Chapter 1

Background

1.1 Melanoma - A statistical overview

Melanoma is a highly aggressive type of skin cancer with a poor prognosis in the advanced stage. It originates from melanocytes, a type of skin cell which are responsible for the production of pigments called melanin. The incidence rates of people suffering from malignant melanoma has increased by 134% since the early 90s and has gone up by 45% since 2006 in the United Kingdom[1]. Skin cancer is the most commonly diagnosed type of cancer in the United States of America with melanoma comprising of less than 1% of total skin cancer cases but the vast majority of skin cancer related deaths[2]. Around 91,270 people are expected to be diagnosed with melanoma in 2018 with 9,320 estimated deaths in the United States of America[2]. Comparatively, melanoma consisted of 10% of all skin cancer diagnoses in 2016 but was responsible for 63% of skin-cancer related deaths in the United Kingdom in 2016[1]. The overall incidence of melanoma in the UK is expected to increase by 7% by 2035[1]. Malignant melanoma mortality rates in the UK have increased by 156% since the 1970's[1]. Melanoma has the second highest increase in mortality (after liver cancer) amongst all cancers for men in the last ten years and the fifth highest increase in mortality for women. The 5 year survival status for melanoma patients is quite high: 99% of all cases in the United States of America[2] and 100% of all cases in the UK were expected to survive for local melanomas[3]. However, this dropped to less than 20% for both for distant metastatic melanomas[2, 3]. 1 in 10 melanoma cases diagnosed were estimated to be metastatic in the UK at diagnosis[1].

1.2 Melanoma through the ages

Melanoma (derived from the Greek words *melas* meaning “dark” and *-oma* referring to abnormal growth or tumours - such as carcinoma or lymphoma) was first coined by Dr Robert Carswell in 1838 as part of his seminal work “Illustrations of the Elementary Forms of Disease”. The earliest chronicles of melanoma as a disease came from Hippocrates in the 5th century B.C. and Rufus of Eupheses in the 1st century A.D. [4]. Metastases in the skeletons of nine mummies from the Pre-Colombian Incas of Peru dated back to 4th century B.C. is commonly cited as the first chronological physical evidence of melanoma[4], a claim that has however been recently disputed. A "cancerous fungous excrescence" was surgically resected by John Hunter, which was eventually confirmed to be a metastatic melanoma tumour. Metastatic melanoma was first described by the French physician Rene Laennec in 1806[5]. He observed metastatic melanoma tumours while studying granulomas in lungs and defined these black masses as melanoses.

The first known description of familial melanoma was provided by Dr. William Norris in 1820[6]. While studying a patient, he observed several salient characteristics: the tumour originated from a mole, the father of the patient(who eventually died of his ailments) also died due to a similar disease and had tumours originating from moles and the son of the patient had similar distribution of moles across his body. These observations led him to conclude that the “disease is hereditary” which was eventually characterised and defined as Familial atypical multiple mole-melanoma syndrome (FAMM) in 1978[7].

The impact of ultraviolet radiation from sunlight on melanoma onset was made in 1956 by Henry Lancaster, who discerned that the intensity of sunlight had a direct influence on the risk of melanoma development, with particular relevance to Caucasian populations[8]. He went on to eventually link features of the skin with playing a principal part in melanoma development.

Different metrics for measuring melanoma progression were designed by Wallace Clark and Alexander Breslow in the late 1960’s. Wallace Clark devised a framework called as Clark’s levels which compares melanoma progression to the level of invasion of the tumour into the skin (Figure 1.1)[9]. A Level I tumour would be restricted to the upper layer of the epidermis while a Level V tumour would be completely invasive and would already be in the subcutaneous tissue. Alexander Breslow, around the same time, determined that melanoma progression is also linked to the size of the tumour as opposed to just the level of invasion, with specific emphasis on the thickness[10]. This was a measurement of the distance between the uppermost layer of the skin to the deepest point of tumour penetration in millimeters and was later defined as Breslow thickness. These metrics of classification are explained in detail

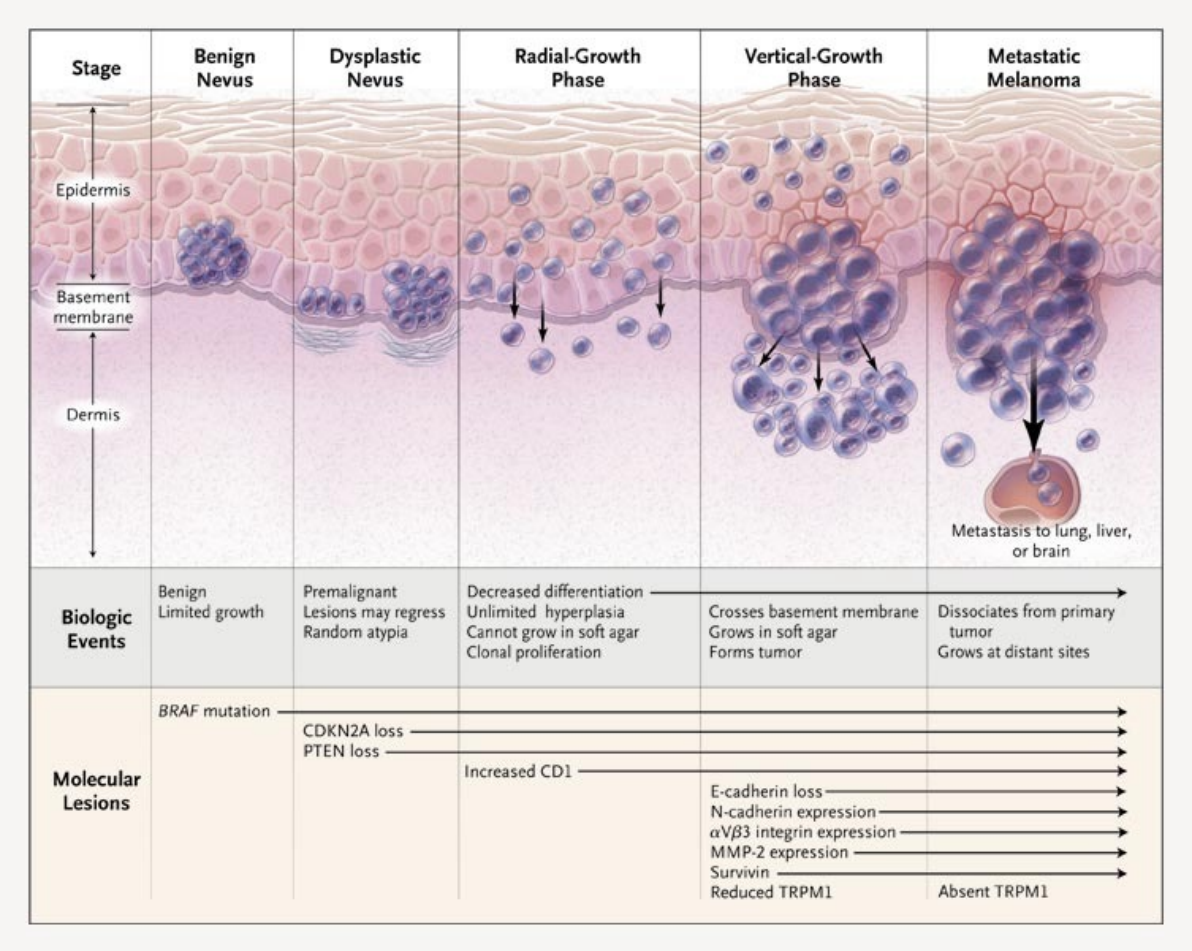

Figure 1.1: The five stages of melanoma progression as described in Clark's model. Reproduced with permission from [11].

in Section 1.4.2.1.

Over time, several insights into the progression of cancer development led to the discovery that genetic mutations in genes related to cancer development, primarily classified into oncogenes and tumour suppressors, play a vital role in tumour formation. One of the earliest group of genes that were determined to play a role in the context of melanoma were the RAS family of genes. *NRAS* was discovered as a novel human transforming gene in 1983[12] with *NRAS* mutations in melanoma cell lines being ascertained soon after[13]. Activating *KRAS* and *HRAS* mutations in melanoma were also eventually discovered. It is estimated that roughly 15-20% of all melanomas have mutations in *NRAS[14]*with 15% of all cancers carrying a RAS mutation[15]. This period also saw the discovery of the RAF family of oncogenes, namely *ARAF*[16], *BRAF[17]* and *CRAF*[18]. These genes play a significant role in the RAS/Raf/MEK/ ERK pathway (Section 1.3.3.1). *BRAF* in particular is responsi-

ble for the activation of *MEK1* and *MEK2*. Multiple large scale genetic and genomic screens have since determined *BRAF* to be the most mutated driver gene in melanoma with roughly 50% of all melanomas carrying a *BRAF* mutation[19]. A majority of these mutations, almost 90%, are a specific substitution at the 600th amino acid where a valine is replaced with a glutamic acid (*V600E*)[19]. It is to be noted that *BRAF* and *NRAS* mutations are usually mutually exclusive and, in a study by the Cancer Genome Atlas network, were identified as being responsible for almost 80% of all melanoma tumours[20]. They also show phenotypic difference as $BRAF^{V600E}$ mutations occur in younger patients having non-chronically sun exposed melanoma[21] while *NRAS* mutations occur in older patients with much higher sun-exposure[14]. The role of *BRAF* and *NRAS* mutations in the development of melanoma is discussed in Section 1.3.3. The relationship between skin characteristics and melanoma development, as predicted by Lancaster[8], were finally established through the discovery of Melanocortin receptor 1 (*MC1R*), a gene that was initially determined as a marker for hair-colour and paleness of skin[22]. Experiments confirmed the presence of several low penetrant-high frequency variants in *MC1R* which not only determined these traits but also, response to tanning/sun exposure[23] and risk of melanoma development[24]. The function of *MC1R* in melanin production is discussed in Section 1.3.1.

The early 1990's also saw the discovery of the most prominent familial melanoma gene, *Cyclin-dependent kinase inhibitor 2A* (Section 1.3.4.2), through linkage analysis[25, 26]. This has since then been determined as the single biggest driver gene in familial melanoma with roughly 40% of all familial melanoma genes carrying a *CDKN2A* driver mutation[27]. The growth of next-generation sequencing in the early 2000's and the availability of whole-exome and whole genome sequencing has helped discover multiple low-frequency driver genes in melanoma. *NF1*, *CDK4*, *BAP1*, *POT1* and *TERT* have all been established as being responsible for a significant number of melanoma cases. These genes are discussed in Sections 1.3.3 and 1.3.4. However, despite these efforts, roughly 20 percent of all melanomas and 50% of all familial melanoma cases have unknown and undiscovered genetic driver mutations. Several ongoing projects across the world are aiming to address this conundrum including this project.

1.3 The biology of melanoma

1.3.1 An introduction to melanocytes

The skin primarily comprises of three layers: The epidermis, the dermis and the subcutaneous tissue (hypodermis). Melanocytes are specific types of skin cells that exist between the inner

most layer of the epidermis (also called the basal layer) and the dermis. There are roughly 1,200 melanocytes per square millimeter of the epidermis in the average human[28]. They are neural crest-derived cells responsible for the production of melanin; a pigment that plays a key role in the protection of keratinocytes in the skin against ultra-violet radiation. Melanin are also responsible for determining the phenotype of the skin colour of the organism[29]. Almost all melanocytes are derived from the neural crest with the exception of retinal pigment epithelium (RPE), which are derived from the neuroepithelium[30]. Although they are primarily present in skin, melanocytes have also been identified in the eye, cochlea, heart, brain and the adipose tissue[31]. Melanin produced by melanocytes can be classified into three types: eumelanin, pheomelanin and neuromelanin.

- Eumelanin is a dark brown-black heterogenous polymer that act as a protective layer to the skin and help by absorbing hazardous solar radiation, particularly UV radiation[32]. Eumelanin is primarily responsible for the beneficial effects of melanin in terms of protection from chronic sun exposure. Eumelanin also functions as a free radical scavenger and superoxide dismutase that reduce reactive oxygen species (ROS)[33, 34].

- Pheomelanin is a reddish-brown pigment that is composed of sulfur-containing benzothiazine and benzothiazole derivatives[35]. It is responsible for the occurrence of red hair and freckles in the general populace. They are unstable in the presence of light and do not offer much protection against radiation; contrarily they may be responsible for the onset of carcinogenesis[32]. This is reflected in the rate of skin cancer in different ethnic populations as the rate of skin cancer development in populations with lighter skin was more than 70 times the rate of skin cancer development in darker populations[32].

- Neuromelanin is a black-brown pigment found within the brain that is a mixture of both eumelanin and pheomelanin[36]. Not much is known about its function.

The primary components of the epidermis which cells called keratinocytes. Production of melanin from melanocytes is instigated when keratinocytes in the epidermis are exposed to sun-light. When keratinocytes are exposed to ultra violet radiation, they produce several products including α-melanocyte stimulating hormone (α-MSH). αMSH in turn binds to a specific receptor called the melanocortin 1 receptor (*MC1R*) which is expressed in melanocytes[37]. The binding of αMSH to *MC1R* initiates the production of melanin, particularly eumelanin. Once this process has been started, mechanisms of melanin synthesis and survival are initiated within the melanocytes. Melanin is then shipped to cellular organelles within the melanocytes called melanosomes[38]. Once the melanosomes mature, they are transferred

through the melanocytic dendrites of the melanocyte to the keratinocytes in the epidermis[39]. A single melanocyte can produce and transfer melanin for up to 40 keratinocytes through the melanocytic dendrites[40].

Mutations in the melanin production process are an integral part of several diseases ranging from pigmental disorders like Neurofibromatosis type 1[41] and Occulocutaneous Albinism[**?**] to skin cancers like melanoma[42].

1.3.2 The progression of melanocytes to metastatic melanoma

Melanoma, as mentioned in Section 1.1, is an aggressive form of skin cancer that develops from uncontrolled proliferation of melanocytes and is referred to as cutaneous melanoma when it occurs in the upper layers of the skin. Cutaneous melanoma can be classified into two major categories based on sun-exposure, chronically sun damaged (CSD) melanoma and non-chronically sun damaged (non-CSD) melanoma. CSD melanoma are associated with increased UV mutations and are observed in parts of the body exposed to the sun while non-CSD are linked to highly penetrant genetic mutations. While there is evidence to suggest that melanoma progression and development predominantly arises de novo[43, 44], a significant percentage of melanoma cases are due to tumours formed from malignant melanocytes associated with pre-existing nevi. The presence of a high number of common and atypical nevi, also known as an atypical mole syndrome, is highly predictive of increased melanoma risk[45]. Previously GWAS studies focussing on nevi counts as a risk phenotype for melanoma have helped identify key driver genes in melanoma[46]. The genes identified from this study are discussed in Section 1.3.4.1.The development of melanoma from melanocytes occurs through several stages, shown in Figure 1.2, and was described in detail by Shain et al[47]. These stages are summarized here.

i) Melanocytic nevus

A melanocytic nevus is a benign neoplasm consisting of melanocytes that appear as a raised, pigmented irregularity. While the majority of such nevi have a low probability of progression to malignancy, increased nevus counts are associated with increased melanoma risk[48]. Melanomas arising from pre-existing nevi are associated with non-CSD melanomas, particularly with superficial spreading melanoma, which are observed with a much lower frequency in CSD melanomas[49]. A specific mutation in the *BRAF* gene called the $BRAF^{V600E}$ variant, described in Section 1.3.3.2, has been identified as a triggering event for nevi formation[50]. While skin related phenotypes such as tanning ability and pigmentation are risk factors for melanoma development, none of the variants in affected genes for these phenotypes play a

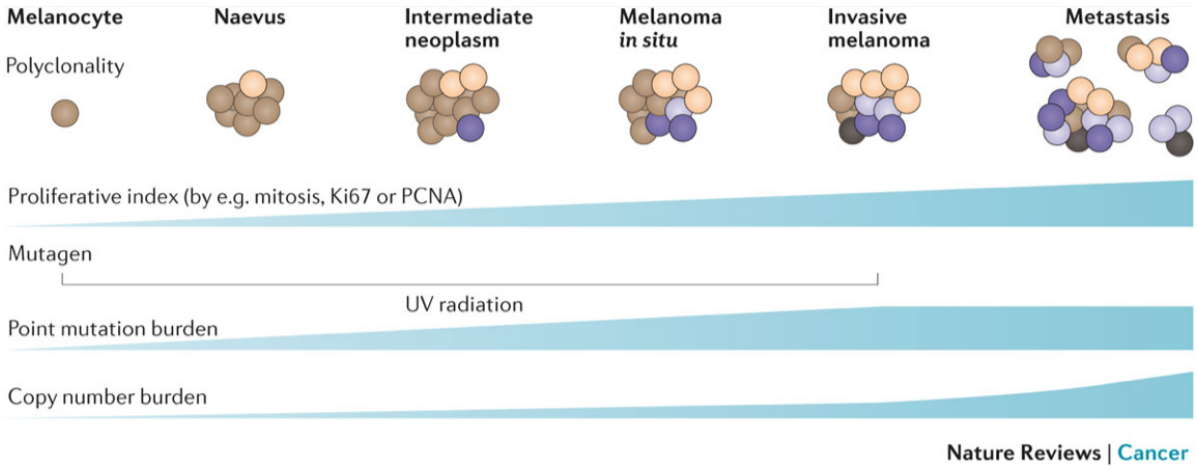

Figure 1.2: Biological characteristics for the progression of melanoma from melanocytes to metastasis. Reproduced with permission from [47].

role in melanocytic nevi formation[51]. However, lighter skin implies weaker protection of the skin against UV radiation which results in higher UV related mutagenesis. This is seen as an implicating factor in nevi development and is therefore linked to increased nevi counts[52–54]. Thus, $BRAF^{V600E}$ variants are seen as the only driver events for melanocytic nevi formation.

ii) Dysplastic nevus

A dysplastic nevus is a transitional category of melanocytic neoplasm with histopathological features that are between a benign nevus and a malignant melanoma[55]. Due to the large variance in the characteristics of such nevi, there is a significant variation in the classification and diagnosis of these neoplasms. They are clinically defined as having a minimum diameter of 5 mm and possessing two of the following characteristics: variable pigmentation, asymmetry and/or irregular or indistinct borders[56]. This type of nevi is commonly observed in familial melanoma pedigrees. While benign melanocytic nevi only carry $BRAF^{V600E}$ mutations, dysplastic nevi carry variants in other pathways associated with melanoma development such as MAPK signalling, telomerase regulation (*TERT* promoter variants) and cell cycle regulation (*CDKN2A* variants)[57]. This indicates that melanocytic nevi may carry pathogenic variants but the proliferation is controlled by regulatory pathways such as the MAPK pathway. Additional mutations that disrupt the MAPK pathway could result in increased proliferation and growth of melanocytic nevi to dysplastic nevi. However, some dysplastic nevi also have a higher proportion of *NRAS* and *BRAF* mutations which are not V600E, implying that they may occur independently without developing from a melanocytic nevus[57].

iii) Melanoma *in situ*

The first point of malignancy is the formation of a melanoma in-situ by the prolifera-

tion of melanocytes with asymmetrical growth contained within the epithelium without basal invasion[58]. Melanoma *in situ* is considered a precursor to invasive melanoma. Depending on the type of growth of melanocytes, two distinct types of melanoma are formed : Superficial spreading melanoma and lentigo maligna melanoma, discussed in Sections 1.4.1.1 and 1.4.1.3 respectively.

Superficial spreading melanoma is a result of a pagetoid growth pattern where the growth spreads both upwards and downwards through the different layers of the epidermis. They are commonly observed in melanomas arising from a pre-existing nevus which may be benign or dysplastic. These types of melanoma also have a higher proportion of $BRAF^{V600E}$ mutations[21].

Lentigo maligna melanoma on the other hand is associated with lentiginous growth and is seen in melanomas linked to chronic sun damage. It is also associated with a low proportion of *BRAF* V600E mutations[21] Such a growth is generally not linked to pre-existing nevi and surfaces independently. Acral melanomas, discussed in Section 1.4.1.4, also exhibit lentiginous melanoma.

iv) Invasive melanoma

When melanocytes from a melanoma *in situ* invade secondary layers of the skin and spread into the dermis from the epithelium, they are termed as invasive melanoma. For melanocytes to reach this stage, they must carry multiple driver mutations over its progression, including variants disrupting the MAPK pathway (described in Section 1.3.3.1) and *TERT* promoter mutations (described in Section 1.3.4.4). However, in addition to such mutations, invasive melanomas have a high fraction of p16INK4a inactivation[57]. This is a tumour suppressor protein generated from *CDKN2A* (discussed in Section 1.3.4.2) which is responsible for cell cycle regulation. With prior evidence towards the loss of *INK4A* developing highly penetrant invasive melanoma, it is increasingly evident that *INK4A* plays a critical role in preventing the transition of melanoma *in situ* to invasive melanoma. However, some invasive melanomas also preserve functional *INK4A*, implying that the G1/S checkpoint is disrupted through other mutations. One such alternative mechanism for the promotion of invasive melanoma are variants in *ARID1A* and *ARID2*, which are important components of the SWI/SNF chromatin-remodeling complex[57]. These genes also function as tumour suppressors in melanoma but function by maintaining genomic stability[59]. Loss of function of *ARID1A* or *ARID2* leads to increased chromosomal aberrations which promotes the proliferation of melanocytes. Additionally, *PTEN* and *TP53* mutations have also been reported in a subset of invasive melanomas but these mutations are in much lower frequencies in primary melanomas, indicating that such mutations are formed in the later stages of tumour progression[57].

v) Metastatic melanoma

A melanoma is said to become metastatic when tumour cells have dispersed into other organs and tissues beyond the site of origin and is the final stage of melanoma progression. Metastatic melanoma is usually observed at the regional lymph nodes and progressively spreads to distal sites. However, circulating tumour cells have previously been observed in melanoma cases without any metastases and even after the patients were considered to be disease free, indicating that the metastatic process does not occur sequentially but rather concurrently[60]. The presence of initial metastases at regional lymph nodes could therefore just be an early hallmark of metastatic progression, which has been seen in other cancers. Additionally, a small percentage of individuals without a recognizable primary tumour have been identified to have metastases. Such types of melanoma occurrences are known as melanomas of unknown primary (MUP)[61]. Several hypotheses exist for the presence of such melanomas. The primary theory of origin is the regression theory proposed by Smith and Stehlin in 1965 which suggests that the disappearance of the primary melanoma is due to spontaneous regression of the tumour post-metastasis[62]. Other cases are attributed to misdiagnosis of the primary tumour or unreported treatment/excision of a suspect lesion[61]. Similar to primary melanomas, MUPs have a high burden of UV radiation-induced mutations[63]. This implies that sun exposure plays a role in the genesis of such tumours.

1.3.3 The landscape of somatic variation in melanoma

A significant proportion of melanocytic tumours are created due to somatic mutations in different parts of the MAPK(Ras-Raf-MEK-ERK) pathway. As previously mentioned in Section 1.2, mutations in *BRAF* and *NRAS* have been observed in 80% of all melanoma tumours. However, *BRAF* and *NRAS* mutations are not solely responsible for the formation of melanocytic nevi; on the contrary, they have observed to be present commonly in benign nevi too. The Cancer Genome Atlas (TCGA) is a project started in 2005 to identify and catalogue the different genetic and genomic variations in different types of cancers. Genomic alterations in cutaneous melanoma were reported by TCGA in 2015; cutaneous melanoma was suggested to be classified into one of 4 subtypes: *BRAF* mutant melanoma, *RAS* mutant melanoma, *NF1* mutant melanoma and triple-wild-type (Triple WT) melanoma[20]. While such a classification is practical from a therapeutive and palliative perspective, there are also other mutations such as those in the triple WT classification. In such cases, the cause for disease onset is not very clear. This section explores the role of the MAPK pathway and the different subtypes of cutaneous melanoma as defined by the TCGA.

1.3.3.1 The MAPK (Ras-Raf-MEK-ERK) Pathway

The MAPK (Ras-Raf-MEK-ERK) cascade reaction is an important signalling pathway, shown in Figure 1.3, that plays a vital role in cancer development[64].

The RAS/MAPK pathway has a critical role in normal development through regulation of cell growth, differentiation, and senescence[64]. A detailed description of the role of the MAPK pathway in cellular proliferation was provided by Zhang and Liu in 2002[65].

The signaling pathway starts with the interaction between a ligand and a epidermal growth factor receptor (EGFR), specifically human epidermal growth factor receptor 2 (HER2). Once active, HER2 links to a protein complex consisting of SOS and GRB2. SOS is a prominent guanine nucleotide exchange factor while GRB2 is an adaptor protein that plays a role in intracellular signalling. Once Sos-Grb2 has been activated, it in turn interacts with proteins from the RAS family, a family of small GTPases, and activates RAS. GTPases toggle between inactive and active conformations which acts as a switch for the signalling chains. This is done through the binding of guanosine phosphates. When guanosine diphosphate (GDP) is bound to the GTPase, they are in their inactive state. This is switched to an active state through based on their binding to the guanine nucleotides GDP or GTP. GTPases are in the "OFF" state when bound to GDP and are activated by guanine nucleotide exchange factors (GEFs), such as SOS. These change the confirmation of the GTPase, release GDP and bind GTP, which changes the structure of the protein and activates other targets downstream of the signalling process. A similar process occurs with the RAS family (*NRAS,KRAS,HRAS*). RAS in its original state is loaded with GDP. When it interacts with the Sos-Grb2 complex, it is shifted from its inactive GDP state to an active GTP state through the action of SOS as a GEF. Active RAS in negatively regulated through neurofibromin 1 (*NF1*), a GTPase activating protein which accelerates the rate of hydrolysis of GTP to GDP, thereby inactivating RAS. The active RAS protein then continues the MAPK pathway by binding to the Ras-binding domain of a family of proteins called the Rapidly Accelerated Fibrosarcoma(RAF) proteins, particularly *BRAF*. This activates *BRAF* which phosphorylates and activates two mitogen activated protein kinase kinases, MEK1 and MEK2 respectively. The activated MEK proteins phosphorylate and activate their targets, a pair of extra-cellular signal related kinases (ERK) called ERK1 and ERK2. ERK1 and ERK2 are responsible as regulators of several key cellular processes including cell proliferation, survival and metastasis. Improper regulation of this key pathway through disruption of its components plays a critical role in the oncogenesis of multiple types of cancer, including cutaneous melanoma[64]. The majority of sporadic melanoma cases are caused due to disruptions in key proteins of this pathway, particularly in *BRAF* and *NRAS*. The role of *BRAF* and *NRAS* in melanoma development are discussed in

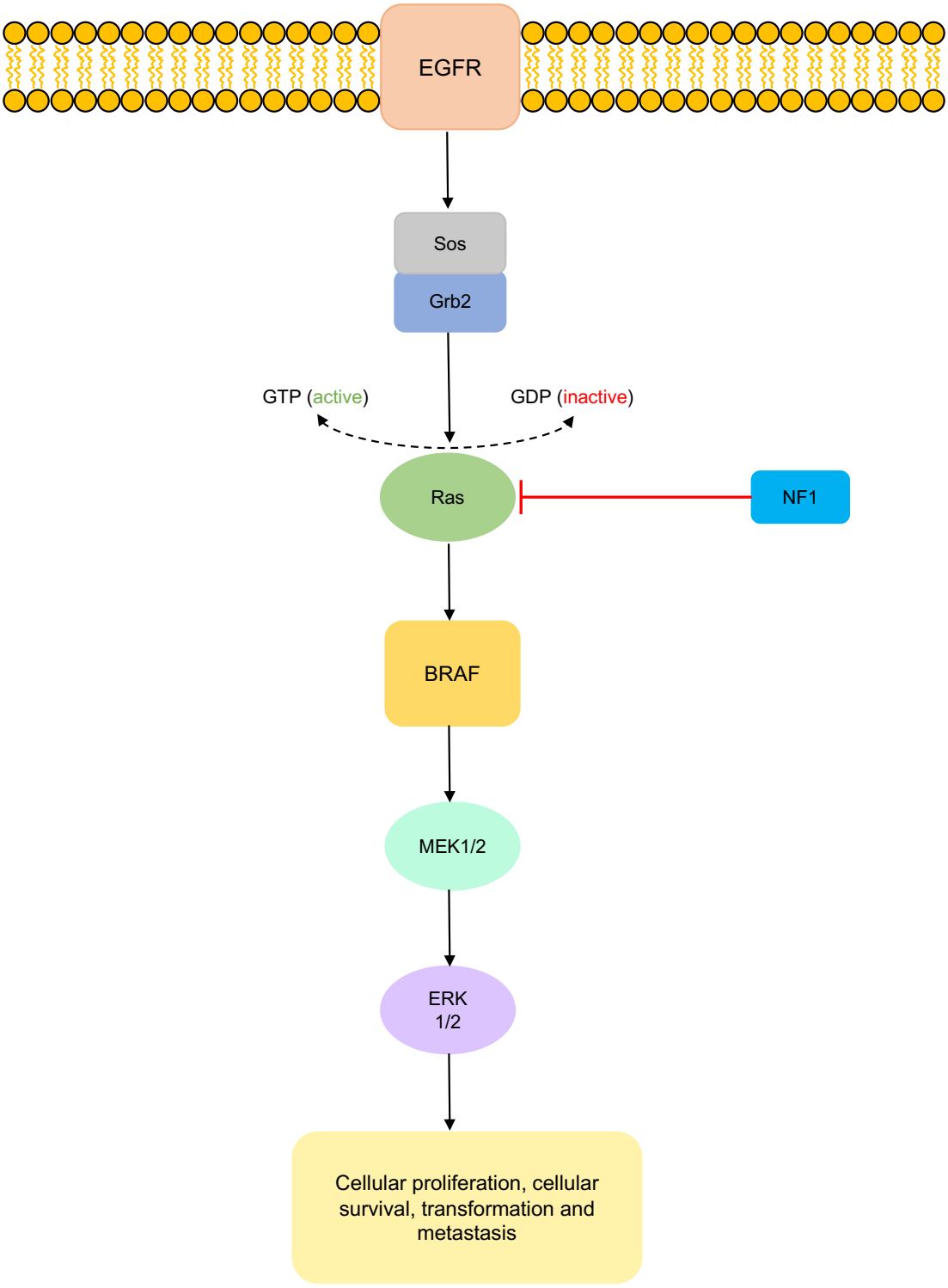

Figure 1.3: The MAPK (Ras-Raf-MEK-ERK) Pathway.

Sections 1.3.3.2 and 1.3.3.3 respectively. The importance of the MAPK pathway in cancer development has resulted in the development of multiple approaches for cancer therapy that target the key components of the pathway[66–70].

1.3.3.2 *BRAF* mutant melanoma

BRAF is a gene that encodes for a signal transduction protein kinase. It is part of the RAF family of genes and functions as an activator of the the MAPK (Ras-Raf-MEK-ERK) pathway. *BRAF* is phosphorylated and activated by RAS which in turn binds to and phosphorylates MEK1 and MEK2. Mutated *BRAF* is responsible for constitutive phosphorylation of MEK1 which continuously activates ERK and results in uncontrolled cellular growth. Activating *BRAF* mutations have been found in variety of human cancers such as cholangiocarcinoma (22%)[71], papillary thyroid cancer (69%)[72], colorectal cancer (12%)[73] and borderline ovarian cancer (28-48%)[74, 75]. However, cutaneous melanoma has the highest proportion of *BRAF* mutations with roughly 50% of cutaneous melanoma tumours estimated to have disruptive *BRAF* mutations. The vast majority of these mutations (over 90%) occur at the amino acid generated by the 600th codon which is a valine. V600E comprises of 90% of all mutations at codon 600 while V600K, V600R and V600D are other less frequent changes that occur at this position, although these frequencies change in different populations[19]. The V600E mutation results in constitutive activation of *BR*AF, leading to uncontrolled cellular growth[15]. This also prevents negative regulation and cellular senescence [76] and contributes to eventual metastasis[77]. *BRAF* mutations in melanoma occur more commonly in relatively younger patients with tumours in non-sun exposed areas. *BRAF* mutations rarely observed much in rarer types of melanoma such as mucosal and acral melanoma. *BRAF* mutation frequencies also differ between primary and metastatic melanomas with a range of 36 to 45% in primary melanomas and 42-55% in metastatic melanomas[78]. $BRAF^{V600E}$ mutations typically occur in younger patients(<55 years), occur at non-sun exposed areas of the body such as the trunk and the extremities, have moderate overall mutational burden and exist mutually exclusive with *NRAS* mutations[21, 79]. In the study conducted by TCGA, 75% of *BRAF* mutant patients were also identified as having *TERT* promoter mutations. The importance and relevance of both sporadic and germline *TERT* mutations, especially promoter mutations, in melanoma development is discussed in detail in Section 1.3.4.4. 91% of *BRAF* mutant samples harboured a UV-signature[20]. Multiple drugs functioning as kinase inhibitors have been designed specifically to target *BRAF* mutations such as vemurafenib, dabrafenib, and trametinib[80]. These drugs however have not proven to be completely effective; further treatment strategies are required for higher effectiveness.

1.3.3.3 *NRAS* mutant melanoma

A series of experiments on transforming retroviruses rats led to the identification of the Rat Sarcoma (RAS) family of genes. They were initially established as oncogenic viruses which cause formation of sarcomas in infected animals and had the potential to transform cells in culture. The Harvey murine sarcoma virus was identified in 1964 [81] with the discovery of the Kirsten murine sarcoma virus in following soon after in 1967[82]. Eventually, the origins of these viruses were traced back to the oncogenes that were responsible for them and these genes were named *HRAS* and *KRAS* respectively. In 1982, a third human *RAS* gene was identified in neuroblastoma-derived DNA and was therefore named as *NRAS*[83]. The three *RAS* family members, *NRAS*, *HRAS* and *KRAS*, have since been identified as being frequently mutated in human cancers, 20% of all tumours harbor activating mutations in one of their *RA*S genes. Mutations in *NRAS*, *KRAS*, or *HRAS* are known to be present in 20% (sometimes as high as 30%), 2%, and 1% respectively of all melanomas tested, making *NRAS* mutant melanoma the second highest subtype of cutaneous melanoma after *BRAF* mutant melanoma[84].

The most commonly reported hotspot mutation in *NRAS* which is observed in 80% of all *NRAS* mutations is a single nucleotide change that affects codon 61, changing a glutamine to a leucine (Q61L)[85]. Other common *NRAS* mutations include Q61R, Q61K, Q61H and G12R, G12D and G12A. *HRAS* mutations also largely affect codon 61 while *KRAS* mutations affected codon 12[86]. Mutations in position 61 are linked to reduced activity of the GTPase protein. This results in the RAS protein being stuck in its activated state and constantly activates Raf proteins. In the study by the Cancer Genome Atlas Network, 93.5% of RAS mutant melanomas were observed to exhibit UV signature[20]. This study also found copy number amplifications in *NRAS* that co-occurred along with *NRAS* mutations within the tumours. 72% of RAS subtype melanoma tumours were observed to have *TERT* promoter mutations. The typical patient with RAS mutant melanoma also tends to be older (>55 years of age) with higher chronic sun exposure than a patient with a *BRAF* mutant melanoma[79, 84]. *BRAF* mutations are more prevalent in benign nevi that are commonly present across the body with 80% of nevi carrying *BRAF* mutations compared to 14% carrying *NRAS* mutations[87]. Congenital nevi, however, have a much larger proportion of *NRAS* mutations; roughly 80% of congenital nevi have *NRAS* mutations[88]. *BRAF* mutations are rarely observed in congenital nevi. Previous attempts at targeted therapies for *NRAS* mutant melanomas by directly interacting with RAS have not been successful. A class of drugs were designed around farnesyltransferase inhibitors, aimed at preventing RAS modification[86]. However, these therapies exhibited severe off-target effects and were therefore not used. Phase 3 trials using binimetinib and dacarbazine were also applied on advanced *NRAS* mutant patients, binimetinib showing a

better response rate but with no difference in overall survival[84].

1.3.3.4 *NF1* mutant melanoma

Neurofibromin 1 (*NF1*) is a gene that was first discovered in the context of studying Neurofibromatosis type I, a genetic disorder that results in the formation of red spots on the eye (Lisch nodules) and in benign skin tumours (neurofibromas). *NF1* was first identified as being in the long arm of chromosome 17 through several concurrent studies in 1989. A precise mapping of *NF1* to a specific genetic locus of 17q11.2 soon followed in 1994. *NF1* encodes a large protein with multiple functional domains, including the GTPase activating protein related domain. This well known domain is responsible for the negative regulating of RAS, which is performed by switching the active state of RAS (Ras-GTP) to its inactive state (Ras-GDP) through hydrolysis. *NF1* functions as a tumour suppressor which is why loss of heterozygosity is also commonly observed in *NF1* mutant tumours through somatic mutations. Unlike *BRAF* and *NRAS*, *NF1* patients do not have a specific phenotypic characterisation, with wide variability being commonly observed. Large-scale next-generation sequencing studies have identified *NF1* as a commonly mutated driver gene in melanoma; 12 to 18% of all melanomas have *NF1* mutations. Although *NF1* mutations sometimes occur alongside *BRAF*/RAS mutations, they are more common in patients with no *BRAF* or *NRAS* mutations. While *NF1* mutations co-occur with hotspot RAS mutations, they do not co-occur with hotspot *BRAF* mutations. *NF1* mutations are also observed at a higher frequency in patients with desmoplastic melanoma, with 45-93% of patients having *NF1* mutations. *NF1* has been found to have somatic mutations in other types of cancers in addition to cutaneous melanoma. 40% of malignant peripheral nerve sheath tumours have *NF1* mutations, with lower frequencies of NF1 mutations also being observed in acute lymphoblastic leukaemia, glioblastoma, lung adenocarcinoma and pancreatic carcinoma. The study by the Cancer Genome Atlas Network focussing on the different subtypes of melanoma found that the average patient with *NF1* mutations was older(>55 years of age), 93% of whom exhibited UV signature and 83.3% of whom carried *TERT* promoter mutations. There was no significant observations with copy-number alterations[20]. While there have been no therapies that target *NF1* directly in melanoma due to its recent discovery, there are other inhibitors targeting the Ras-Raf-MEK-ERK pathway which is regulated by *NF1*. Melanomas with wild-type *BRAF*/*NRAS* are highly sensitive to the MEK inhibitor trametinib, with *NF1* protein expression also being sensitive to MEK inhibition, potentially implicating trametinib as a useful therapy for *NF1* patients[89]. It is also predicted that *NF1* mutations play a role in *BRAF* targeted therapy by increasing resistance to *BRAF* inhibitors. *BRAF* mutant murine tumours that are also mutated in *NF1* are identified

as being resistant to *BRAF* inhibitors. These tumours are however sensitive to combinatorial inhibition of MAPK/ERK and mTOR pathways[90]. Increased understanding of the role of *NF1* in cutaneous melanoma onset offers a potential novel target for new therapies which aim at inhibiting the Ras-Raf-MEK-ERK pathway.

1.3.3.5 Triple-Wild type melanoma

Cutaneous melanoma patients that do not have hotspot or any mutations in the tumour in either *BRAF*, *NRAS* or *NF1* are collectively called as triple-wild type melanoma patients[20]. Information on triple-wildtype melanoma is relatively sparse due to low frequency of cases and non-specific mechanism of melanoma development. The study by the Cancer Genome Atlas Network[20] identified several driver mutations in genes that weren't *BRAF*, *RAS* or *NF1* within the triple-wildtype subtype. As compared to the three main subtypes, triple-wildtype tumours also exhibited a much lower rate of UV signature with only 30% of tumours displaying the signature. Similarly, only 6.7% of triple wild type patients carried *TERT* promoter mutations. There was a significant increase in copy-number alterations and structural changes as compared to the other subtypes. Melanoma onset and progression in triple-wildtype patients may be driven by one of several other driver genes; therapeutic treatment of this subtype is therefore not as straightforward.

1.3.4 Germline familial melanoma genes and their clinical impact

1.3.4.1 The role of GWAS in melanoma research

The most common approach used in the investigation of large datasets of genomic data is to use a genome wide association study. These studies are used to help identify association of specific genetic SNPs or loci with a disease by comparing the genetic mutational burden across thousands of affected and unaffected individuals. While GWAS has been used as a tool quite significantly in a lot of other disorders such as type 2 diabetes[91], its use in melanoma research has been limited until recently. An initial GWAS performed in 2009 identified 3 markers involved in pigmentation and nevi formation including *TYR, ASIP* and *MC1R[92]. MTAP* and *PLA2G6* were also identified in this study and verified through follow up meta-analyses of multiple melanoma related GWAS studies in 2015[51] and 2020 [93]. *MTAP* is a gene that is frequently disrupted in cancers due to its proximity to *CDKN2A [94]* while compound mutations in *PLA2G6* lead to early-onset Parkinsons [95], a disease closely associated with melanoma[96]. The discovery of *PLA2G6* as a melanoma marker through GWAS showed that neurological diseases are linked to melanoma risk. Another GWAS study focussing on

loci related to tanning response to sun exposure helped identify another 14 novel loci related to this risk phenotype[97]. A GWAS study focussing on hair colour in Europeans determined 124 loci associated with hair colour, another low risk phenotype for melanoma[98]. Such GWAS studies have therefore expanded our knowledge of both low risk and high risk genetic markers in sporadic and familial melanoma. The number of risk markers for melanoma also increased from 3 in 2009 to 68 SNPS in 54 distinct loci in 2020 including genes in DNA damage repair pathways, telomerase pathways and pigmentation pathways, all of which have subsequently yielded novel candidate genes. Improvements in GWAS studies and inclusion of loci from additional risk phenotypes such as naevi count, hair colour and sun exposure will also improve the accuracy of polygenic risk score estimations in the future. Thus, while GWAS analyses might not directly translate into explaining the genetic origins of a disease, they play a significant role in providing insight into the mechanisms involved in disease onset.

1.3.4.2 Cyclin-dependent kinase inhibitor 2A *(CDKN2A)*

A major proportion of familial melanoma cases have a cause attributed to germline mutations in Cyclin-dependent kinase inhibitor 2A (*CDKN2A*). It was linked to melanoma for the first time through linkage analysis in 1994[25, 26]. It is estimated that it is responsible for up to 40% of familial melanoma cases with percentages varying from 20% to 57% in different parts of the world[27]. The *CDKN2A* gene lies on chromosome 9, and encodes two separate tumour suppressor protein products: p16INK4A and p14ARF (ARF = alternative reading frame). These two proteins arise from alternate splicing of *CDKN2A*, as shown in Figure 1.4. While both protein products share exons 2 and 3, they do not have a common amino acid sequence due to being encoded in an alternate reading frame.

The p16INK4a protein regulates the cell cycle by inhibiting the activity of *CDK4* and *CDK6*, two cyclin dependent kinases which are responsible for phosphorylation of the Retinoblastoma protein (RB)[100]. By controlling this activity, *CDKN2A* prevents the phosphorylation of RB and arrests the cell cycle at the G1-S phase. Mutant p16INK4A leads to early phosphorylation of RB which leads to improper progression into the S-phase of the cell cycle[101]. By contrast, p14ARF inhibits the binding of *HDM2* to the p53 tumour suppressor, thereby controlling the negative regulation of p53[102]. Both p53 and p16INK4A are known to play vital roles in cellular damage response and cellular senescence, both of which are integral pathways in cancer onset and progression. Germline in both p16INK4A([103]) and p14ARF([104]) have been shown to lead to melanoma; such mutations significantly increase the risk of melanoma development with a penetrance of up to 80% in the 8th decade of life. However, it is to be noted that *CDKN2A* mutations are extremely rare in population based melanoma cases. The

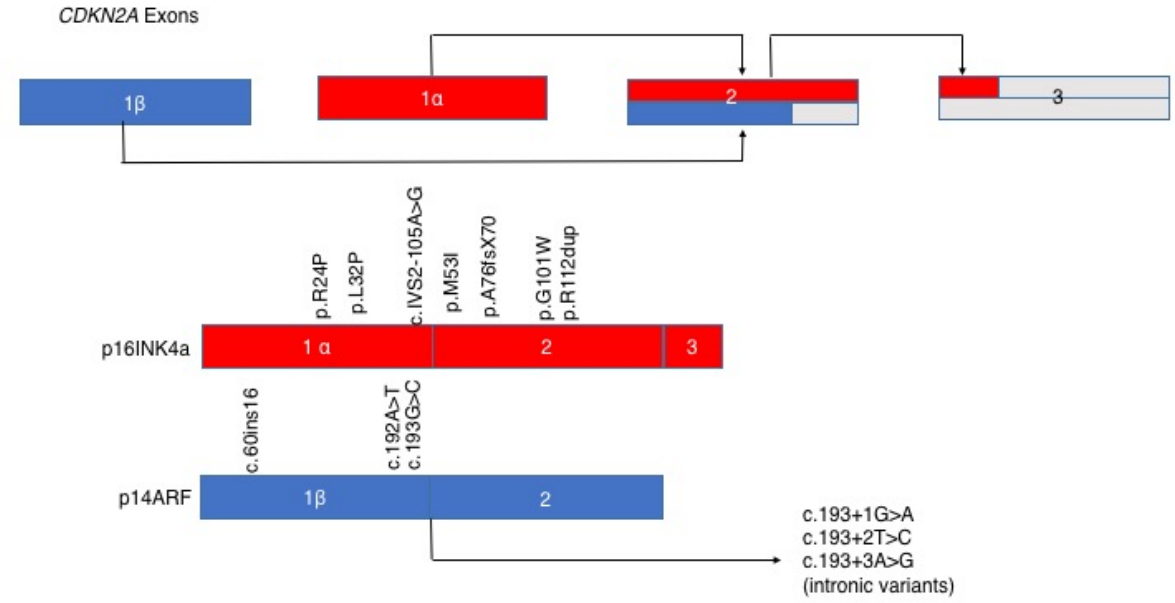

Figure 1.4: Creation of p16INK4A and p14ARF through alternate splicing of *CDKN2A* with the locations of founder mutations in melanoma in each protein also given. Adapted from [99].

collation of multiple studies comparing germline *CDKN2A* mutation carriers with non-carriers has also determined that *CDKN2A* mutation carriers have a lower median age of melanoma diagnosis[27, 105–108].

An increased incidence of pancreatic cancer has been observed in patients from familial melanoma pedigrees germline mutations in *CDKN2A* across several studies*[103, 109–113]*. In addition to giving rise to melanoma, somatic *CDKN2A* variants and disruption have also been observed in a several other types of cancer including oral pediatric lymphoblastic leukaemia, oral squamous cell carcinoma, head and neck squamous cell carcinoma, colon cancer and bladder cancer, indicating that *CDKN2A* has a role to play in general tumour formation and cancer development[114–122]. Germline *CDKN2A* mutations also resulted in increased risk for pancreatic, lung, head and neck cancers[123].

1.3.4.3 Cyclin dependent Kinase 4 *(CDK4)*

Once *CDKN2A* was discovered as a familial melanoma gene, efforts were focussed on discovering other potential familial genes by determining which other genes interact with it. This led to the discovery of the next familial melanoma gene, cyclin dependent Kinase 4 (*CDK4*). *CDK4* plays a key role in controlling cell cycle progression as it is responsible for the phos-

phorylation of RB. Mutations in *CDK4* which inhibit the activity of p16INK4A - thereby leading to early phosphorylation of RB - have also shown to be causative of melanoma. Due to the similar mechanisms involved in *CDK4* and *CDKN2A* mutations, they result in identical phenotypes, namely, early onset CMM, multiple primary melanomas and distinct nevi[124]. This complicates the process of distinguishing between *CDKN2A* and *CDK4* mutations as the cause of melanoma; families with melanoma that test negative for *CDKN2A* mutations should still be tested for mutations in *CDK4*. All *CDK4* driver mutations identified to date affect the 24th codon[124]. This supports the idea that the arginine amino acid generated by this codon normally is essential for the binding of the p16 tumour suppressor to *CDK4* which in turn prevents the phosphorylation of RB1. Alteration of this amino acid would therefore prevent *CDK4* inactivation. However, while *CDK4* mutations are necessary to understand the role of the p16 pathway in melanoma development, they are extremely rare.

1.3.4.4 Telomere maintenance pathway

Telomeres are terminal DNA structures at the ends of chromosomes responsible for genomic stability and integrity[125]. In humans, telomeres comprise of 9-15 kb double stranded repeats of a "TTAGGG" sequence which ends with a single stranded overhang called as the G-tail or G-overhang[126]. Cells that lack telomere length maintenance mechanisms progressively lose telomeric sequence with every round of cell division. This continues until the length of the protective telomere ends become critically short at which point they stop protecting the cells from DNA damage repair, leading to replicative senescence [126]. In this way, telomere length helps provide a mechanism for controlling the replicative lifespan of cells. Telomere replication, regulation and maintenance are primarily controlled through two protein complexes : the telomerase complex and the shelterin complex[127]. This section discusses the functions of these complexes, the roles of specific genes within these complexes in familial melanoma onset and previously identified germline mutations that implicate telomere dysregulation as a mechanism for melanoma development.

i) Telomerase and telomerase reverse transcriptase *(TERT)*

The maintenance of telomere length in germline cells and stem cells is controlled through the telomerase complex. Telomerase consists of two core subunits: telomerase reverse transcriptase (*TERT*) and telomerase RNA component (*TERC*)[127]. These components are responsible for telomerase extension through the addition of multiple "TTAGGG" repeats at the ends of the chromosomes using the G-overhang as the substrate[126]. *TERC* generates

the RNA component of telomerase while the addition of telomeric repeats is controlled by the highly regulated and conserved *TERT*. Additional accessory proteins including Dyskerin, GAR1, NHP2 and NOP10 are recruited to promote TERC accumulation[128]. TERC provides the template for DNA synthesis and is 451 bp long in *homo sapiens*. It also binds to TERT, which acts as the active site for catalysis and binds to the telomeric DNA[129]. The structure of the telomerase complex is shown in Figure 1.6.

Telomere length as a risk phenotype for melanoma was first reported in 2007 where increased telomere lengths were observed in circulating white blood cells which were associated with higher naevi count, another risk phenotype for melanoma[130]. In 2013, a novel promoter mutation 57 bp upstream of the translation start site which created an ETS binding motif was observed to be segregating with the disease in a German pedigree[131]. This was also shown to double the transcription rate of *TERT*. Members of the pedigree with this mutation had also developed multiple other types of cancer including ovarian cancer, renal cell carcinoma and bladder cancer, implying that *TERT* promoter mutations might play a role in general cancer development as opposed to being specific to melanoma. However, follow-up studies showed that germline *TERT* promoter mutations were sparse in the context of familial melanoma[132]. Additional recurrent somatic mutations affecting the *TERT* promoter at positions 124bp and 146 bp upstream of the translation start site have since been identified in multiple melanoma cell lines though whole genome sequencing [133]. These mutations, along with the germline promoter mutation, are shown in Figure 1.5. Somatic *TERT* promoter mutations have also been identified in an array of other cancers including bladder cancer[134], glioblastoma[135], thyroid cancer[136], mesothelioma[137], hepatocellular carcinoma[138] and squamous cell carcinoma[139]. The confirmation of telomere length as a risk phenotype for melanoma after initial studies focussing on precursor phenotypes indicate that this is a cogent approach in the discovery of novel cancer genes and pathways.

ii) Protection of telomeres protein 1 *(POT1)* and the shelterin complex

The shelterin complex are a set of six proteins that protect telomeres from degradation. These proteins regulate the interactions of the telomeres with the telomerase complex, help protect telomeres from the DNA damage repair pathway and maintain genomic integrity[141]. The constituent components of the shelterin complex include the following six core proteins *:* adrenocortical dysplasia protein homolog *(ACD)*, protection of telomeres 1 *(POT1), TERF2*-interacting protein 1 *(TERF2IP,* also called *Rap1),* telomeric repeat-binding factors 1 and 2 *(TERF1* and *TERF2)* and *TERF1*-interacting protein 2 *(TINF2)* [127].

TERF1 and *TERF2* are responsible for the production of double-stranded DNA binding

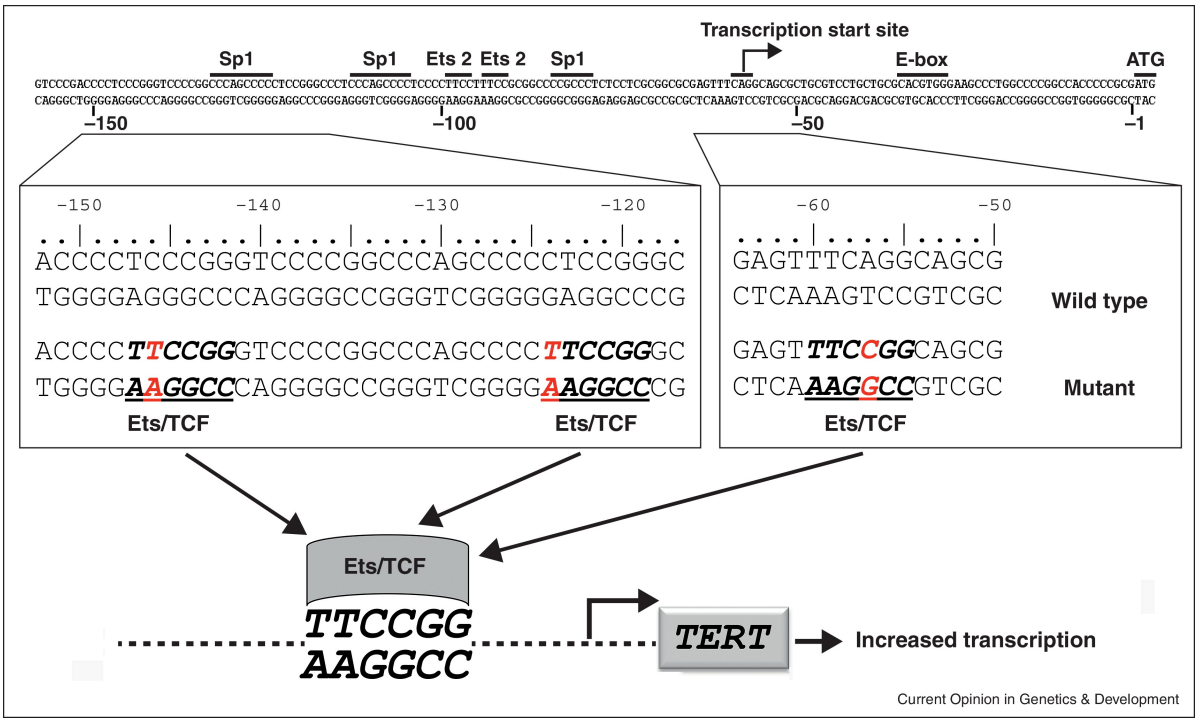

Figure 1.5: Previously identified *TERT* promoter mutations in sporadic and familial melanoma. Reused with permission from [140].

proteins that identify and bind to the telomeric repeats. *POT1* is the most conserved component of the shelterin complex which creates a protein that binds to single stranded telomeric DNA; *ACD* helps in the recruitment of POT1 to the telomere by binding to POT1 and creating a sub-complex[127, 141]. *TERF2IP* does not directly interact with telomeric repeats and instead interacts with *TERF2*[142]. The shelterin complex is bound together as a single entity through *TINF2*. *TINF2* constructs a protein which binds to TERF1, TERF2 and the ACD/POT1 complex[127, 141]. The shelterin complex helps in the generation of a protective structure at the end of the telomere called as telomeric loop or T-loop. The presence of this loop provides a distinct protective cap to telomeres which distinguishes telomeres from double-strand breaks and protects them from the DNA damage repair pathway[126]. The arrangement of the different proteins within the shelterin complex and their interactions with each other resulting in the formation of the T-loop are shown in Figure 1.6.

Large-scale exome sequencing of members belonging to 105 pedigrees from Australia, The UK and The Netherlands was performed for identification of novel germline variants[143]. Three missense mutations (Y89C, Q94E, R273L) affecting the OB domain of *POT1* and a splice set variant in *POT1* were identified from this dataset. These mutations resulted in longer telomeres, which predisposed the individuals with the mutation to developing cuta-

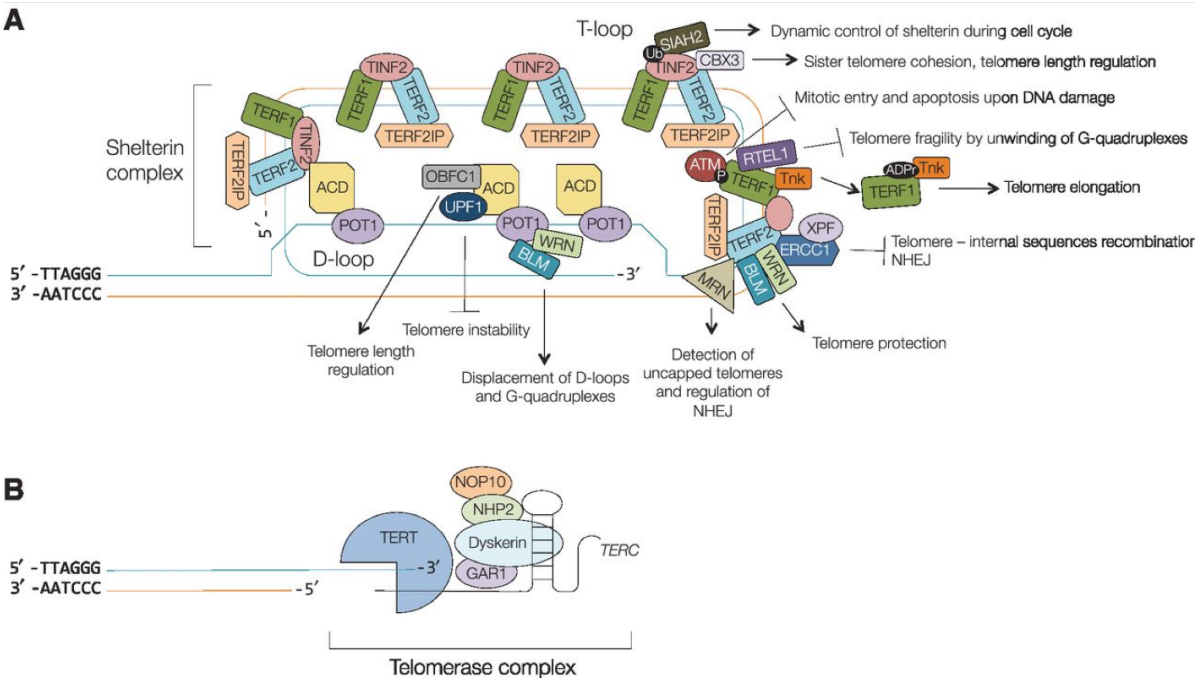

Figure 1.6: a) The structure of the shelterin complex showing the interactions between the different components (including the association of *ACD* with *POT1* and *TERF2IP* with *TERF2*), alongside their function in telomerase maintenance and T-loop formation.b) Structure of the telomerase complex showing the interaction between *TERC*, *TERT* and the other proteins necessary for the addition of telomeric repeats. Reproduced with permission from [127].

neous melanoma[143]. Concurrently, a founder mutation and additional rare variants in *POT1* were also identified from another whole-exome sequencing study[144]. With the discovery of inactivating mutations in *POT1*, the remaining components of the shelterin complex were also investigated for their possible role in the melanoma progression. Additional exome and whole genome sequencing of 510 affected families resulted in the discovery of 6 families with mutations in *ACD* and 4 families with mutations in *TERF2IP*[145]. *ACD* binds with *POT1* and interacts with the telomerase complex. It was observed that when the binding domains of *ACD* and *POT1* are mutated, they fail to form a functioning shelterin complex, which leads to increased telomere length due to an active telomerase complex. The mutations in *TERF2IP* were found to disrupt its binding capacity with *TERF2* which in turn affects the repair of the double strand break at the telomere.

Thus, the presence of mutations in genes of the shelterin complex which are associated with familial melanoma strengthen the relevance of telomere dysregulation as a mechanism in melanoma development.

1.3.4.5 BRCA1-associated protein-1 *(BAP1)*

A heterodimer of breast cancer 1(*BRCA1*) and *BRCA1*-associated RING domain (*BARD1*), which has E3 ubiquitin ligase activity, controls the DNA damage repair pathway. *BRCA1*-associated protein-1 (*BAP1*) acts as a deubiquitination enzyme and helps in deubiquitinating *BARD1* and regulating the E3 ligase activity of this complex*[146]*. Inhibition of *BAP1* leads to impaired DNA damage repair process and results in S-phase retardation. *BAP1* also interacts with the Yin Yang1 (YY1) transcription factor to control the transcription of genes involved in cellular proliferation*[146]*.

Inactivating somatic mutations were first identified in *BAP1* in metastasizing uveal melanoma tumours[147]. Additional studies have associated germline *BAP1* mutations with malignant mesothelioma and with distinct morphological neoplasms related to melanocytic tumours leading to cutaneous melanoma and renal cell carcinomas[148–150]. While missense mutations do occur, they were comparatively rare compared to nonsense germline *BAP1* mutations. An extensive study of all known nonsense and missense germline *BAP1* mutations identified 104 unique nonsense variants and 36 unique missense variants from 181 families[151]. In 2013, 15% of *BAP1* mutation carriers developed cutaneous melanoma*[146]* but the larger study in 2018 identified a 24% occurrence of cutaneous melanoma in probands and a 12% occurrence of cutaneous melanoma in non-proband variant carriers[151]. This implies that *BAP1* is a medium-penetrance risk gene for cutaneous melanoma. A list of all truncating germline variants in *BAP1* are shown in Figure 1.7.

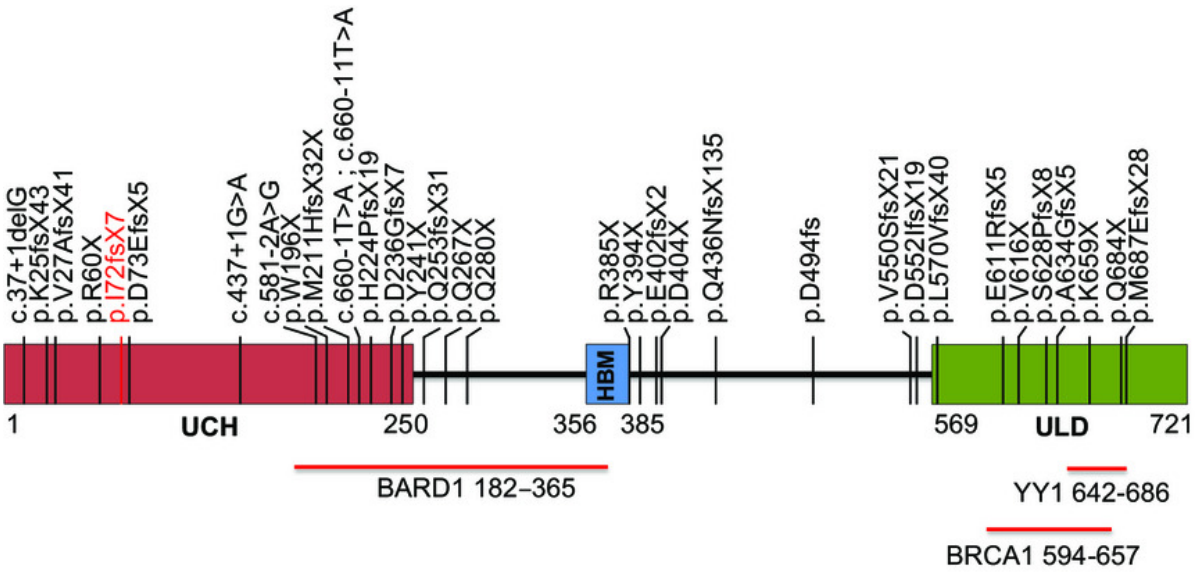

Figure 1.7: Germline truncating *BAP1* mutations along with the protein domains of *BAP1*. The binding sites of *BARD1*, *BRCA1* and *YY1* are shown in red. Reused with permission from [99].

A large scale population study was performed in 2017 by O'Shea et al. to identify the frequency of germline *BAP1* mutations in sporadic melanoma[152]. 1,977 melanoma cases and 754 controls were sequenced for this study with only 30 mutations identified in *BAP1*, shown in Figure 1.8. Only 2 of these were truncating mutations, indicating that germline *BAP1* mutations are extremely rare in sporadic melanoma and strengthening the claim that *BAP1* is a medium-penetrance risk gene for cutaneous melanoma.

1.4 Classification of melanoma

1.4.1 Melanoma subtypes

Wallace Clark not only defined a staging system for the progression of a tumour from a benign nevus malignancy, but had also previously established three specific malignant melanoma subtypes based on the type of tumour : superficial spreading melanoma (SSM), nodular melanoma (NM) and lentigo maligna melanoma (LM)[9]. A fourth subtype of melanoma called acral lentiginous melanoma (ALM) was later identified[154]. Additional, rare melanoma subtypes were identified following this. The characteristics of these subtypes were described by Anand Rotte and Madhuri Bhandaru[155]; these characteristics are summarised here:

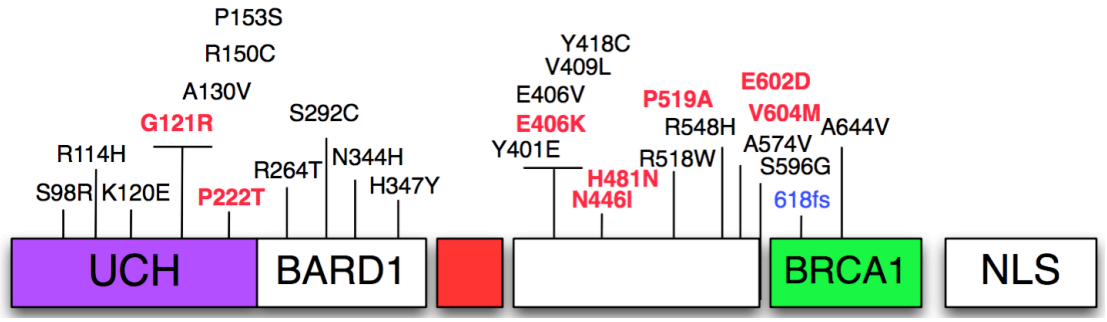

Figure 1.8: List of germline mutations in *BAP1* observed in sporadic melanoma. The nonsense mutations are highlighted in red. Reproduced with permission from [152].

1.4.1.1 Superficial spreading melanoma

This is the most common type of melanoma, accounting for roughly 60% of all melanoma cases. Patients with SSM are usually less than 60 years old (with a median age of 55) and have a high rate of $BRAF^{V600E}$ mutations. SSM tumours are not caused due to chronic sun-damage and are usually located in non-sun exposed areas of the body. The tumour appears as a flat, discoloured region which eventually enlarges radially. When it invades the dermal region of the skin, it forms an elevated lump in the region.

1.4.1.2 Nodular melanoma

This is the second most common subtype of melanoma and is observed in 15-30% of all melanoma cases. It is usually observed in the sun-exposed areas (head and neck) of older patients (>60 years of age) with high sun-exposure. NM is the quickest growing subtype in terms of tumour depth and is linked to poor prognosis due to late detection. NM tumours are rigid, symmetric and do not show a lot of colour variation. They also do not tend to change colour on growth, which sometimes contributes to its late detection. On occasions, these tumours ulcerate and potentially start bleeding.

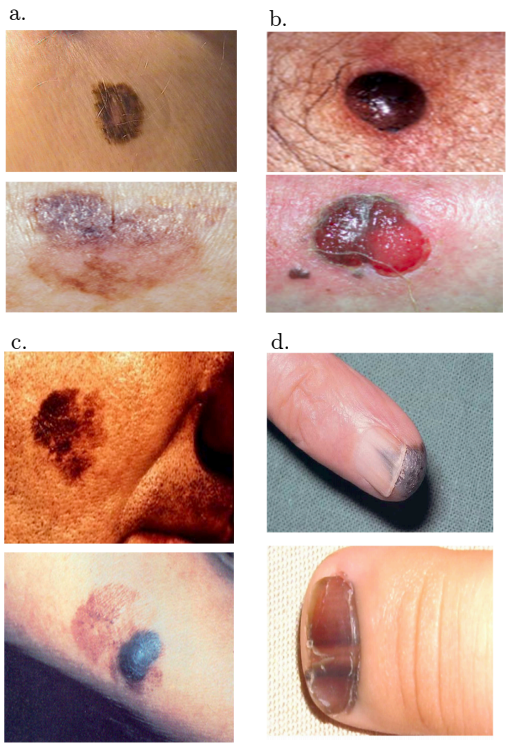

Figure 1.9: The different types of tumours observed in common melanoma subtypes is depicted here. These subtypes are a) Superficial spreading melanoma b) Nodular melanoma c) Lentigo maligna melanoma d) Acral lentiginous melanoma. Images obtained from [153]. Images downloaded and reused under the under Creative Commons Attribution (CC BY) license (http://creativecommons.org/ licenses/by/4.0/).

1.4.1.3 Lentigo maligna melanoma

This subtype refers to the development of melanomas in sun-damaged skin caused by chronic sun exposure. As a result, these are commonly found on parts of the body with high sun-exposure and in older people (>60 years of age). LM accounts for 4-10% of all melanoma cases. The tumour initially presents as a black coloured region of discolouration that is flat and has rhomboidal structures. tumour growth is lentiginous in the epidermal region.

1.4.1.4 Acral lentiginous melanoma

This subtype is responsible for 2-3% of all melanoma cases but is the most common type of melanoma in non-Caucasian populations., ranging from 9% in Hispanic Whites to 36% in people of African descent. LM, similar to ALM, is also diagnosed over the age of 60 on average. It is observed in the extremities of the limbs such as the palms, soles or under the nails. The development of ALM in such irregular sites leads to a later diagnosis as compared

to the other subtypes and therefore has a worse survival rate as well. These lesions have specific ridge like patterns and occasionally exhibit symmetric arrangement of globules under dermoscopic observation.

1.4.1.5 Rare melanoma subtypes

In additional to these primary subtypes, there are several extremely rare subtypes of melanoma that are seen in around 1% of all melanoma cases. Some of these subtypes are:

- Desmoplastic melanoma: Rare subtype of melanoma more frequently seen in men. Presents as scar like nodules and lacks prominent clinical features, preventing early detection.
- Nevoid melanoma: Rare subtype of nodular melanoma. tumour presents as a common or Spitz nevus.
- Verrucous melanoma: Rare subtype of cutaneous melanoma often mistaken for seborrhoeic keratosis . More frequent in women. tumour presents as a wart covered lesion and usually develop on the extremities.
- Mucosal melanoma: Comprise of less than 1% of all melanoma cases; present in mucosal surfaces of the body such as nasal passages, sinuses, vagina, bowel, urethra and anus.
- Giant congenital melanocytic nevus: Considered as a precursor for malignant melanoma, consist of melanocytic lesions present at birth that grow in size over time and may eventually progress to melanoma.

Such diversity in melanoma characterisation indicate that cancer progression and development through melanocytes can take one of several paths depending on several factors including gender, age, ethnicity and sun exposure. Such a complex process demands a clear and well-defined system of stratification to categorize patients into distinct groups which would enable better disease treatment and improve survival.

1.4.2 Cancer staging systems

1.4.2.1 A history of staging systems for cancer

Initial diagnoses of melanoma up to the 1950's were made by identifying distinct attributes on the skin, at which point they were often at an advanced stage with a poor prognosis. To

improve patient survival and treatments, different staging systems have been developed over time to characterise and stratify patients based on the nature of their tumour and cancer progression. These staging systems have also improved along with the knowledge of the disease and have become more accurate over time.

The first major observation related to prognosis and tumour types was made in 1953 by Allen and Spitz where they had reported that melanoma tumours with higher depth had worse survival[156]. The first attempt at developing a stage system for melanoma based on the primary melanomas was done by Petersen in 1962 where patients were classified based on what stage of the dermis had been invaded by the tumour[157]. Further studies confirmed that there were three distinct groups of patients based on primary tumour, regional lymph metastasis and distant metastasis status of the patients. This led to another three stage system proposed in 1964 by Mcneer and Dasgupta where patients were classified into one of three categories: primary tumour with no metastasis, metastasis confined to regional lymph nodes and multiple, distant metastases[158].

An additional staging system based on tumour depth was established by Wallace Clark in 1969 which led to the Clark levels staging of melanoma invasion[9], which was also independently verified in 1970[159]. Based on this system, cutaneous melanoma was subdivided into five categories:

1. Level I : tumours are restricted to upper membrane of the epidermis, also called as in-situ melanoma.

2. Level II : tumours have extended from basement membrane to the papillary dermis but have not extended into the reticular dermis.

3. Level III : tumours have bridged the interface between the papillary and reticular dermis.

4. Level IV : tumours have completely extended into the reticular dermis.

5. Level V: tumours have invaded the subcutaneous tissue.

As mentioned in Section 1.2, Alexander Breslow came up with Breslow's thickness as a measurement parameter for tumour thickness at the same time[10]. This was found to be more effective as a prognostic measure and has since been incorporated into almost all future major melanoma staging systems. The late 1970's saw melanoma staging systems based on the primary tumours, lymph nodes and metastasis status which incorporated both the Clarks levels and Breslow thickness. These systems were created by the American Joint Committee on Cancer and the Union for International Cancer Control; these were eventually merged into a single

TNM staging system. This system has been regularly updated with multiple editions across the years with several subcategories and subgroups included based on analysis of thousands of patients to better classify different patients and is the current standard.

1.5 Genetic testing and therapies for familial melanoma

Unlike somatic mutations, germline variations affect and are present in all cells of the body. By testing for specific predominant germline mutations, we can estimate a risk for predisposition to familial melanoma. This process is referred to as genetic testing or gene panel testing. The risk of predisposition is higher if there are multiple family members affected with familial melanoma or if an individual in the family has had multiple primary melanomas as this suggests an underlying germline genetic cause over a sporadic mutation. The primary genes that were tested for familial melanoma were *CDKN2A*(p16) and *CDK4*. Recent improvements to multi-gene panel testing include the addition of several other genes of relevance to cancer development. The current NHS cancer gene panel includes 156 genes; the familial m elanoma gene panel amongst these genes include the aforementioned *CDKN2A* and *CDK4* but now also include other familial melanoma driver genes such as *BAP1, BRCA2, POLE, POT1* and *TERT*. Genetic testing is also free in the NHS if there are more than 2 related individuals with melanoma, if a patient has multiple primary melanoms or if melanoma and pancreatic cancer exist in the same pedigree. The addition of knowledge from sequencing studies such as this project regarding the variants in other medium to low penetrance genes in the development of familial melanoma will further refine the quality of genetic testing for familial melanoma in the future. The quality and access to genetic testing varies significantly from place to place; additional education is required for dermatologists and oncologists to refer patients to a clinical geneticist and genetic testing when required.

The primary method of treatment for stage 1 and stage 2 melanomas is surgical excision, where the melanoma is removed along with a small area of the skin around it. Patients are continued to be monitored for a few years to ensure that the melanoma does not return. Stage 3 and 4 melanomas are usually treated with a combination of immunotherapy and targeted gene therapy. Targeted gene therapy involves the use of drugs including dabrafenib, vemurafenib and trametinib in the treatment of specific types of melanoma. Particularly, dabrafenib and vemurafenib are used for the treatment of cases with *BRAFV600E* mutations in sporadic cases[80], while trametinib, which is a MEK inhibitor is used for *NRAS* and *NF1* affected melanoma cases[89].

Immunotherapy involves the use of drugs that target specific components of the immune

complex such as cytotoxic T lymphocyte associated antigen 4 (CTLA4*)* and programmed death ligand 1 (PDL1). Cancer cells produce antigens which are detected by the immune system and help identify them. Components of the immune system called as cytotoxic T lymphocytes (CTLs) in the lymph node target these cancer cells based on the antigens and kill them. However, additional inhibitory signals are sometimes created by the dendritic cells that act as a checkpoint and are aimed at preventing the immune system from attacking the body. These inhibitory signals are detected by CTLA4 which are present on CTLs and this results in the cytotoxic reaction of the CTLs being deactivated, resulting in the growth of the cancer cells. Drugs such as ipilimumab can bind and block the function of CTLA4, thus allowing CTLs to function as normal and kill cancer cells[160]. Another similar mechanism is the role of T cells in targeting cancer cells. The surface of T cells contain a protein called programmed cell death 1 (PD1). In order to prevent the T cells from targeting normal cells, a protective protein called PDL1 binds to PD1 to inhibit T cells. However, some cancer cells can also produce PDL1 which prevents the T cells from killing these cancer cells. Nivolumab is a drug that works on this mechanism and obstructs PDL1 from binding to PD1, thereby activating T cells[161].

The presence of germline variants usually has no bearing on the treatment of the disease as they have access to either targeted gene therapy or immunotherapy depending on the presence or absence of *BRAF* mutations. However, patients with germline mutations do tend to have better survival compared to sporadic melanoma cases due to the less aggressive nature of the tumours and increased surveillance of the disease.

1.6 Proposed approaches of sequence analysis undertaken for the identification of novel genes

A variety of approaches ranging from linkage analysis and positional cloning to whole-exome sequencing have helped identify several high penetrance familial melanoma genes including *CDKN2A*, *CDK4, BAP1*, *POT1* and *TERT*. However, germline variants in these genes are jointly responsible for up to 40% of all known familial melanoma cases, leaving the remaining with an unexplained genetic cause. Such cases can be explained with one of three alternatives:

- They have a mutation in a yet-to-be-discovered high penetrance gene.
- They have a mutation in a high penetrance gene but in a region that is still unexplored such as transcription factor binding sites or the 5’ UTR.

- They have a high burden of mutations in several low penetrance genes which cumulatively lead to melanoma development.
- They have an increased risk of cancer due to a combination of genetic and epigenetic factors.

With respect to the pathways involved in melanoma development, most of the genes identified so far are involved in cell cycle regulation (*CDKN2A*, *CDK4* and *BAP1*) or genomic stability through telomere maintenance(*TERT*, *POT1*). Disruption of these processes can lead to multiple types of cancer, as is evident from the different disorders these driver genes cause. The set of families with an unknown cause indicates that there could be either be other potential driver genes within these pathways or genes in uncharted novel pathways that play a role in the onset of familial melanoma. Additionally, while epigentic factors might play a role in oncogenesis, the high number of melanomas in each pedigree indicates towards a genetic mutation because the cause of disease.

This PhD project was directed at exploring these possibilities and shed some light on answering these questions. The project comprises of sequencing and analysing a large cohort of familial melanoma pedigrees comprising of both whole genome and exome sequences. This approach was chosen for the following advantages:

- The falling costs of sequencing (whole genome sequencing in particular) has allowed for the sequencing of hundreds of samples, increasing the statistical power of the project in the ability to observe low-frequency mutations.
- The availability of whole genome sequences enables the exploration of the non-coding region of the genome for structural variants, regulatory region variants and promoter mutations, all of which are relatively unexplored in the context of familial melanoma.
- The presence of publicly available datasets such as ExAC and gnomAD which provide a highly accurate estimate of allele frequencies in normal populations which is vital for the identification of low-frequency, high-penetrant mutations.
- Sequencing the whole genome/exome allows us to neutrally examine the entire human gene set for novel driver genes and variants as opposed to focussing on a specific subset of candidate genes which could be potentially biased.

The chapters in this thesis have been organized into different components based on the type of analysis performed on the dataset, consisting of 4 distinct cohorts. This is shown in Figure 1.10.

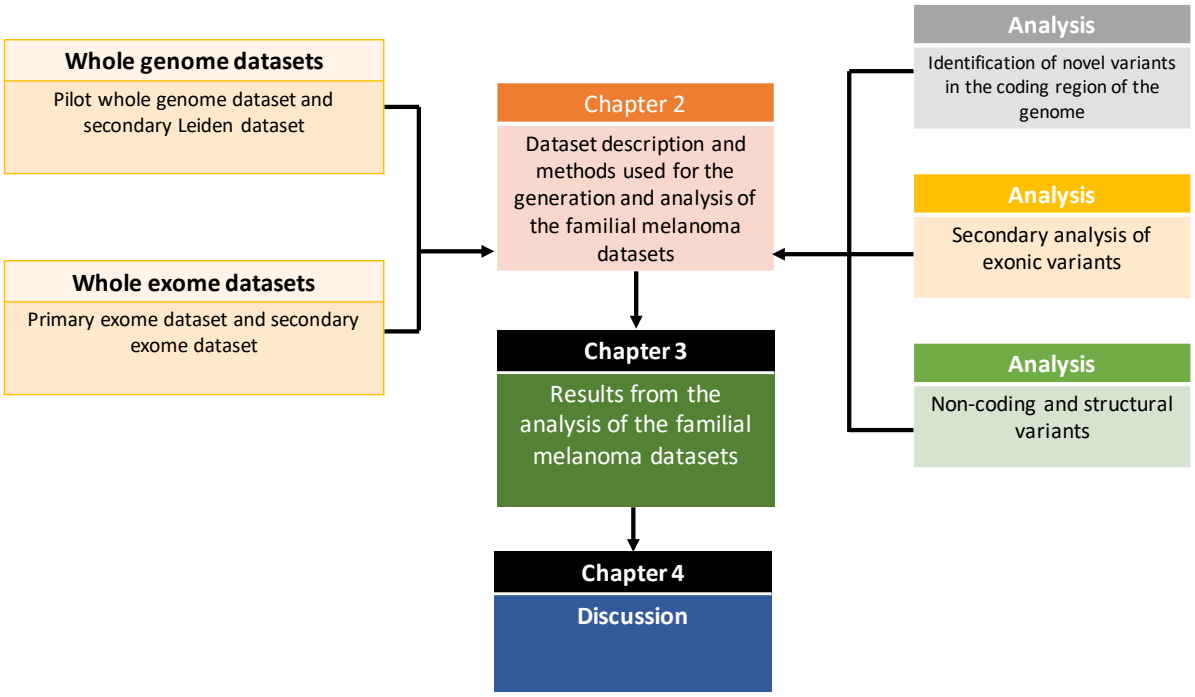

Figure 1.10: Outline of PhD project.

Two of these cohorts consist of whole genome sequences while the other two are whole-exome datasets and the samples were obtained from 8 different locations/institutions across the world. The composition and criteria for sample selection in each of these datasets are explained in Chapter 2. Chapter 2 also contains information on the background and methods designed for the different analyses performed on the dataset. This includes an association analysis and a joint association-linkage analysis of coding region variants aimed at identifying novel high penetrance melanoma susceptibility genes. This is followed by the exploration of the variants in the dataset in known melanoma predisposition genes and additional secondary analysis of exonic variants that complement the association and linkage analysis. The final approach that is discussed comprises the search for novel mutations in the non-coding region, particularly with structural variants and mutations in known transcription factor binding sites. The complete workflows for all of these different processes are described in detail in Chapter 2. The results from these different analysis are included altogether in Chapter 3. Finally, in Chapter 4, the relevance of the results from each of the preceding chapters as well as the future directions of this project are discussed.

1.7 Overarching aims of the project

The primary goal of this project is to determine novel variants that predispose individuals carrying these variants to the development of familial melanoma. This goal incorporates several key distinct aims which are individually listed here.

- To obtain the samples of familial melanoma patients from multiple locations/sources and to analyse these samples - through exome or whole genome sequencing.
- To incorporate all the individual datasets sequenced through different methods into a single, consistent dataset.
- To perform variant calling uniformly across the dataset and to annotate each mutation with their predicted consequences on protein function.
- To perform preliminary analyses on the dataset to eliminate potential pre-existing biases related to an increased burden of common risk factors and population stratification.
- To identify rare, deleterious variants in data from cases and controls by filtering on several criteria.
- To utilise a rare variant association analysis for the identification of genes with a higher mutation burden in cases compared to controls.
- To design and execute a joint approach combining association analysis and linkage analysis that employs both variant data from the sequencing and the relatedness data from the pedigrees can be utilised in determining novel candidates for familial melanoma development.
- To establish methods that can determine variants related to cancer development which cannot be identified through a rare-variant association and linkage analysis.
- To determine which of these variants have high segregation within our cases and to account for the presence of potential phenocopies within the pedigrees.
- To identify variants in known melanoma predisposition genes by annotating their clinical significance using ClinVar and to explore potentially pathogenic variants associated with cancer.
- To establish the location of transcription factor binding motifs across the genome.

- To ascertain rare non-coding variants that lie within transcription factor binding motifs.
- To determine a suitable control dataset and to identify genes with increased burden of non-coding variants within transcription factor binding motifs in cases compared to controls and to discern rare variants within the non-coding region of the genome.
- To establish a workflow for the identification of structural variants within the cases.
- To identify novel structural variants disrupting known cancer genes.