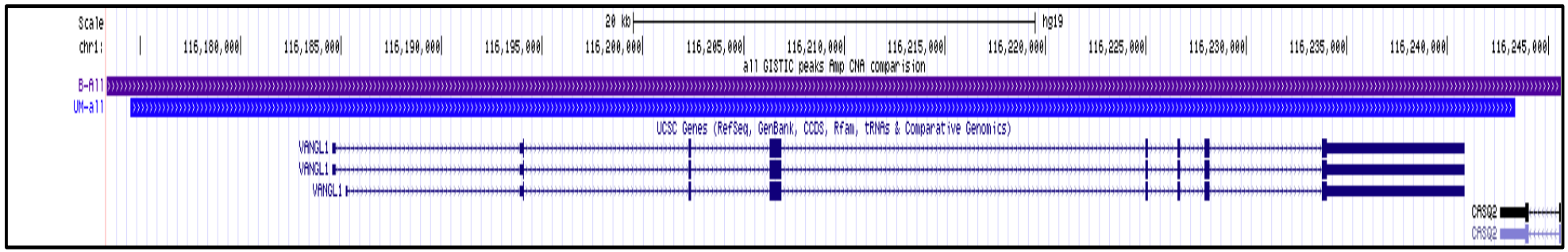
1E) Amplification Peak: 8q24.3 – peak2



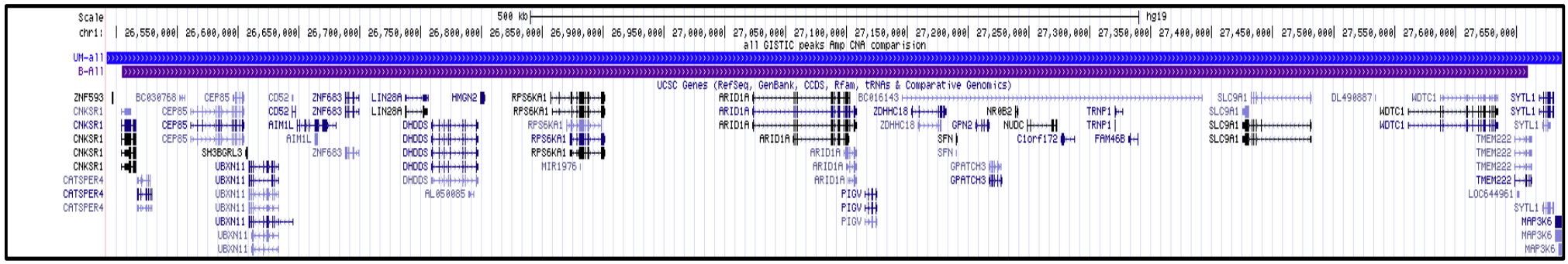
2A) Deletion Peak: Chromosome 6q16.2 – peak1 & peak2



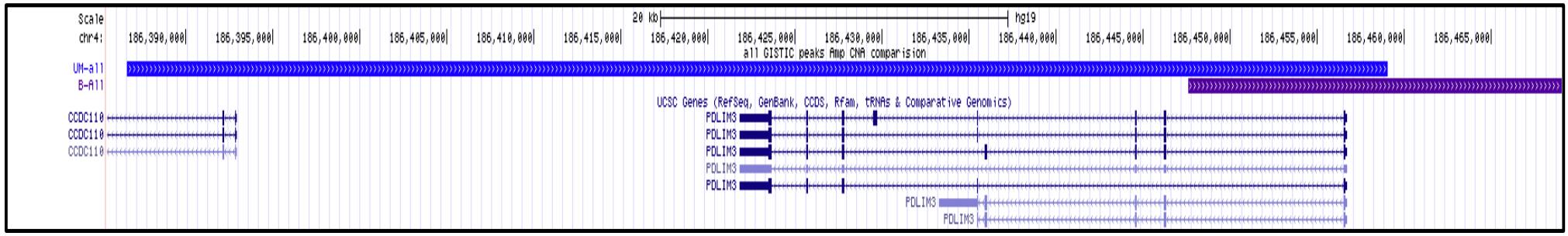
2B) Deletion Peak: Chromosome 1p13.1



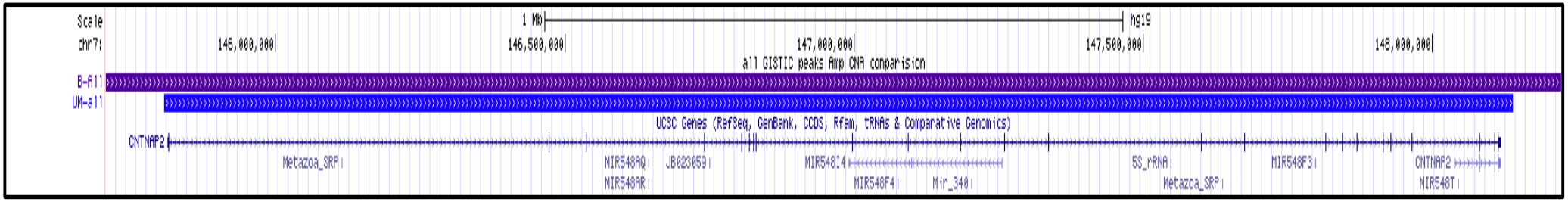
2C) Deletion Peak: Chromosome 1p35.3



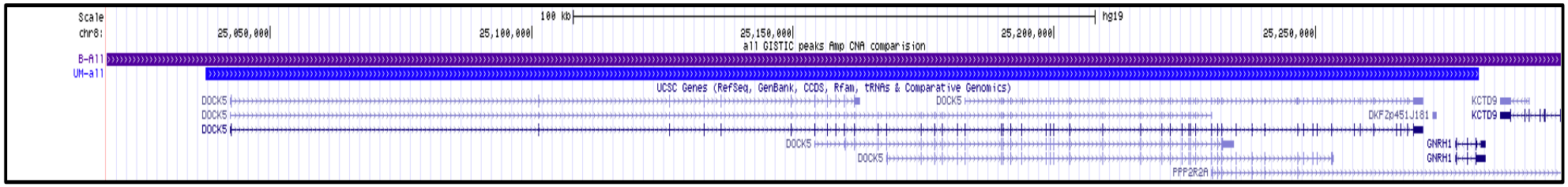
2D) Deletion Peak: Chromosome 4q35.1



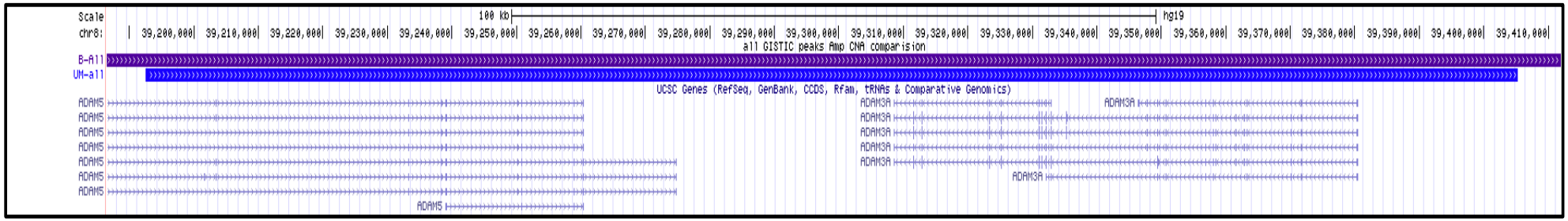
2F) Deletion Peak: Chromosome 7q35



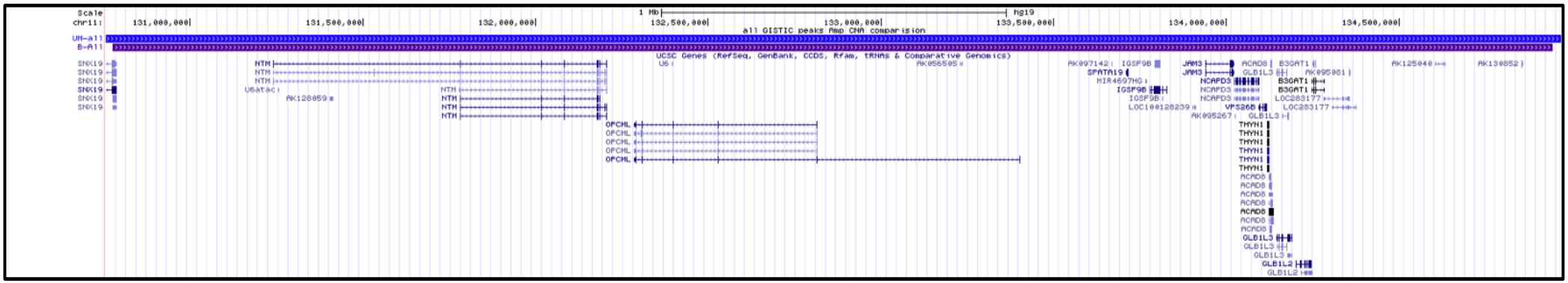
2G) Deletion Peak: Chromosome 8p11.2



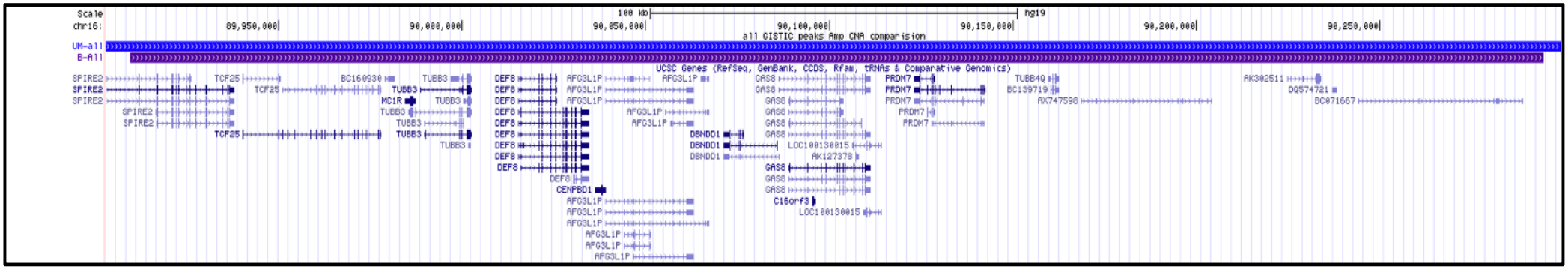
2H) Deletion Peak: Chromosome 8p21.2



2I) Deletion Peak: Chromosome 11q24.3



2J) Deletion Peak: Chromosome 16q24.3



Appendix 4 – Variant table for all significant genes from MutSigCV analysis

Gene	Sample with mutation	Cohort	Chromosome	Start	End	Variant_Classification	Variant_Type	Reference allele	Mutant allele	Mutant allele frequency	Protein change	CADD score	Keep/Discard variants with reason for discarding variant
AOC3	A9EL	TCGA	17	41003601	41003603	In_Frame_Del	DEL	CCA	-	0.030	p.P81del		Discard, recurrent Indel in single cohort - likely artefact
AOC3	A87Y	TCGA	17	41003602	41003603	Frame_Shift_Del	DEL	CA	-	0.047	p.P81fs		Discard, recurrent Indel in single cohort - likely artefact
AOC3	A880	TCGA	17	41003602	41003603	Frame_Shift_Del	DEL	CA	-	0.090	p.P81fs		Discard, recurrent Indel in single cohort - likely artefact
AOC3	A9E5	TCGA	17	41003602	41003603	Frame_Shift_Del	DEL	CA	-	0.116	p.P81fs		Discard, recurrent Indel in single cohort - likely artefact
AOC3	A9EO	TCGA	17	41003602	41003603	Frame_Shift_Del	DEL	CA	-	0.064	p.P81fs		Discard, recurrent Indel in single cohort - likely artefact
AOC3	A9EX	TCGA	17	41003602	41003603	Frame_Shift_Del	DEL	CA	-	0.064	p.P81fs		Discard, recurrent Indel in single cohort - likely artefact
AOC3	AA8Q	TCGA	17	41003602	41003603	Frame_Shift_Del	DEL	CA	-	0.070	p.P81fs		Discard, recurrent Indel in single cohort - likely artefact
AOC3	AA8S	TCGA	17	41003602	41003603	Frame_Shift_Del	DEL	CA	-	0.039	p.P81fs		Discard, recurrent Indel in single cohort - likely artefact
AOC3	A9EW	TCGA	17	41003704	41003704	Missense_Mutation	SNP	G	C	0.500	p.R115T	8.491	Discard, CADD < 10
AOC3	MM127	WASH-U	17	41008337	41008337	Missense_Mutation	SNP	G	A	0.480	p.A688T	19.14	Keep
BAP1	MM103	WASH-U	3	52436306	52436306	Nonstop_Mutation	SNP	A	T	0.867	p.*730R	17.92	Keep , Known driver
BAP1	MM098	WASH-U	3	52436617	52436617	Splice_Site	SNP	C	T	0.492	0.000	16.16	Keep , Known driver
BAP1	A9EO	TCGA	3	52436620	52436620	Missense_Mutation	SNP	T	A	0.977	p.E685V	20.9	Keep , Known driver
BAP1	A9EI	TCGA	3	52436650	52436659	Frame_Shift_Del	DEL	ATGAACTCAT	-	0.767	p.DEFI672fs		Keep , Known driver
BAP1	MM046	WASH-U	3	52436865	52436868	Frame_Shift_Del	DEL	CACT	-	0.857	p.KC637fs		Keep , Known driver
BAP1	A985	TCGA	3	52437189	52437217	Frame_Shift_Del	DEL	CAGGCCTCAC CATCCCCGTCT TCTCTCTG	-	0.800	p.SREKTGMVR PG609fs		Keep , Known driver
BAP1	MM162	WASH-U	3	52437431	52437431	Splice_Site	SNP	C	G	0.855	0.000	17.27	Keep , Known driver
BAP1	AA8T	TCGA	3	52437669	52437670	Frame_Shift_Ins	INS	-	A	0.909	p.E498fs		Keep , Known driver
BAP1	A8KN	TCGA	3	52437802	52437803	Frame_Shift_Del	DEL	TT	-	0.875	p.K453fs		Keep , Known driver
BAP1	A9EU	TCGA	3	52437840	52437840	Nonsense_Mutation	SNP	G	A	0.875	p.Q441*	16.83	Keep , Known driver
BAP1	A984	TCGA	3	52437899	52437899	Frame_Shift_Del	DEL	T	-	0.783	p.K421fs		Keep , Known driver

BAP1	A980	TCGA	3	52439229	52439229	Frame_Shift_Del	DEL	G	-	0.286	p.P339fs		Keep , Known driver
BAP1	MM070	WASH-U	3	52439264	52439274	Frame_Shift_Del	DEL	GTGGGATGGGG	-	0.900	p.APSH323fs		Keep , Known driver
BAP1	MM161	WASH-U	3	52439813	52439814	Frame_Shift_Del	DEL	CT	-	0.667	p.R300fs		Keep , Known driver
BAP1	A9EV	TCGA	3	52440323	52440363	Frame_Shift_Del	DEL	GGCCTCATACTTGATCCTGCGGTCGGGCACCACTGCCATCA	-	0.423	p.LMAVVPDRRIKYEA230fs		Keep , Known driver
BAP1	A9EF	TCGA	3	52440383	52440383	Nonsense_Mutation	SNP	G	C	0.975	p.Y223*	39	Keep , Known driver
BAP1	MM056	WASH-U	3	52440916	52440916	Nonsense_Mutation	SNP	C	T	1.000	p.W196*	39	Keep , Known driver
BAP1	MM151	WASH-U	3	52440917	52440925	Splice_Site	DEL	CAGGGCCCT	-	0.933	p.EGP194del		Keep , Known driver
BAP1	AA9A	TCGA	3	52441217	52441217	Missense_Mutation	SNP	C	G	1.000	p.G185R	33	Keep , Known driver
BAP1	A888	TCGA	3	52441252	52441252	Missense_Mutation	SNP	T	C	0.160	p.Y173C	24.1	Keep , Known driver
BAP1	MM066	WASH-U	3	52441263	52441263	Missense_Mutation	SNP	G	C	0.500	p.H169Q	18.26	Keep , Known driver
BAP1	A9EE	TCGA	3	52441417	52441418	Frame_Shift_Ins	INS	-	AAGGCAAAGCTTCA	0.222	p.-145fs		Keep , Known driver
BAP1	MM121	WASH-U	3	52441470	52441470	Missense_Mutation	SNP	C	G	0.310	p.G128R	27.6	Keep , Known driver
BAP1	A8KK	TCGA	3	52441978	52441978	Frame_Shift_Del	DEL	G	-	0.850	p.P124fs		Keep , Known driver
BAP1	MM116	WASH-U	3	52441995	52441995	Frame_Shift_Del	DEL	G	-	0.900	p.F118fs		Keep , Known driver
BAP1	AA8N	TCGA	3	52442512	52442512	Missense_Mutation	SNP	T	C	0.778	p.N78S	20	Keep , Known driver
BAP1	A9F8	TCGA	3	52442542	52442542	Missense_Mutation	SNP	T	C	0.923	p.D68G	19.57	Keep , Known driver
BAP1	A88A	TCGA	3	52442548	52442579	Frame_Shift_Del	DEL	ACCAAGGTAGAGACCTTTCGCCGGGACCGGCG	-	1.000	p.RRSRRKVSTLV56fs		Keep , Known driver
BAP1	MM120	WASH-U	3	52442600	52442600	Frame_Shift_Del	DEL	G	-	0.810	p.L49fs		Keep, Known driver
BAP1	A9EX	TCGA	3	52443574	52443574	Nonsense_Mutation	SNP	G	A	0.809	p.Q40*	34	Keep , Known driver
BAP1	A9F1	TCGA	3	52443593	52443593	Nonsense_Mutation	SNP	G	C	0.743	p.Y33*	28.4	Keep , Known driver
BAP1	A8KL	TCGA	3	52443612	52443613	Frame_Shift_Ins	INS	-	C	0.928	p.V27fs		Keep , Known driver
BAP1	MM100	WASH-U	3	52443748	52443782	Splice_Site	DEL	GGGTGAAGAGGCCTGGGTGGGGCGACAAGAGGAGG	-	0.500	p.ASSCR13fs		Keep , Known driver
BAP1	MM179	WASH-U	3	52443876	52443876	Nonsense_Mutation	SNP	C	A	1.000	p.E7*	37	Keep , Known driver
BAP1	MM179	WASH-U	3	52443877	52443877	Silent	SNP	C	A	1.000	p.L6L		Discard - Silent variant
C7orf49	A8KD	TCGA	7	134851406	134851408	In_Frame_Del	DEL	TCC	-	0.032	p.143_144EE>E		Discard, recurrent Indel - likely artefact
C7orf49	A9EH	TCGA	7	134851406	134851408	In_Frame_Del	DEL	TCC	-	0.032	p.143_144EE>E		Discard, recurrent Indel - likely artefact
C7orf49	A983	TCGA	7	134851408	134851409	In_Frame_Ins	INS	-	TCC	0.537	p.143_143E>EE		Discard, recurrent Indel - likely artefact

DNAJB7	MM038	WASH-U	22	41257836	41257836	Missense_Mutation	SNP	C	T	0.042	p.V55I	24.5	Discard, same cohort - recurrent SNP
DNAJB7	MM038	WASH-U	22	41257836	41257836	Missense_Mutation	SNP	C	G	0.042	p.V55L	27.4	Discard, same cohort - recurrent SNP
DNAJB7	MM050	WASH-U	22	41257836	41257836	Missense_Mutation	SNP	C	G	0.048	p.V55L	27.4	Discard, same cohort - recurrent SNP
DNAJB7	MM070	WASH-U	22	41257836	41257836	Missense_Mutation	SNP	C	T	0.056	p.V55I	24.5	Discard, same cohort - recurrent SNP
DNAJB7	MM080	WASH-U	22	41257836	41257836	Missense_Mutation	SNP	C	T	0.041	p.V55I	24.5	Discard, same cohort - recurrent SNP
DNAJB7	MM134	WASH-U	22	41257836	41257836	Missense_Mutation	SNP	C	T	0.045	p.V55I	24.5	Discard, same cohort - recurrent SNP
DNAJB7	AA8S	TCGA	22	41257997	41257997	Start_Codon_SNP	SNP	A	G	0.543	p.M1T	13.65	Keep
EIF1AX	A9ZY	TCGA	X	20152122	20152122	Missense_Mutation	SNP	A	T	0.923	p.W70R	22.5	Keep , Known driver
EIF1AX	A87U	TCGA	X	20156713	20156713	Missense_Mutation	SNP	C	T	1.000	p.G15D	18.11	Keep , Known driver
EIF1AX	A9F7	TCGA	X	20156713	20156713	Missense_Mutation	SNP	C	T	0.225	p.G15D	18.11	Keep , Known driver
EIF1AX	A9EC	TCGA	X	20156731	20156731	Missense_Mutation	SNP	C	T	0.558	p.G9D	18.41	Keep , Known driver
EIF1AX	MM078	WASH-U	X	20156731	20156731	Missense_Mutation	SNP	C	A	1.000	p.G9V	17.25	Keep , Known driver
EIF1AX	MM159	WASH-U	X	20156731	20156731	Missense_Mutation	SNP	C	G	1.000	p.G9A	16.97	Keep , Known driver
EIF1AX	A8KE	TCGA	X	20156732	20156732	Missense_Mutation	SNP	C	G	0.879	p.G9R	17.8	Keep , Known driver
EIF1AX	A884	TCGA	X	20156735	20156735	Missense_Mutation	SNP	C	G	0.530	p.G8R	18.33	Keep , Known driver
EIF1AX	A8KO	TCGA	X	20156735	20156737	In_Frame_Del	DEL	CTT	-	1.000	p.7_8KG>R		Keep , Known driver
EIF1AX	A9EY	TCGA	X	20156735	20156735	Missense_Mutation	SNP	C	T	0.417	p.G8R	18.91	Keep , Known driver
EIF1AX	A880	TCGA	X	20156740	20156740	Splice_Site	SNP	C	T	0.931	p.G6D	18.74	Keep , Known driver
EIF1AX	AA8R	TCGA	X	20156740	20156740	Splice_Site	SNP	C	T	0.900	p.G6D	18.74	Keep , Known driver
EIF1AX	MM086	WASH-U	X	20156740	20156740	Splice_Site	SNP	C	T	1.000	p.G6D	18.74	Keep , Known driver
EIF1AX	MM038	WASH-U	X	20156741	20156741	Splice_Site	SNP	C	T	0.836	0.000	19.87	Keep , Known driver
EIF1AX	MM105	WASH-U	X	20156742	20156742	Splice_Site	SNP	T	C	0.800	0.000	18.5	Keep , Known driver
EIF1AX	MM018	WASH-U	X	20159749	20159749	Missense_Mutation	SNP	T	A	1.000	p.N4Y	16.34	Keep , Known driver
EIF1AX	MM050	WASH-U	X	20159752	20159752	Missense_Mutation	SNP	T	C	0.333	p.K3E	18.92	Keep , Known driver
FAM133B	MM016	WASH-U	7	92195326	92195326	Splice_Site	SNP	G	T	0.994	0.000	18.93|18.93	Keep
FAM133B	MM055	WASH-U	7	92195342	92195342	Nonsense_Mutation	SNP	G	A	0.023	p.R215*	19.59	Discard, same cohort - recurrent SNP
FAM133B	MM065	WASH-U	7	92195342	92195342	Nonsense_Mutation	SNP	G	A	0.015	p.R215*	19.59	Discard, same cohort - recurrent SNP
FAM133B	MM080	WASH-U	7	92195342	92195342	Nonsense_Mutation	SNP	G	A	1.000	p.R215*	19.59	Discard, same cohort - recurrent SNP, read depth =1
GNA11	A985	TCGA	19	3115012	3115012	Missense_Mutation	SNP	C	T	0.451	p.R183C	20.4	Keep , Known driver
GNA11	MM138	WASH-U	19	3115012	3115012	Missense_Mutation	SNP	C	T	0.509	p.R183C	20.4	Keep , Known driver

GNA11	A87W	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.370	p.Q209L	15.05	Keep , Known driver
GNA11	A87Y	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.269	p.Q209L	15.05	Keep , Known driver
GNA11	A881	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.465	p.Q209L	15.05	Keep , Known driver
GNA11	A882	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.385	p.Q209L	15.05	Keep , Known driver
GNA11	A884	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.483	p.Q209L	15.05	Keep , Known driver
GNA11	A888	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.340	p.Q209L	15.05	Keep , Known driver
GNA11	A88A	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.391	p.Q209L	15.05	Keep , Known driver
GNA11	A8K8	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.479	p.Q209L	15.05	Keep , Known driver
GNA11	A8KA	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.389	p.Q209L	15.05	Keep , Known driver
GNA11	A8KF	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.465	p.Q209L	15.05	Keep , Known driver
GNA11	A8KG	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.457	p.Q209L	15.05	Keep , Known driver
GNA11	A8KH	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.517	p.Q209L	15.05	Keep , Known driver
GNA11	A8KL	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.364	p.Q209L	15.05	Keep , Known driver
GNA11	A980	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.219	p.Q209L	15.05	Keep , Known driver
GNA11	A9E8	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.341	p.Q209L	15.05	Keep , Known driver
GNA11	A9EF	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.525	p.Q209L	15.05	Keep , Known driver
GNA11	A9EH	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.563	p.Q209L	15.05	Keep , Known driver
GNA11	A9EI	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.365	p.Q209L	15.05	Keep , Known driver
GNA11	A9EJ	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.491	p.Q209L	15.05	Keep , Known driver
GNA11	A9EQ	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.339	p.Q209L	15.05	Keep , Known driver
GNA11	A9ET	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.429	p.Q209L	15.05	Keep , Known driver
GNA11	A9EU	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.267	p.Q209L	15.05	Keep , Known driver
GNA11	A9EV	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.278	p.Q209L	15.05	Keep , Known driver
GNA11	A9F3	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.393	p.Q209L	15.05	Keep , Known driver
GNA11	A9F5	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.455	p.Q209L	15.05	Keep , Known driver
GNA11	A9F7	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.525	p.Q209L	15.05	Keep , Known driver
GNA11	A9ZX	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.407	p.Q209L	15.05	Keep , Known driver
GNA11	A9ZY	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.441	p.Q209L	15.05	Keep , Known driver
GNA11	AA8N	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.378	p.Q209L	15.05	Keep , Known driver
GNA11	AA8P	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.440	p.Q209L	15.05	Keep , Known driver
GNA11	AA8R	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.492	p.Q209L	15.05	Keep , Known driver
GNA11	AA8T	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.383	p.Q209L	15.05	Keep , Known driver
GNA11	AA9A	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.333	p.Q209L	15.05	Keep , Known driver
GNA11	AB0B	TCGA	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.384	p.Q209L	15.05	Keep , Known driver
GNA11	MM048	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.333	p.Q209L	15.05	Keep , Known driver
GNA11	MM049	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.517	p.Q209L	15.05	Keep , Known driver
GNA11	MM066	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.400	p.Q209L	15.05	Keep , Known driver
GNA11	MM085	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.461	p.Q209L	15.05	Keep , Known driver

GNA11	MM098	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.384	p.Q209L	15.05	Keep , Known driver
GNA11	MM100	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.324	p.Q209L	15.05	Keep , Known driver
GNA11	MM103	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.471	p.Q209L	15.05	Keep , Known driver
GNA11	MM105	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.450	p.Q209L	15.05	Keep , Known driver
GNA11	MM113	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.758	p.Q209L	15.05	Keep , Known driver
GNA11	MM120	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.506	p.Q209L	15.05	Keep , Known driver
GNA11	MM131	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.417	p.Q209L	15.05	Keep , Known driver
GNA11	MM134	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.459	p.Q209L	15.05	Keep , Known driver
GNA11	MM159	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.493	p.Q209L	15.05	Keep , Known driver
GNA11	MM161	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.303	p.Q209L	15.05	Keep , Known driver
GNA11	MM162	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.532	p.Q209L	15.05	Keep , Known driver
GNA11	MM179	WASH-U	19	3118942	3118942	Missense_Mutation	SNP	A	T	0.608	p.Q209L	15.05	Keep , Known driver
GNAQ	A87T	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.485	p.Q209P	25.2	Keep , Known driver
GNAQ	A87U	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.473	p.Q209P	25.2	Keep , Known driver
GNAQ	A880	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.522	p.Q209L	29.2	Keep , Known driver
GNAQ	A883	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.381	p.Q209P	25.2	Keep , Known driver
GNAQ	A885	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.444	p.Q209P	25.2	Keep , Known driver
GNAQ	A8K7	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.409	p.Q209L	29.2	Keep , Known driver
GNAQ	A8K9	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.435	p.Q209P	25.2	Keep , Known driver
GNAQ	A8KB	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.533	p.Q209P	25.2	Keep , Known driver
GNAQ	A8KE	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.636	p.Q209P	25.2	Keep , Known driver
GNAQ	A8KK	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.336	p.Q209P	25.2	Keep , Known driver
GNAQ	A8KM	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.231	p.Q209P	25.2	Keep , Known driver
GNAQ	A8KN	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.351	p.Q209P	25.2	Keep , Known driver
GNAQ	A8KO	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.456	p.Q209L	29.2	Keep , Known driver
GNAQ	A983	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.417	p.Q209P	25.2	Keep , Known driver
GNAQ	A9E5	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.316	p.Q209P	25.2	Keep , Known driver
GNAQ	A9E7	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.537	p.Q209L	29.2	Keep , Known driver
GNAQ	A9E9	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.450	p.Q209P	25.2	Keep , Known driver
GNAQ	A9EA	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.513	p.Q209L	29.2	Keep , Known driver
GNAQ	A9EC	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.488	p.Q209P	25.2	Keep , Known driver
GNAQ	A9EE	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.542	p.Q209P	25.2	Keep , Known driver
GNAQ	A9EK	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.432	p.Q209P	25.2	Keep , Known driver
GNAQ	A9EL	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.422	p.Q209L	29.2	Keep , Known driver
GNAQ	A9EM	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.291	p.Q209L	29.2	Keep , Known driver
GNAQ	A9EO	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.323	p.Q209P	25.2	Keep , Known driver
GNAQ	A9EW	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.4	p.Q209P	25.2	Keep , Known driver
GNAQ	A9EX	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.441	p.Q209P	25.2	Keep , Known driver

GNAQ	A9EY	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.532	p.Q209P	25.2	Keep , Known driver
GNAQ	A9EZ	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.575	p.Q209P	25.2	Keep , Known driver
GNAQ	A9F0	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.429	p.Q209L	29.2	Keep , Known driver
GNAQ	A9F1	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.405	p.Q209P	25.2	Keep , Known driver
GNAQ	A9F2	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.463	p.Q209L	29.2	Keep , Known driver
GNAQ	A9F4	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.459	p.Q209P	25.2	Keep , Known driver
GNAQ	A9F8	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.459	p.Q209P	25.2	Keep , Known driver
GNAQ	AA8M	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.481	p.Q209P	25.2	Keep , Known driver
GNAQ	AA8Q	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.301	p.Q209P	25.2	Keep , Known driver
GNAQ	AA8S	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.644	p.Q209P	25.2	Keep , Known driver
GNAQ	AA9E	TCGA	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.435	p.Q209L	29.2	Keep , Known driver
GNAQ	MM016	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.542	p.Q209L	29.2	Keep , Known driver
GNAQ	MM032	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	C	0.556	p.Q209R	28	Keep , Known driver
GNAQ	MM038	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.407	p.Q209L	29.2	Keep , Known driver
GNAQ	MM046	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.488	p.Q209P	25.2	Keep , Known driver
GNAQ	MM050	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.286	p.Q209P	25.2	Keep , Known driver
GNAQ	MM056	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.548	p.Q209L	29.2	Keep , Known driver
GNAQ	MM065	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.652	p.Q209P	25.2	Keep , Known driver
GNAQ	MM070	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.337	p.Q209P	25.2	Keep , Known driver
GNAQ	MM078	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.462	p.Q209L	29.2	Keep , Known driver
GNAQ	MM080	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.559	p.Q209P	25.2	Keep , Known driver
GNAQ	MM086	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.667	p.Q209P	25.2	Keep , Known driver
GNAQ	MM089	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.214	p.Q209L	29.2	Keep , Known driver
GNAQ	MM101	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.5	p.Q209L	29.2	Keep , Known driver
GNAQ	MM116	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.273	p.Q209P	25.2	Keep , Known driver
GNAQ	MM133	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	A	0.654	p.Q209L	29.2	Keep , Known driver
GNAQ	MM134	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.018	p.Q209P	25.2	Keep , Known driver
GNAQ	MM150	WASH-U	9	80409488	80409488	Missense_Mutation	SNP	T	G	0.359	p.Q209P	25.2	Keep , Known driver
GNAQ	A985	TCGA	9	80412493	80412493	Missense_Mutation	SNP	C	T	0.065	p.R183Q	37	Keep , Known driver
GNAQ	A9ES	TCGA	9	80412493	80412493	Missense_Mutation	SNP	C	T	0.578	p.R183Q	37	Keep , Known driver
GNAQ	A8KI	TCGA	9	80537255	80537255	Missense_Mutation	SNP	C	A	0.380	p.G48V	26.8	Keep , Known driver
GNAQ	A8KI	TCGA	9	80537256	80537256	Nonsense_Mutation	SNP	C	A	0.372	p.G48*	39	Keep , Known driver
HAUS6	A8K7	TCGA	9	19058431	19058431	Missense_Mutation	SNP	T	G	0.500	p.E778D	14.52	Keep
HAUS6	A884	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.070	p.L588F	16.52	Keep
HAUS6	A8KJ	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.065	p.L588F	16.52	Keep
HAUS6	A982	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.056	p.L588F	16.52	Keep
HAUS6	A985	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.103	p.L588F	16.52	Keep
HAUS6	A9E5	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.048	p.L588F	16.52	Keep

HAUS6	A9E8	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.130	p.L588F	16.52	Keep
HAUS6	A9E9	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.044	p.L588F	16.52	Keep
HAUS6	A9EH	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.125	p.L588F	16.52	Keep
HAUS6	A9EM	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.071	p.L588F	16.52	Keep
HAUS6	AA8O	TCGA	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.111	p.L588F	16.52	Keep
HAUS6	MM080	WASH-U	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.029	p.L588F	16.52	Keep
HAUS6	MM113	WASH-U	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.111	p.L588F	16.52	Keep
HAUS6	MM138	WASH-U	9	19060087	19060087	Missense_Mutation	SNP	C	G	0.091	p.L588F	16.52	Keep
HAUS6	MM101	WASH-U	9	19102559	19102559	Missense_Mutation	SNP	C	A	0.115	p.A31S	2.492	Discard, CADD < 10
HDGFRP3	A884	TCGA	15	83819992	83819993	Frame_Shift_Ins	INS	-	AAAAAAA	0.026	p.S194fs		Discard, recurrent Indel in single cohort - likely artefact
HDGFRP3	A8KI	TCGA	15	83819992	83819993	Frame_Shift_Ins	INS	-	AAAAAAAA	0.023	p.S194fs		Discard, recurrent Indel in single cohort - likely artefact
HDGFRP3	A8KM	TCGA	15	83819992	83819993	Frame_Shift_Ins	INS	-	AAAAAAAA	0.037	p.S194fs		Discard, recurrent Indel in single cohort - likely artefact
HDGFRP3	A9EX	TCGA	15	83819992	83819993	Frame_Shift_Ins	INS	-	AAAAAAA	0.021	p.S194fs		Discard, recurrent Indel in single cohort - likely artefact
HDGFRP3	A9F0	TCGA	15	83819992	83819993	Frame_Shift_Ins	INS	-	AAAAAAAA	0.043	p.S194fs		Discard, recurrent Indel in single cohort - likely artefact
HDGFRP3	A9EQ	TCGA	15	83819996	83819996	Missense_Mutation	SNP	T	C	0.384	p.T193A	7.809	Discard, CADD < 10
KIAA2013	A8KF	TCGA	1	11983371	11983371	Silent	SNP	C	A	0.103	p.G403G		Discard - Silent variant
KIAA2013	A87T	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.040	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A87Y	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.036	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A8K8	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.060	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A8KH	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.106	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A8KI	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.054	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A8KO	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.041	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A983	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.046	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A9EL	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.056	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A9ES	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.098	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A9EZ	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.073	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact

KIAA2013	A9F5	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.079	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	A9ZX	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.054	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	AA8T	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.143	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
KIAA2013	AA9E	TCGA	1	11985396	11985396	Missense_Mutation	SNP	A	G	0.118	p.L300P	11.11	Discard, recurrent SNP in single cohort -likely artefact
LCE2A	A888	TCGA	1	152671515	152671556	In_Frame_Del	DEL	CAGCTCTGGGGGCTGCTGCGGCTCCAGCTCTGGGGGCTGCTG	-	0.021	p.SSGGCCGSSSGGCC47del		Discard, recurrent Indel - likely artefact, seen in only single cohort
LCE2A	A8K7	TCGA	1	152671515	152671556	In_Frame_Del	DEL	CAGCTCTGGGGGCTGCTGCGGCTCCAGCTCTGGGGGCTGCTG	-	0.023	p.SSGGCCGSSSGGCC47del		Discard, recurrent Indel - likely artefact, seen in only single cohort
LCE2A	A8KB	TCGA	1	152671515	152671556	In_Frame_Del	DEL	CAGCTCTGGGGGCTGCTGCGGCTCCAGCTCTGGGGGCTGCTG	-	0.013	p.SSGGCCGSSSGGCC47del		Discard, recurrent Indel - likely artefact, seen in only single cohort
LCE2A	A8KD	TCGA	1	152671515	152671556	In_Frame_Del	DEL	CAGCTCTGGGGGCTGCTGCGGCTCCAGCTCTGGGGGCTGCTG	-	0.011	p.SSGGCCGSSSGGCC47del		Discard, recurrent Indel - likely artefact, seen in only single cohort
LCE2A	AB0B	TCGA	1	152671515	152671556	In_Frame_Del	DEL	CAGCTCTGGGGGCTGCTGCGGCTCCAGCTCTGGGGGCTGCTG	-	0.016	p.SSGGCCGSSSGGCC47del		Discard, recurrent Indel - likely artefact, seen in only single cohort
MFF	A8KD	TCGA	2	228197195	228197195	Missense_Mutation	SNP	C	T	0.466	p.P81L	22.5	Keep
MFF	A885	TCGA	2	228197303	228197304	Splice_Site	INS	-	AATCCG	1.000	p.118_119insRI		Discard, recurrent Indel - likely artefact
MFF	MM089	WASH-U	2	228197303	228197304	Splice_Site	INS	-	AATCCG	0.037	p.118_119insRI		Discard, recurrent Indel - likely artefact
MFF	MM133	WASH-U	2	228197303	228197304	Splice_Site	INS	-	AATCCGAGC	1.000	p.118_119insRAI		Discard, recurrent Indel - likely artefact
MFF	A9EA	TCGA	2	228205096	228205097	Splice_Site	INS	-	ACCTGTGTTGCGTGGTGGGTCTGCTGCC	1.000	p.L147fs		Discard, recurrent Indel - likely artefact
MFF	MM127	WASH-U	2	228205096	228205097	Splice_Site	INS	-	ACCTGTGTTGC	1.000	p.L147fs		Discard, recurrent Indel - likely artefact

NUP50	A8KA	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.037	p.T123A	0.204	Discard, CADD < 10
NUP50	A8KB	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.062	p.T123A	0.204	Discard, CADD < 10
NUP50	A8KI	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.043	p.T123A	0.204	Discard, CADD < 10
NUP50	A8KO	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.031	p.T123A	0.204	Discard, CADD < 10
NUP50	A982	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.029	p.T123A	0.204	Discard, CADD < 10
NUP50	A9E5	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.067	p.T123A	0.204	Discard, CADD < 10
NUP50	A9EK	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.070	p.T123A	0.204	Discard, CADD < 10
NUP50	A9EW	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.049	p.T123A	0.204	Discard, CADD < 10
NUP50	A9EY	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.055	p.T123A	0.204	Discard, CADD < 10
NUP50	A9ZY	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.056	p.T123A	0.204	Discard, CADD < 10
NUP50	AA8N	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.046	p.T123A	0.204	Discard, CADD < 10
NUP50	AA8S	TCGA	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.095	p.T123A	0.204	Discard, CADD < 10
NUP50	MM050	WASH-U	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.056	p.T123A	0.204	Discard, CADD < 10
NUP50	MM085	WASH-U	22	45574145	45574145	Missense_Mutation	SNP	A	G	0.042	p.T123A	0.204	Discard, CADD < 10
NUP50	A880	TCGA	22	45574520	45574520	Missense_Mutation	SNP	G	A	0.444	p.D248N	14.01	Keep
NUP50	MM056	WASH-U	22	45580363	45580363	Missense_Mutation	SNP	C	T	0.423	p.P412S	16.83	Keep
NUP50	A9EQ	TCGA	22	45580380	45580380	Silent	SNP	G	A	0.413	p.T417T		Discard, CADD < 10
RBMX2	A885	TCGA	X	129545492	129545493	Frame_Shift_Ins	INS	-	AAAA AGGA	0.16	p.-159fs		Discard, Likely Artefact
RBMX2	A983	TCGA	X	129545492	129545493	Frame_Shift_Ins	INS	-	AAAA AGGA	0.250	p.-159fs		Discard, Likely Artefact
RBMX2	A9EW	TCGA	X	129545492	129545493	Frame_Shift_Ins	INS	-	AAAA AGGA	0.111	p.-159fs		Discard, Likely Artefact
RBMX2	A9F3	TCGA	X	129545492	129545493	Frame_Shift_Ins	INS	-	AAAA AGGA	0.094	p.-159fs		Discard, Likely Artefact
RBMX2	AA9A	TCGA	X	129545492	129545493	Frame_Shift_Ins	INS	-	A	0.111	p.K159fs		Discard, Likely Artefact
RGPD3	A8KJ	TCGA	2	107039738	107039738	Missense_Mutation	SNP	T	C	0.146	p.N1562S	5.547	Discard, CADD < 10
RGPD3	A87U	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.028	p.S937R	5.022	Discard, CADD < 10
RGPD3	A87W	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.031	p.S937R	5.022	Discard, CADD < 10
RGPD3	A87Y	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.041	p.S937R	5.022	Discard, CADD < 10
RGPD3	A881	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.059	p.S937R	5.022	Discard, CADD < 10
RGPD3	A883	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.057	p.S937R	5.022	Discard, CADD < 10
RGPD3	A888	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.045	p.S937R	5.022	Discard, CADD < 10
RGPD3	A88A	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.020	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8K7	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.033	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8K8	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.033	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8K9	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.053	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KA	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.069	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KD	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.018	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KE	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.035	p.S937R	5.022	Discard, CADD < 10

RGPD3	A8KF	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.021	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KG	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.058	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KI	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.054	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KJ	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.054	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KL	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.013	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KM	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.040	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KN	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.033	p.S937R	5.022	Discard, CADD < 10
RGPD3	A8KO	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.031	p.S937R	5.022	Discard, CADD < 10
RGPD3	A983	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.030	p.S937R	5.022	Discard, CADD < 10
RGPD3	A984	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.025	p.S937R	5.022	Discard, CADD < 10
RGPD3	A985	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.028	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9E7	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.053	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9E9	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.018	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EA	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.031	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9ED	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.035	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EE	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.057	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EF	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.023	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EH	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.026	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EI	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.019	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EJ	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.024	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EK	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.058	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EL	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.035	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EO	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.084	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9ET	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.032	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EV	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.107	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EW	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.035	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EX	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.034	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EY	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.033	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9EZ	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.051	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9F0	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.038	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9F1	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.040	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9F2	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.029	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9F4	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.033	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9F5	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.059	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9F8	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.076	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9ZX	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.038	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9ZY	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.037	p.S937R	5.022	Discard, CADD < 10
RGPD3	AA8M	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.047	p.S937R	5.022	Discard, CADD < 10

RGPD3	AA8N	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.067	p.S937R	5.022	Discard, CADD < 10
RGPD3	AA8O	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.034	p.S937R	5.022	Discard, CADD < 10
RGPD3	AA8P	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.020	p.S937R	5.022	Discard, CADD < 10
RGPD3	AA8Q	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.065	p.S937R	5.022	Discard, CADD < 10
RGPD3	AA8R	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.040	p.S937R	5.022	Discard, CADD < 10
RGPD3	AA8T	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.057	p.S937R	5.022	Discard, CADD < 10
RGPD3	AA9E	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.033	p.S937R	5.022	Discard, CADD < 10
RGPD3	AB0B	TCGA	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.025	p.S937R	5.022	Discard, CADD < 10
RGPD3	MM113	WASH-U	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.019	p.S937R	5.022	Discard, CADD < 10
RGPD3	MM121	WASH-U	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.033	p.S937R	5.022	Discard, CADD < 10
RGPD3	MM138	WASH-U	2	107041612	107041612	Missense_Mutation	SNP	A	C	0.100	p.S937R	5.022	Discard, CADD < 10
RGPD3	A9E5	TCGA	2	107042530	107042530	Missense_Mutation	SNP	G	C	0.040	p.P874A	9.375	Discard, CADD < 10
RGPD3	AA8R	TCGA	2	107042546	107042547	Splice_Site	INS	-	GA	0.035	0.000		Discard, recurrent Indel - likely artefact
RGPD3	MM113	WASH-U	2	107042546	107042547	Splice_Site	INS	-	AAA	0.017	0.000		Discard, recurrent Indel - likely artefact
SF3B1	A9EC	TCGA	2	198267370	198267370	Missense_Mutation	SNP	T	G	0.438	p.T663P	26.8	Keep , Known driver
SF3B1	A881	TCGA	2	198267372	198267372	Missense_Mutation	SNP	T	C	0.493	p.H662R	25.2	Keep , Known driver
SF3B1	A885	TCGA	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.435	p.R625H	35	Keep , Known driver
SF3B1	A8KH	TCGA	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.411	p.R625H	35	Keep , Known driver
SF3B1	A9E9	TCGA	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.386	p.R625H	35	Keep , Known driver
SF3B1	A9EJ	TCGA	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.415	p.R625H	35	Keep , Known driver
SF3B1	A9ES	TCGA	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.341	p.R625H	35	Keep , Known driver
SF3B1	A9EZ	TCGA	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.429	p.R625H	35	Keep , Known driver
SF3B1	A9F4	TCGA	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.426	p.R625H	35	Keep , Known driver
SF3B1	AA9E	TCGA	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.468	p.R625H	35	Keep , Known driver
SF3B1	MM049	WASH-U	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.459	p.R625H	35	Keep , Known driver
SF3B1	MM101	WASH-U	2	198267483	198267483	Missense_Mutation	SNP	C	T	0.56	p.R625H	35	Keep , Known driver
SF3B1	MM131	WASH-U	2	198267483	198267483	Missense_Mutation	SNP	C	A	0.412	p.R625L	35	Keep , Known driver
SF3B1	A8K9	TCGA	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.514	p.R625C	19.68	Keep , Known driver
SF3B1	A8KA	TCGA	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.547	p.R625C	19.68	Keep , Known driver
SF3B1	A8KB	TCGA	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.377	p.R625C	19.68	Keep , Known driver
SF3B1	A985	TCGA	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.500	p.R625C	19.68	Keep , Known driver
SF3B1	A9EA	TCGA	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.429	p.R625C	19.68	Keep , Known driver
SF3B1	A9EW	TCGA	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.421	p.R625C	19.68	Keep , Known driver
SF3B1	MM010	WASH-U	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.425	p.R625C	19.68	Keep , Known driver
SF3B1	MM032	WASH-U	2	198267484	198267484	Missense_Mutation	SNP	G	C	0.400	p.R625G	18.65	Keep , Known driver
SF3B1	MM065	WASH-U	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.414	p.R625C	19.68	Keep , Known driver

SF3B1	MM133	WASH-U	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.481	p.R625C	19.68	Keep , Known driver
SF3B1	MM134	WASH-U	2	198267484	198267484	Missense_Mutation	SNP	G	A	0.515	p.R625C	19.68	Keep , Known driver
WHAMM	A883	TCGA	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.080	p.M237T	0.01	Discard, CADD < 10
WHAMM	A8KF	TCGA	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.069	p.M237T	0.01	Discard, CADD < 10
WHAMM	A8KI	TCGA	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.064	p.M237T	0.01	Discard, CADD < 10
WHAMM	A8KL	TCGA	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.111	p.M237T	0.01	Discard, CADD < 10
WHAMM	A8KN	TCGA	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.100	p.M237T	0.01	Discard, CADD < 10
WHAMM	A982	TCGA	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.057	p.M237T	0.01	Discard, CADD < 10
WHAMM	A9EJ	TCGA	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.111	p.M237T	0.01	Discard, CADD < 10
WHAMM	A9EQ	TCGA	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.069	p.M237T	0.01	Discard, CADD < 10
WHAMM	MM050	WASH-U	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.091	p.M237T	0.01	Discard, CADD < 10
WHAMM	MM101	WASH-U	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.133	p.M237T	0.01	Discard, CADD < 10
WHAMM	MM103	WASH-U	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.136	p.M237T	0.01	Discard, CADD < 10
WHAMM	MM161	WASH-U	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.125	p.M237T	0.01	Discard, CADD < 10
WHAMM	MM179	WASH-U	15	83481955	83481955	Missense_Mutation	SNP	T	C	0.031	p.M237T	0.01	Discard, CADD < 10
WHAMM	MM010	WASH-U	15	83482024	83482024	Missense_Mutation	SNP	T	A	0.063	p.I260N	20.8	Keep
WHAMM	MM050	WASH-U	15	83482024	83482024	Missense_Mutation	SNP	T	A	0.139	p.I260N	20.8	Keep
WHAMM	MM055	WASH-U	15	83482024	83482024	Missense_Mutation	SNP	T	A	0.058	p.I260N	20.8	Keep
WHAMM	MM065	WASH-U	15	83482024	83482024	Missense_Mutation	SNP	T	A	0.052	p.I260N	20.8	Keep
WHAMM	MM100	WASH-U	15	83482024	83482024	Missense_Mutation	SNP	T	A	0.108	p.I260N	20.8	Keep
WHAMM	MM113	WASH-U	15	83482024	83482024	Missense_Mutation	SNP	T	A	0.117	p.I260N	20.8	Keep
WHAMM	MM032	WASH-U	15	83495511	83495511	Missense_Mutation	SNP	C	A	0.688	p.D518E	17.39	Keep
WHAMM	AA8P	TCGA	15	83499532	83499532	Missense_Mutation	SNP	A	G	0.500	p.K608R	6.165	Discard, CADD < 10

Appendix 5 - Candidate gene fusions after filtering steps

sample	fusion genes	chrom1	base1	strand1	chrom2	base2	strand2	gap (kb)	spanning pairs	spanning reads	inframe	aligns	rearrangement	contig	contig break	classification	known (polymorphism)
MM0171	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	0	10	FALSE	TRUE	TRUE	HWI-D00467	36	MediumConfidence	Yes
MM010	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	2	7	FALSE	TRUE	TRUE	HW-ST997_0	42	HighConfidence	Yes
MM091	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	0	7	FALSE	TRUE	TRUE	HWI-D00467	49	MediumConfidence	Yes
MM065	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	2	6	FALSE	TRUE	TRUE	HW-ST997_0	29	HighConfidence	Yes
MM0144	KANSL1:ARL17A	chr17	44171926	-	chr17	44648235	-	476.305	0	4	FALSE	TRUE	TRUE	HWI-D00467	50	MediumConfidence	Yes
MM0171	TRIO:DNAJC21	chr5	14143991	+	chr5	34950275	+	20806.29	0	5	FALSE	TRUE	FALSE	HWI-D00467	42	MediumConfidence	
MM0171	MYO10:DST	chr5	16783444	-	chr6	56600087	-	Inf	3	53	FALSE	TRUE	TRUE	HWI-D00467	62	HighConfidence	-
MM0171	MYO10:DST	chr5	16794755	-	chr6	56600087	-	Inf	3	11	FALSE	TRUE	TRUE	HWI-D00467	58	HighConfidence	-
MM080	TFG:GPR128	chr3	1E+08	+	chr3	1E+08	+	90.458	1	5	TRUE	TRUE	TRUE	HW-ST997_0	53	HighConfidence	Yes
MM082	RP11-680G10.1:GSE1	chr16	85391249	+	chr16	85682158	+	290.914	5	3	TRUE	TRUE	FALSE	HW-ST997_0	65	HighConfidence	-
MM0173	RP11-680G10.1:GSE1	chr16	85391249	+	chr16	85667520	+	276.274	1	3	TRUE	TRUE	FALSE	HWI-D00467	47	HighConfidence	-
MM0171	KPNA1:QRICH1	chr3	1.22E+08	-	chr3	49070661	-	73085.36	0	4	FALSE	TRUE	FALSE	HWI-D00467	44	MediumConfidence	-