An Investigation into the Mechanisms of Angiogenesis and Breast Cancer Metastasis

Ivonne Cesarina Olivares García

MD, MSc

University of Wolverhampton

Faculty of Science and Engineering

PhD Thesis   2023

An Investigation into the Mechanisms of Angiogenesis and Breast Cancer Metastasis

Ivonne Cesarina Olivares García

MD, MSc

A thesis submitted in partial fulfilment of the requirements of the University of Wolverhampton for the degree of Doctor of Philosophy (PhD)

June 2023

This work and any part thereof has not previously been presented in any form to the university or to any other body whether for assessment, publication or for any other purpose (unless otherwise indicated). Save for any express acknowledgments, references and/or bibliographies cited in the work, I can confirm that the intellectual content of this work is the result of my own efforts, not those of any other person.

The right of Ivonne Olivares to be identified as the author of this work is asserted in accordance with ss. 77 and 78 of the Copyright, Design and Patent Act 1988. At this date, the author owns the copyright.

Signature:

Date: 19/06/23

Acknowledgements

Abstract

Breast cancer survival rates have increased over the years due to early detection and therapeutic efficacy. However, after many years of what appears to be disease-free health, cancer can return in the form of a secondary metastatic tumour. Formation of new blood vessels is a crucial stage in the progression of primary tumours to metastatic tumours. In many cases primary breast tumours that metastasise to the brain occur. Brain tumours are heterogenous malignancies with a low survival rate. Tumours often development therapy resistance by secreting different pro-angiogenic growth factors that allow them to overcome the effect of anti-angiogenic drugs. Thus, characterising the mechanisms that promote tumour angiogenesis may help to stop the development of tumour metastasis. Here it was determined that ATF2 is a downstream molecule activated (phosphorylated) by major pro-angiogenic factors and that its suppression in HUVECs resulted in increased upregulation of Notch signalling pathway (major regulator of angiogenesis) ligand DLL1 and DLL4.

Additionally, we investigated genomic changes that may be involved in the development and progression of breast to brain metastasis (BBM). Whole exome sequencing (WES) of 26 breast to brain metastases was carried out. Bioinformatic analysis of WES data identified recurrent genomic alterations in several genes that may be associated with BBM development including an ARFEFG protein family member gene, BIG3. Functional analysis of BIG3 showed that this gene is involved in the regulation of neurotransmitter receptors subunits in the BIG3-knockout MCF-7 breast cancer cell line. Neurotransmitter regulation has been shown as one mechanism through which brain tumours integrate into the neuronal signalling network to promote colonisation of the brain.

Results presented here show that investigating the possible mechanisms behind tumour angiogenesis, and the molecular changes that may be involved in metastatic tumour development and brain colonisation offers a better understanding of tumour mechanisms for growth and survival which can offer an opportunity for the development of new therapeutic targets.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Ad-ATF2_AA - Dominant-Negative Version of ATF2

Ad-GFP - recombinant adenovirus containing the green fluorescent protein gene

ARF - ADP-Ribosylation Factor

ATF2 - Activating Transcription Factor-2

ATP - Adenosine Triphosphate

BBB - Blood Brain Barrier

BBM - Breast to Brain Metastasis

bFGF - Basic Fibroblast Growth Factor

BRCA2 -Breast Cancer Susceptibility Gene 2

CaMK - Ca^{2+}/Calmodulin-Dependent Protein Kinase

CCLE - Cancer Cell Line Encyclopedia

CDD - Conserved Domain Database

CNS - Central Nervous System

CoSMIC - Catalog of Somatic Mutations in Cancer

COX2 - Cyclooxygenase-2

DLL1 – Delta-like Protein 1

DLL4 - Delta-like Protein 4

DMSO - Dimethyl sulfoxide

DNA - Deoxyribonucleic Acid

EBNA1 - Epstein Barr Nuclear Antigen 1

EGF - Epidermal Growth Factor

EGFR - Epidermal Growth Factor Receptor

EMT - Epithelial Mesenchymal Transition

ER - Estrogen Receptor

ExAC - Exome Aggregation Consortium

FBS - Fetal Bovine Serum

GABAR - Gamma Aminobutyric Acid Receptors

GFP - Green Fluorescent Protein

HDFa - Human Dermal Fibroblasts

HGF - Hepacyte Growth Factor

HIF - Hypoxia Inducible Factor

HPRT1 - Hypoxanthine Phosphoribosyltransferase 1

HRECs - Human Retinal Microvascular Endothelial Cells

HUVEC - Human Umbilical Vein Endothelial Cells

MAF - Minor Allele Frequency

MOI - Multiplicity of Infection

NCBI - National Center for Biotechnology Information

NGS - Next Generation Sequencing

NMF - Non-Negative Matrix Factor

NR - Not Reported on ExAC database

PCR - Polymerase Chain Reaction

PPAR - Peroxisome Proliferator-Activated Receptor

RNA - Ribonucleic Acid

RT - Reverse Transcription

SIFT - Sorting Intolerant From Tolerant software

Si-NT - Si-Non-Target

SNAI2 - Snail Family Transcriptional Repressor 2

SNV - Single Nucleotide Variant

TGCA - The Cancer Genome Atlas

TNBC - Triple Negative Breast Cancer

VEGF - Vascular Endothelial Growth Factor

WES - Whole Exome Sequencing

WT - Wild Type

Chapter 1 Introduction

1.1 Cancer

1.1.1 Definition, incidence and significance, treatments

Cancer is a term that covers a multitude of disease indications (Hanahan, 2022). Worldwide, cancer is the primary cause of death after heart disease (Mattiuzzi and Lippi, 2019). In 2020 there were 19.3 million novel cancer cases and around 10 million cancer deaths globally (Ferlay et al., 2021). In the United Kingdom, 1 in 2 people are predicted to get cancer in their lifetime (Ahmad, Ormiston-Smith and Sasieni, 2015). For more cancer related statistics refer to Figure 1.1.

Cancer development comprises a series of biological capabilities termed the hallmarks of cancer that result in the creation of the right environmental conditions for tumour formation and malignant progression (Villa et al., 2019). Currently there are six original hallmarks, two enabling capabilities, and two emerging characteristics (See Figure 1.2) that result from the accumulation of genomic and epigenomic changes that characterise the mechanisms behind tumour development, progression, and metastasis (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011; Hanahan 2022).

1.1.2 Background – genetic research

Cancer is usually described as a state where normal cells shift to a neoplastic growth state (Hanahan, 2022). The concept that genetic alterations lead to tumorigenesis implies that there would be a high heritability, i.e., increased risk for cancer to be passed on from parent to offspring (Shen and Laird, 2013; Caiado, Silva-Santos and Norell, 2016). It also seemed possible that the genetic basis of cancer could be used to explain racial disparities in cancer risk, for example, breast cancer in African

American women tends to be more aggressive, with worse clinical outcomes (Ahmad et, al., 2017).

Global Cancer Statistics

Male

Estimated Cases

Prostate 1,193,715
Colorectum 800,665
Lung 734,708
Stomach 427,889
Bladder 353,001

Estimated Deaths

Lung 1,188,679
Liver 577,522
Colorectum 515,637
Stomach 502,788
Prostate 375,304

Female

Estimated Cases

Breast 1,835,883
Colorectum 650,766
Lung 435,811
Cervix uteri 410,809
Thyroid 361,802

Estimated Deaths

Breast 684,996
Lung 607,465
Colorectum 419,536
Cervix uteri 341,831
Stomach 266,005

(World Health Organization, 2020)

Figure 1.1: Worldwide cancer statistics. Lung cancer in men and breast cancer in women accounts for most of the cancer related deaths. Most of the cancer related mortality in men and women are caused by lung, breast, colorectum, liver, stomach, cervix, prostate, and thyroid cancers. Brain/ CNS are shown for comparison. Picture designed using BioRender.com.
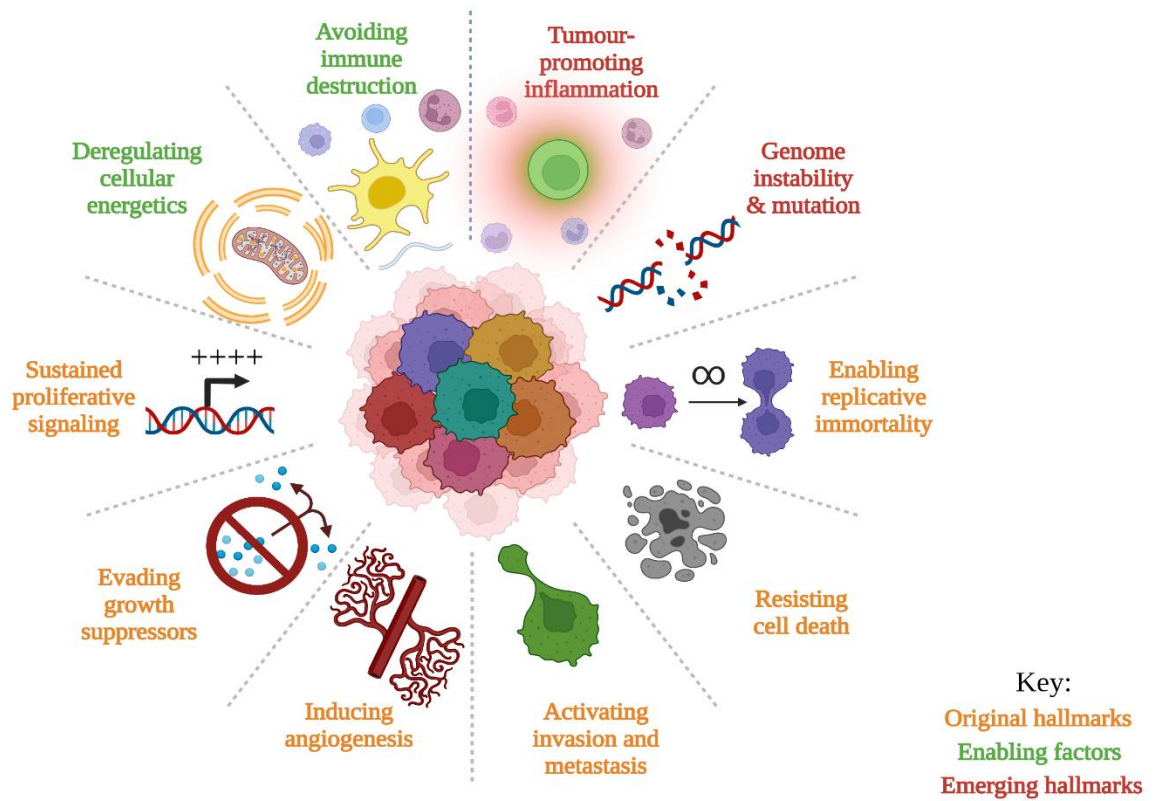
Figure 1.2: Schematic representation of the Hallmarks of cancer. These characteristics represent six original hallmarks previously described by Hanahan and Weinberg (2000) and subsequently validated by the same authors in 2011 and by Hanahan in 2022. All these capabilities enable tumour development, progression, and metastasis. Figure created using BioRender.com template.

The accumulation of mutations in the cells results in genomic instability; and unstable genomes favour tumour development by the activation of oncogenes and the loss of tumour suppressor genes (Negrini, Gorgoulis and Halazonetis, 2010; Duijf et al., 2019). Researchers have delved deeply into studying the role of specific genes, but this has not been enough to explain the diversity and prevalence of a variety of cancers (Wallis and Nam, 2015; Tang et al., 2018; Wang et al., 2020). Tumour cells often show great heterogeneity, which cannot simply be explained by genetic models of tumorigenesis as they can only provide a one-dimensional view (Caiado, Silva-Santos and Norell, 2016). In fact, cancer is often described as a collection of heterogenous cells and stroma (Li, Seehawer and Polyak, 2022). This heterogeneity may even be underestimated, as biopsies often only take a small sample of the variety of cells present in the tumour (Hausser and Alon, 2020).

These findings have been followed up by further investigations that have shown that cancer development is heavily influenced by social/cultural/environmental factors (Ferreira and Esteller, 2018; Darwiche, 2020). So, it was considered valuable to delve deeper, beyond the expression of genes and DNA, to the level of the proteins and amino acids that the genes code for; because DNA, RNA and protein interactions lead to changes in the cellular landscape (Xiao et al., 2019; Sharma et al., 2021).

Epigenetics studies have demonstrated how genes and their products interact to create healthy and disease phenotypes (Zhang, Lu and Chang, 2020). These interactions include biomedical changes to the DNA, nucleosomes, histones, and non-coding RNAs, which can be hereditary without altering the DNA sequence (Dai, Ramesh and Locasale, 2020). Epigenetic alterations have been identified to play a key role in the potentiation, initiation, and progression of tumorigenesis (Flavahan, Gaskell and Bernstein, 2017). This suggests that the study of epigenetic changes in

tumours could have beneficial results for the development of therapeutic strategies that consider the many forces involved in tumour development.

Prior to second generation therapies, tumour treatment was originally a combination of surgery, chemotherapy and/or radiotherapy (Miller et al., 2019). Such treatments were not target specific and the drugs' cytotoxic effects would impact a wide variety of cells in the body, therefore, as therapy delayed death but did not always prevent it, a more targeted approach, looking at cancer at the genetic level, was considered as the next stage of drug development (Falzone, Salomone and Libra, 2018).

The advances in the fields of molecular and cell biology have made possible the development of more successful target therapy that increases survival rates and quality of life for most patients because it achieves a better therapeutic response with less cytotoxic effects (Lee, Tan, and Oon, 2018; Min and Lee, 2022). However, the development of resistance is still a major issue due to tumour heterogeneity and the activation of several mechanisms that reduce or inhibit the efficacy of the target agent (Boumahdi and de Sauvage, 2019).

However, despite having early diagnosis and successful courses of treatment for the primary tumour, ensuring metastatic tumour prevention and early detection is still extremely challenging (Roma-Rodrigues et al., 2019; Soffietti et al., 2020). Existing therapies have proven to be inefficient to target them, as metastatic tumours have genomic and environmental characteristics that differ from the primary malignancy (Ganesh and Massagué, 2021) and most cancer related deaths do not occur due to the primary tumour but as a consequence of metastatic tumour development (Zhang et al., 2022). Hence the need to understand the mechanisms underlying metastatic tumour formation that could lead to the development of more appropriate therapeutic

approaches. Therefore, research such as the present study is important, to narrow down further the drivers of metastatic cancer, and to be able to target them more accurately, perhaps before resistance occurs.

1.2 Breast Cancer

1.2.1 Statistics

According to worldwide cancer statistics female breast cancer is the most diagnosed form (2.3 million cases), followed by lung (2.2) and prostate tumours (1.4) (Ferlay et al., 2021). Male breast cancer accounts for 1% of all tumours in men, and 1 per cent of breast cancer overall (Gucalp et al., 2018). However, cancer of the breast is not as lethal as other forms: such as lung (mortality of 1.79 million), liver (830,000) and stomach cancers (769,000) (Ferlay et al., 2021). The estimated incidence of female breast cancer in the United States for 2022 will be 287, 850 cases, with the estimated number of deaths at 43,250 (Siegel et al., 2022). This discrepancy between incidence and fatality could be due to early detection and better clinical management (Ahmad, 2019). Breast cancer rates second in cancer deaths in women, (Wu, Sarkissyan and Vadgama, 2016) behind lung cancer (Siegel et al., 2022).

1.2.2 Breast cancer biology

The breast tissue undergoes many changes throughout the female life cycle, changing with the variation in hormones that come with puberty, with pregnancy and lactation. Therefore, these cells must be capable of considerable growth, invasion, and multi-lineage potential (Dravis et al., 2018). As with other types of cancer, breast carcinoma is not only one disease, there are also several varieties, with heterogeneity of the

tissues, at the molecular and cellular level, they present different symptoms and have wildly different outcomes (Peng et al., 2020).

Histologically, breast cancer is classified based on how much the cells look like normal breast tissue and the rate at which they are growing. Following this criteria breast cancer can be Grade 1 or well differentiated breast cancer. Cell look similar to the mammary tissue and grow slowly. Grade 2 or moderately differentiated. Cells look less like breast tissue and growing at a faster rate. Grade 3 or poorly differentiated. Cells look completely different to breast tissue. These cells are growing very fast and are more likely to metastasise (Rakha et. al.,2010).

Molecular biomarker heterogeneity classifies breast cancer into three clinical subtypes, ER positive, progesterone receptor positive and HER-2 positive based on oestrogen, progesterone, and HER-2 receptor status respectively (Gu, Dustin, and Fuqua, 2016). Molecular heterogeneity allows for the classification of breast cancer into four molecular subtypes: HER-2 positive; ER negative, progesterone receptor negative and HER-2 positive. Triple negative breast cancer (TNBC); ER negative, progesterone receptor negative, and HER-2 negative. Luminal A: ER positive, progesterone receptor positive and HER-2 negative. Luminal B; ER positive, progesterone receptor positive and HER-2 positive or negative (Watkins, 2019). Each subtype is associated with specific risk factors, metastatic tendency and require different therapeutic approaches (Waks and Winer, 2019).

Molecular based analyses have made possible the classification of tumour associated genes that allow for the prediction of recurrence risk and therapeutic response (Turashvili and Brogi, 2017). Genomic and transcriptomic analyses have detailed somatic mutations in several cancer genes including TP53, PIK3CA, PTEN, TBX3,

MLL3, RUNX1 and CBFB that have frequently been found to be associated with the breast cancer subtype (Koboldt et al., 2012; Curtis et al., 2012; Fusco et al., 2021). Identification of tumour associated genes has led to a better understanding of the molecular basis of breast cancer. However, this characterisation has not solved the problem of therapeutic resistance and tumour metastases (Grote et al., 2021). It has been reported that patients with the same breast cancer subtype, undergoing the same course of treatment, have different responses that may lead to resistance to therapy and metastasis (Turner et al., 2021). Regardless of the advances in early detection and therapeutic treatment that have increased patients' overall survival, resistance to therapy and metastatic disease are still almost impossible to prevent (Peng et al., 2020). The tumour-associated immune cells, the conditions of the tumour microenvironment, and tumour genomic alterations trigger cellular reprogramming that promote intratumor heterogeneity, which can lead to resistance to therapy and metastatic tumour progression (Chung et al., 2017).

1.3 Metastasis

1.3.1 Definition and description

The capability of cancer to metastasise is one of the hallmarks of cancer that occur when the cancerous cells go from the primary tumour to other parts of the body through the vascular and lymphatic systems (Guan, 2015; Hanahan, 2022). Metastatic tumours account for most of the cancer related deaths predominantly due to the advances in early detection and effective therapeutic approach of the primary malignancies (Gao et al., 2019).

Metastatic development is not a random event, it follows an organ specific pattern (organotropism) (Pretzsch et al., 2019). The seed and soil hypothesis of metastatic

development departs from the premise that a supportive niche is essential for successful metastatic growth (Liu et al., 2017; Langley and Fidler, 2011). It has been demonstrated that the distant organs targeted for tumour metastasis are pre-conditioned by the primary tumour by secreting tumour-derived soluble factors such as VEGF-A, TNF-α, TGF-β, PDGF and the activation of regulatory signals that recruit bone marrow derived myeloid and hematopoietic cells, both of which are required for implantation of the tumour cells in the new microenvironment (Peinado, Lavotshkin and Lyden, 2011; Gao et al., 2019; Lu et al., 2020). Ultimately, this results in the creation of a pre-metastatic tumour microenvironment that is free of tumour cells but has acquired cancer supporting characteristics that will protect and promote metastatic cell growth and survival (Izraely and Witz, 2020). Furthermore, tumour dormancy occurs, which is the capability of the circulating tumour or micro metastases to remain in small numbers following surgical removal of the primary tumour (Gomis and Gawrzak, 2017). Dormant cells remain undetected for years or decades, during this period the cell will acquire new mutations that will support their adaptation in the new microenvironment and metastasis development. Additionally, the cells will be resistant to therapeutic treatment and immune attacks (Recasens and Munoz, 2019).

1.3.2 The Metastatic Cascade

The metastatic cascade is a multistep process influenced by the pre-metastatic niche (Dujon et al., 2021). It includes a series of well-regulated events (Figure 1.3) that include invasion into the local tissue and the circulatory and lymphatic systems; followed by migration and extravasation from the blood stream into the adjacent tissue that if the cells survive will result in the development of metastatic tumours in the target organs (see picture 1.2) (Chin and Wang, 2016; Dujon et al., 2021).

The initial stages of the metastatic cascade are local tissue invasion and migration by the primary tumour cells (Hapach et al., 2019). Invasion is characterized by increased cell motility caused by the dysregulated expression of extracellular matrix (ECM) adhesion molecules, and the upregulation of proteases from the extracellular space result in breakdown of the ECM which promote tumour cell invasion and growth (Rankin, Nam and Giaccia, 2016). Migration requires the release of growth factors, cytokines and cell adhesion molecules that will promote migration into the neighbouring tissue and protect from immunological surveillance (Popper, 2020). In addition, low oxygen tension in the tumour microenvironment also influences tumour metastasis via upregulation hypoxia inducible factor (HIF) by the tumour cells. HIF regulates epithelial-mesenchymal transcription, via upregulation of several transcription factors that support tissue invasion, migration, and metastasis (Mujcic et al., 2014). Hypoxia also regulates the breast cancer cells release of exosomes (King, Michael, and Gleadle, 2012). Exosomes are small membrane vesicles that regulate cell interactions, and in cancer exosomes are major mediators of the crosstalk between tumour and stroma cells (Steinbichler et al., 2017; Tian, Liu, and Li, 2019).

It has been demonstrated that the development and survival of metastatic tumours is highly dependent on the capability of the invading tumour cells to survive immune cell attacks in the new microenvironment (Bates et al., 2018). Immune evasion is mediated by the tumour supporting macrophages that have been found surrounding circulating tumour cells, they release cytokines such as IL-8 that impair the native immune response via the activation of the STAT3 pathway (Wu et al., 2019). Also, tumour cells hijack the antigen presenting cells to get immune protection (Bates et al., 2018).

Another way by which tumour cells avoid immune destruction is via regulatory T-cell downregulation of natural killer cells and cytotoxic T-cells (Drapela and Gomes, 2021).
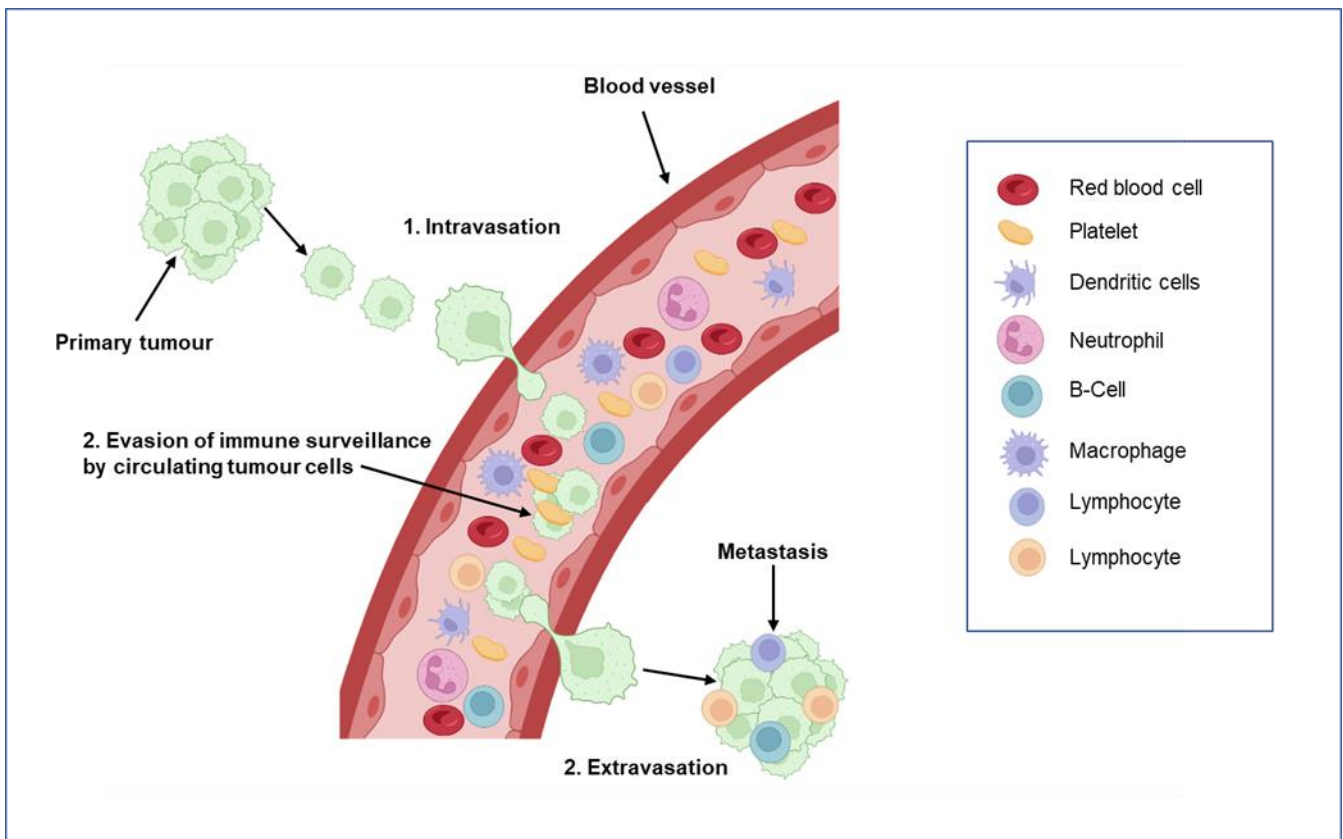
Figure 1.3: The metastatic cascade. Schematic representation of the central steps of metastatic tumour cell dissemination which involve the intravasation of tumour cells into the circulatory system, evasion of immune system surveillance and the final step of extravasation into a distant tissue. Figure adapted from Xiaoming et, al., 2020 using BioRender.com.

1.3.3 Epithelial-mesenchymal transition

Epithelial-mesenchymal transition (EMT) is a critical tissue transformation during the development of the embryo (Chen et al., 2010). It is an essential part of cancer invasion, tumour metastasis, and therapeutic resistance (Zhang and Weinberg, 2018). Malignant cells of epithelial origin escape their primary microenvironment by expressing fibroblast-like characteristics and increased cell adhesion and motility which can result in the development of metastases (Hao, Baker, and Ten Dijke, 2019).

EMT is a highly controlled cellular transforming stage characterised by dysregulated expression of adhesion molecules, cell cytoskeleton reorganisation, and degradation of the basement membrane by metalloproteinases; all which result in increased cell motility and metastatic capability (Loh et al., 2019). The regulation of these processes involves cytokines and transcription factors such as: fibroblast growth factor (FGF), epidermal growth factor (EGF) receptor, bone morphogenic protein (BMP), hepatocyte growth factor (HGF), transforming growth factor (TGF-β), Twist1, Snail1, and ZEB (Celià-Terrassa et al., 2017). These molecules are formed and secreted by the tumour microenvironment in response to oxygen deprivation, acidosis, and nutrient scarcity (Parlani, Jorgez and Friedl, 2022). Furthermore, a low oxygen tumour microenvironment results in the activation of hypoxia inducible factor (HIF) activity that will initiate the epithelial-mesenchymal phenotype which will result in decreased cell polarity resulting in increased cancer invasiveness and metastases (Ribatti, Tamma and Annese, 2020). In human metastatic cancers, overexpression, and upregulation of EMT transcription factors have been directly linked with tumorigenesis and metastatic progression (Dongre and Weinberg, 2018; Parlani, Jorgez and Friedl, 2022). The maintenance of the mesenchymal phenotype in human carcinomas is primarily regulated by Snail and ZEB transcription factors (Kang et al., 2021). In breast

cancer, Snail is associated with increased tumour invasiveness and poor survival outcomes (Jin et al., 2019).

Similarly, EMT is characterised by overexpression of vimentin and N-cadherin (Lamouille, Xu and Derynck, 2014). E-cadherin, claudins and cytokeratin are downregulated, resulting in loss of epithelial integrity (Reddy et al., 2005), (Figure 1.4). E-cadherin downregulation is also associated with the increased metastatic potential of solid tumours (Sommariva and Gagliano, 2020). However, this molecule has also been found to be upregulated in patients with glioblastoma, ovarian and breast carcinomas, resulting in increased tumour invasion which suggests that E-cadherin signalling consequence may be tumour dependent (Putzke et al., 2011; Reddy et al., 2005; Van Roy, 2014).

Drugs that target epithelial-mesenchymal transition have proven to be successful in preventing metastatic tumour formation in several cancer types (Liu, Smith, and Wang, 2022). Breast cancer patients treated with Eribulin, a synthetic drug with antimitotic effects, showed to be beneficial stopping EMT and metastatic progression (Yardley et al., 2019; FUJII et al., 2020).

Dissemination of cancer cells requires changeable conversion between the epithelial and mesenchymal states (Gallardo et al., 2022). Studies have found that metastatic tumours in distant organs present a more differentiated epithelial cell phenotype which suggests that there is reacquisition of epithelial genes via mesenchymal-epithelial transition (MET) (Chaffer, Thompson, and Williams, 2007). Mesenchymal-epithelial transition (MET) is crucial for organogenesis, and it has been determined to be important for metastatic development (Owusu-Akyaw et al., 2018; Cho et al., 2019).
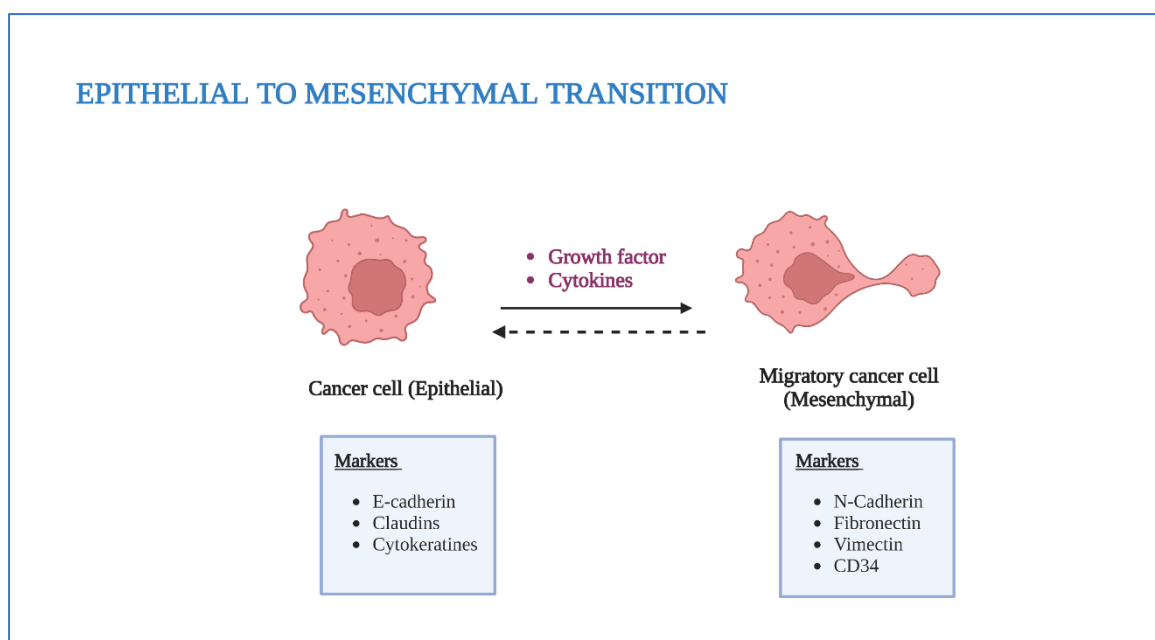
Figure 1.4: Epithelial-Mesenchymal Transition. Reversible tissue transformation during which epithelial cells acquire mesenchymal characteristics. In cancer cells several transcription factors and adhesion molecules regulate EMT downregulation and overexpression of epithelial and mesenchymal markers, respectively which results in increased migration, cell invasiveness and metastatic potential. Figure created using BioRender.com.

1.4 Angiogenesis

1.4.1 Definition

Angiogenesis is when new blood vessels grow via two mechanisms, either by pre-existing endothelial cells sprouting, or by intussusception which consists of the splitting of existing vasculature to create new blood vessels (Carmeliet and Jain, 2011; Petel-Hett and D'Amore, 2011). Both physiological and pathological states require the growth of new blood vessels (Felmeden, Blan et al., 2003). Physiological angiogenesis

requires an appropriate balance between anti-angiogenic and pro-angiogenic factors, whereas pathological angiogenesis is the result of upregulation or downregulation of angiogenic stimuli.

1.4.2 Molecular mechanism of angiogenesis

The development of new blood vessels requires the coordinated action of several signalling molecules that promote endothelial cell migration and proliferation towards the stimulatory signal (Yadav, 2015). VEGF, ANG2, FGF and chemokines are several of the survival proteins released by endothelial cells in response to low oxygen levels in their microenvironment (Liu et al., 2018). VEGF-A regulates both physiological and pathological angiogenesis (Nagy et. al., 2007) and is the most well-characterized VEGF family member (Otrock et.al., 2007). VEGF regulates angiogenesis in endothelial cells by signalling through VEGFR-1 and VEGFR-2 found in the cell surface (Ferrara et al., 2009). Schematic representation of the different VEGF signalling is shown in Figure 1.5.

During angiogenesis, VEGF is responsible for enabling vascular permeability which releases the plasma proteins into the surrounding tissue, establishing a temporary extracellular matrix (ECM) for the new blood vessel (Carmeliet and Jain, 2011). Also, it promotes the tip cell phenotype by upregulating DLL4 expression which results in down-regulation of VEGFR-2 in the stalk cells; ensuring that the leading role during angiogenesis goes to the tip cells (Jakobsson et al., 2010). Furthermore, VEGF promotes the expansion and branching of the new blood vessel (Hiratsuka et al., 2005).

Inhibition of VEGF in animal studies has resulted in the suppression of tumour growth (Carmeliet and Jain, 2011, Ellis and Hicklin, 2008). The VEGF/VEGFR axis is essential

for vessel development, hence several antiangiogenic therapies have been developed for targeting their interaction (Park et al., 2018, Kim et.al., 2018). The approach consists of blocking either the ligand or the receptor with an antibody, or the tyrosine kinase domain with tyrosine kinase inhibitors (Vasudev and Reynolds, 2014).
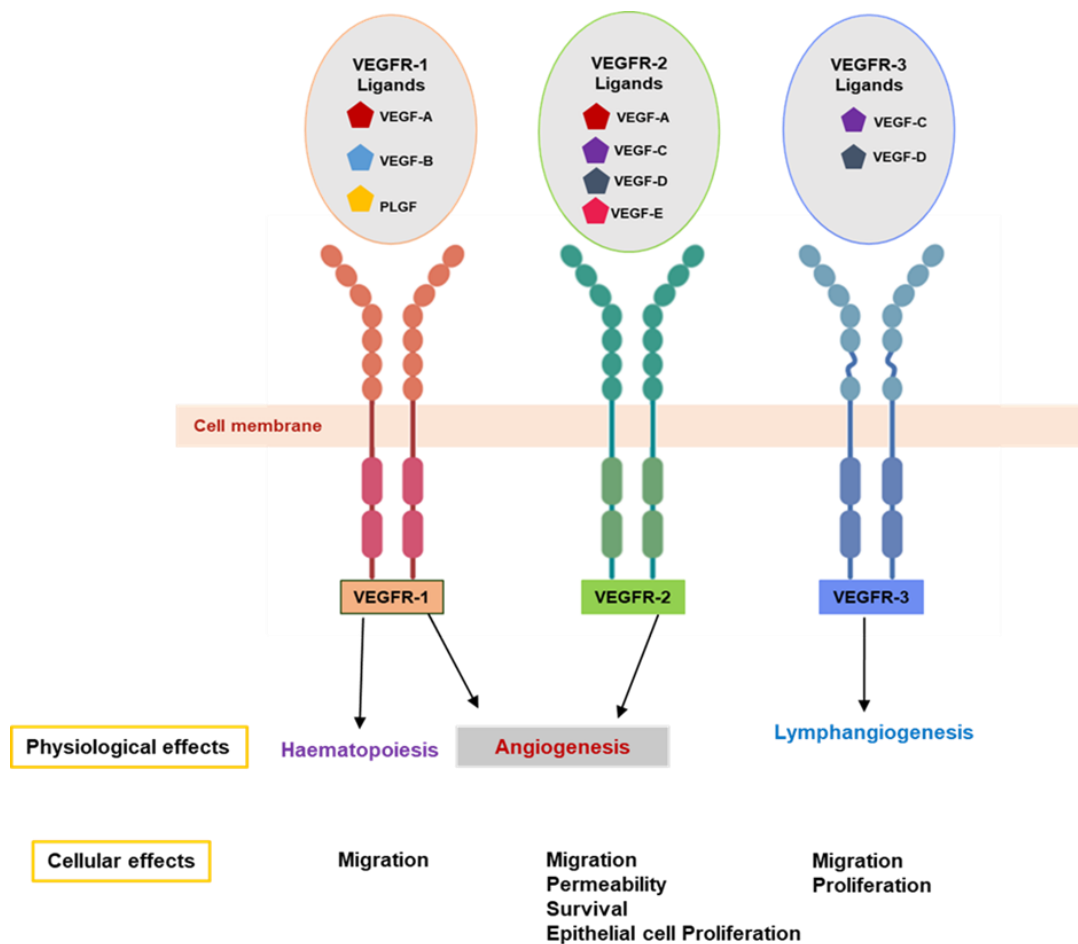
Figure 1.5: Schematic representation of VEGF receptors and their specific ligands. All VEGFR-A isoforms bind to Ig domain 1 and 2 in both VEGFR-1 and VEGFR-2. Expression of VEGFR1 is found in monocytes, macrophages, haematopoietic stem cells and the vascular endothelium. VEGFR2 expression is found in both the vascular endothelium and the lymphatic endothelium. VEGFR3 expression is predominantly limited to the lymphatic endothelium. In VEGFR-3 a disulfide bridge

replaces the fifth Ig domain. (Adapted from Holmes et al., 2007, and Abhinand et al., 2016).

1.4.3 Tumour angiogenesis

The development of new blood vessels is deemed one of the hallmarks of cancer by Hanahan (2022). Adequate vascular supply is not only essential for tumours' initial development and continuous growth but without new blood vessel formation tumours can only grow 1-2 mm; it also enables tumour cells to travel in the blood circulation and metastasise to distant organs (Yadav et al., 2015). In tumour angiogenesis, the newly developed blood vessels infiltrate the mass of cancer cells delivering nutrients and oxygen (Morbidelli et.al., 2018).

1.4.4 Tumour angiogenic switch

The angiogenic switch is referred to as the shift in the equilibrium of pro- and anti-angiogenic factors that can take place at any stage of tumour evolution (Yadav et al., 2015) and it will exhibit different characteristics within the same cancer tissue, depending on the anatomic location (Lugano, Ramachandran and Dimberg, 2019).

Tumour angiogenesis is a very deregulated process due to the continuous presence of angiogenic factors in the tumour microenvironment (De Palma, Biziato and Petrova, 2017). The nascent blood vessels often present an abnormal phenotype resulting in inadequate blood flow within the tumour (Krishna Priya et al., 2016; Hanahan and Weinberg, 2011). Tumour stroma often presents with areas of persistent or sporadic hypoxia due to inadequate tumour vascularization (Aguilar-Cazares et al., 2019).

Tumour vascularization is carried out by three main mechanisms: vascular co-option, vascular mimicry, or through trans-differentiation of cancer circulating cells into endothelial cells (Kuczynski et al., 2019) regulated by the interaction between tumour-

secreted angiogenic factors and their receptors found in endothelial cells (Hillen and Griffioen, 2007; Siveen et al., 2017). There are a number of promotors and inhibitors that direct the tumour angiogenic switch including VEGF, bFGF, HIF and IGF (Schneider et al., 2017). VEGF-A is a well-known angiogenic inducer that plays a major role in tumour angiogenesis (Macedo et al., 2017; Wang et al., 2018).

Higher micro vessel density and lower patient survival rate are associated with high levels of VEGF-A in tumours (Toi et al, 1994). Similarly, uninterrupted blood vessel growth in tumours and the development of resistance to anti-angiogenic therapy has been associated with persistent upregulation of bFGF (Zahra, Sajib and Mikelis, 2021).

1.4.5 Anti-angiogenic therapy for cancer

Inhibition of VEGF in animal studies has resulted in the suppression of tumour growth (Carmeliet and Jain, 2011, Ellis and Hicklin, 2008). The VEGF/VEGFR axis is essential for vessel development hence several antiangiogenic therapies have been developed for targeting their interaction (Park et al., 2018, Kim et.al., 2018). The approach consists of blocking either the ligand or the receptor with an antibody, or the tyrosine kinase domain with tyrosine kinase inhibitors (Vasudev and Reynolds, 2014).

1.4.6 Suggested Tumour evasion mechanisms of anti-VEGF therapy

Targeting tumour angiogenesis is critical to stop solid tumours from growing and proliferating which has resulted in anti-angiogenic therapy being the standard of treatment to prevent cancer development and dissemination (Lu and Bergers, 2013). During tumour development, the VEGF ligand-receptor signalling pathway is crucial for new blood vessel formation. Therefore, anti-VEGF therapy is routinely used to target tumour angiogenesis (Yadav et al., 2015). However, blocking VEGF activity often proves ineffective because tumours develop resistance to anti-VEGF therapy,

and activate VEGF-independent pro-angiogenic pathways (Itatani et al., 2018). Several pro-angiogenic factors have been suggested to be the alternative growth factor driving tumour angiogenesis (Teleanu et al., 2019). The development of new drugs against each pro-angiogenic molecule represents a potential approach for targeting tumour angiogenesis. However, this strategy might not be the best method as tumours may divert to a different pro-angiogenic molecule to escape treatment. Hence, identifying a downstream molecule targeted by all major pro-angiogenic factors could represent a more efficient molecular target for anti-angiogenic therapy. Based on preliminary testing by Dr. Angel Armesilla's group laboratory at the University of Wolverhampton, it has been hypothesized that ATF2 may be the common downstream molecule for all major pro-angiogenic pathways.

1.5 Transcription Factor ATF2

1.5.1 Background information

ATF2 is part of the AP1 family of transcription factors that regulate several cellular functions through gene expression regulation in response to hypoxia and DNA damage (Bhoumik et al, 2007). The ATF2 coding gene is located in chromosome 2q32 and conformed by 15 exons that code for a 505 amino acid protein (Lou and Ronai 2012, Ozawa et al., 1991). Structurally, ATF2 consists of an N-terminal Zinc finger domain (ZF), a transcriptional activation domain (TAD), and C-terminus basic leucine zipper domain (bZIP) (Lopez-Bergami et.al., 2010) (Fig 1.1.6). ATF2 is found ubiquitously expressed throughout the body with expression being particularly rich in the brain (Takeda et al., 1991). The ATF2 gene is essential during early development. Complete loss of the ATF2 gene causes postnatal death in mice (Ackermann et.al.,

2011, Chen et al., 2008) whereas partial deregulation is associated with cancer development (Bruhat et la, 2007; Papassava et al., 2004).

1.5.2 Mechanism of ATF2 activation

ATF2 activation is required in response to stress signals. Activation of ATF2 is carried out by different stimuli including ultraviolet (UV) radiation, growth factors and cytokines (Watson, Ronai and Lau, 2017). Transcriptional activation is regulated by the TAD at the N-terminal. Phosphorylation by stress-activated protein kinases of amino acid residues Thr69 and Thr71 (Figure 1.6) is essential for ATF2 transcriptional activity (Lopez-Bergami et.al., 2010, Raingeaud et al.,1995). Following phosphorylation, the autoinhibitory molecular interactions are lifted thus allowing ATF2 to dimerize with other AP1 (Watson, Ronai and Lau, 2017). Activated ATF2 regulates the expression of hundreds of genes when it translocates into the nucleus of the cell (Lau and Ronai, 2012, Li and Green, 1996, Livingston et al., 1995). ATF2 homodimerization has limited efficacy, therefore, heterodimerization with other AP1 proteins is required for efficient ATF2 transcriptional activity (Lau and Ronai, 2012). Depending on the cell type and the nature of the stimulus, ATF2 will heterodimerize with specific AP1 partners which can result in either induction or repression of transcription (Lopez-Begami et al., 2010; Lau and Ronai, 2012; Gong et al. 2002,). Studies in endothelial cells showed that treatment with VEGF or EGF caused c-Jun/ATF2 heterodimerization which resulted in increased cell survival by activating anti-apoptotic protein BCL2L1 (Salameh et al., 2010). Another study showed that decreased polyamide levels in intestinal epithelial cells resulted in an increase in JUND/ATF2 heterodimerization causing transcriptional repression of essential cell cycle kinase CDK4 (Xiao et al., 2010). ATF2 transcriptional control of essential genes such as RB1, cyclin A, cyclin D, and Bcl2 (among others) drives ATF2 regulation of cell cycle progression and cell survival (Ma Q et. al., 2007).

Additionally, phosphorylation of ATF2 at different amino acid residues will also have a role in ATF2 regulation of gene expression. ATF2 phosphorylation by the ataxia-telangiectasia gene (ATM) is responsible for ATF2- mediated response to DNA damage due to ionizing radiation (Bhoumik et al., 2005). Acetylation of ATF2 at lysine 357 and 374 (Lys357 and Lys374) by the histone acetyltransferase p300/CREB-binding protein also modulates ATF2 transcriptional activity, however, the exact mechanism of action is not well understood (Karanam et al., 2007).
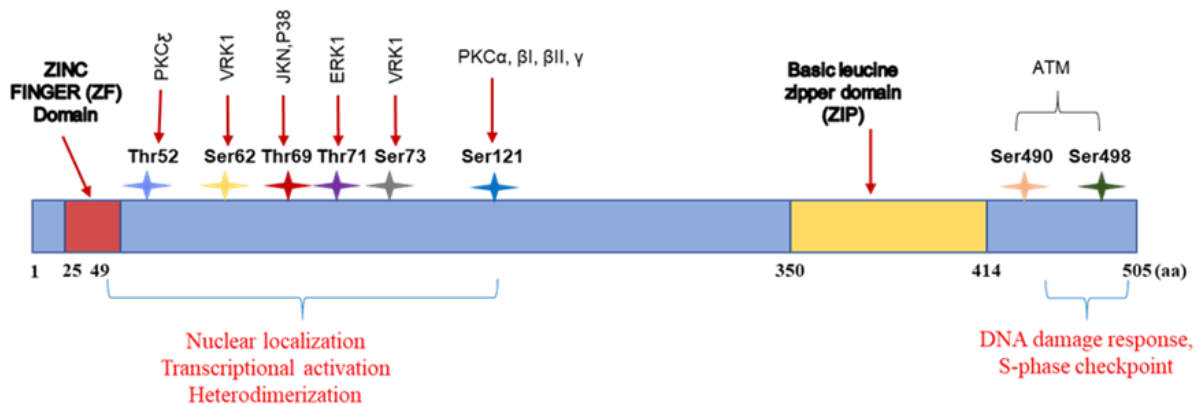
Figure 1.6: Graphical representation of ATF2 phosphorylation sites. Following Phosphorylation (Activation) at Thr69 by JNK and P38, and at Thr71 by ERK1; ATF2 protein regulates the expression of several genes with help from other AP1 transcription factors when it translocates into the cell's nucleus. (Adapted from Lau and Ronai, 2012).

1.5.3 The role of ATF2 during angiogenesis

The role played by ATF2 during angiogenesis is not well understood. Research studies have shown that ATF2 is required for VEGF-stimulated endothelial cell responses

such as cell migration and tubule formation. ATF2 knockdown resulted in inhibition of VEGF-dependent cell migration and tubule formation in endothelial cells (Fearnley et al., 2014). AFT2 has also been found to be part of a pro-angiogenic stimulation in an in vitro study in nasopharyngeal carcinoma. The study found that EBNA1 (Epstein–Barr viral nuclear antigen 1), a protein produced by the Epstein Barr Virus gene, increases the expression of AP1 proteins. EBNA1 binds to the promoter of these proteins including c-Jun and ATF2, leading to an increase in microtubule formation due to enhanced expression of AP1 target genes such as VEGF and interleukin 8 (O'Neil et al., 2008). Furthermore, ATF2 transcriptional activity is necessary for VEGF-dependent endothelial cell expression of adhesion molecule VCAM-1 (Fearnley et al., 2014). VCAM, like other endothelial adhesion molecules, promotes the accumulation of leukocytes in inflamed tissues, a process trailed by the development of new blood vessels (Imhof and Aurrand-Lions, 2006). Additionally, preliminary results from our laboratory show that the transcription factor ATF2 is activated by major pro-angiogenic factors (VEGF, HGF, bFGF, HB-EGF) in endothelial cells. Therefore, studying ATF2 transcriptional targets and their potential role in new blood vessel formation can provide a better understanding of the angiogenic process in tumours, and potentially lead to the development of target therapy to stop tumour angiogenesis.

1.6 Breast to Brain Metastasis

1.6.1 Background

Brain metastases are the most commonly reported type of brain tumour (Bailleux, Eberst and Bachelot, 2020). They mainly originate from tumours of the lung and the breast (See Figure 1.7) (Bastianos et al 2015; Morgan, Giannoundis and Palmieri,

2021). Estimates of breast to brain metastasis incidence are often underestimated as the diagnosis is usually made based on the presentation of symptomatic disease rather than early detection (Suh et al., 2020). Thus, twenty to thirty percent of patients with primary breast cancer will develop metastases and about ten to sixteen percent of those will be diagnosed in symptomatic patients and about thirty percent in autopsy reports (McMullin et al., 2014; Huang et al., 2021; Zimmer, Van Swearingen and Anders, 2020). The gold standard for treatment includes the combination of whole brain radiation followed by surgical excision where possible, or stereotactic radiosurgery (Mills et al., 2020). Even though several advances have been made to deliver drugs to targets inside the brain, the used of chemotherapeutic agents remains limited due to inadequate penetration of the blood brain barrier (Park et al., 2020; Tashima et al., 2022).

Major risk factors predisposing patients to the development of breast to brain metastases include young age, an oestrogen receptor negative breast cancer, Her-2+ overexpression, nodal invasion, and at least metastases to two secondary organs at the time of diagnosis (Koniali et al., 2020).

Cancers that most commonly metastasise to the brain
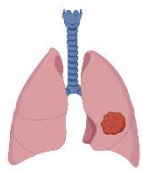
Lung cancer
~60%*

Breast cancer
~11%*

Melanoma
~6%*
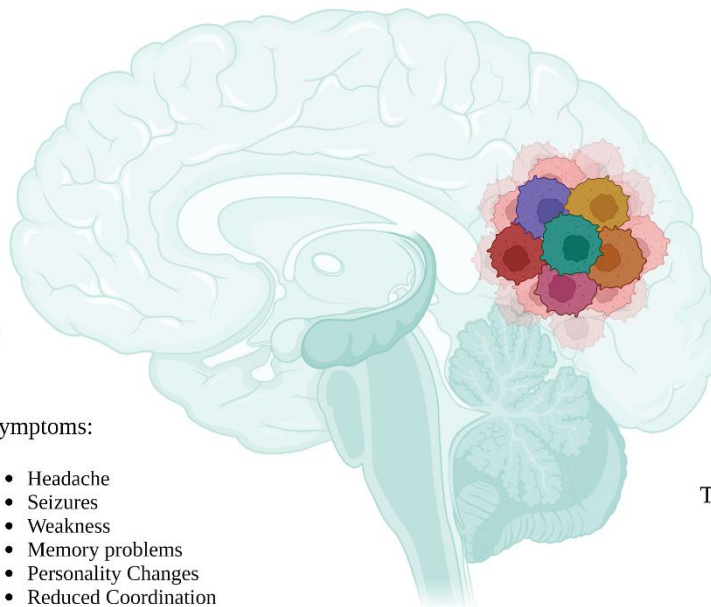
Symptoms:

- Headache
- Seizures
- Weakness
- Memory problems
- Personality Changes
- Reduced Coordination
- Loss of balance

Treatments:

- Radiotherapy
- Steroids
- Surgery
- Chemotherapy
- Hormone therapy
- Targeted drugs
- Immunotherapy

*Percentage of overall brain metastases cases

Figure 1.7: Top three cancers that metastasise to the brain. Brain metastases are a common complication of primary tumours. They are the most commonly reported brain tumours. Sixty percent of all brain metastatic tumours are from primary lung carcinoma, making them the most common type of tumours that metastasise to the brain, this is followed by breast with around eleven percent of cases and melanoma accounting for six percent (Habbous et al., 2020). Figure created with BioRender.com.

1.6.2 The molecular biology of breast to brain metastasis

Metastatic tumours seem to favour certain organs depending on the originating primary malignancy even though tumour cells disseminate throughout the body (Louie et al., 2013; Altaleb, 2021). This indicates that there are molecular and biochemical patterns during tumour evolution that need to be discovered to support the characterisation of early diagnostic markers for early and/or efficient therapeutic targets.

Gene expression analysis has shown that metastatic tumours and their progenitors have similar genomic alterations (De Mattos-Arruda et al., 2018). However, it has also been established that metastatic tumours continue to evolve away from the primary tumour of origin and that they harbour more and new genomic aberrations (Tao et al., 2020). Schematic representation of breast to brain metastatic tumour development is shown in Figure 1.8.

Studies have shown some genomic alterations that are specific for the development of breast to brain metastasis (Morgan, Giannoudis and Palmieri, 2021, Giannoudis et al., 2021; Rinaldi et al., 2020; Brastianos et al., 2015). BARD1 and RAD51 are genes implicated in DNA repair mechanisms; overexpression of these genes was found in metastatic brain tumours for breast cancer but not in their primary originating counterpart (Duchnowska et al., 2015; Woditschka et al., 2014). These genes are believed to provide cancer cells with the mechanisms to overcome the genotoxic effect of reactive oxygen species in the brain (Woditschka et al., 2014).

Breast to Brain Metastasis
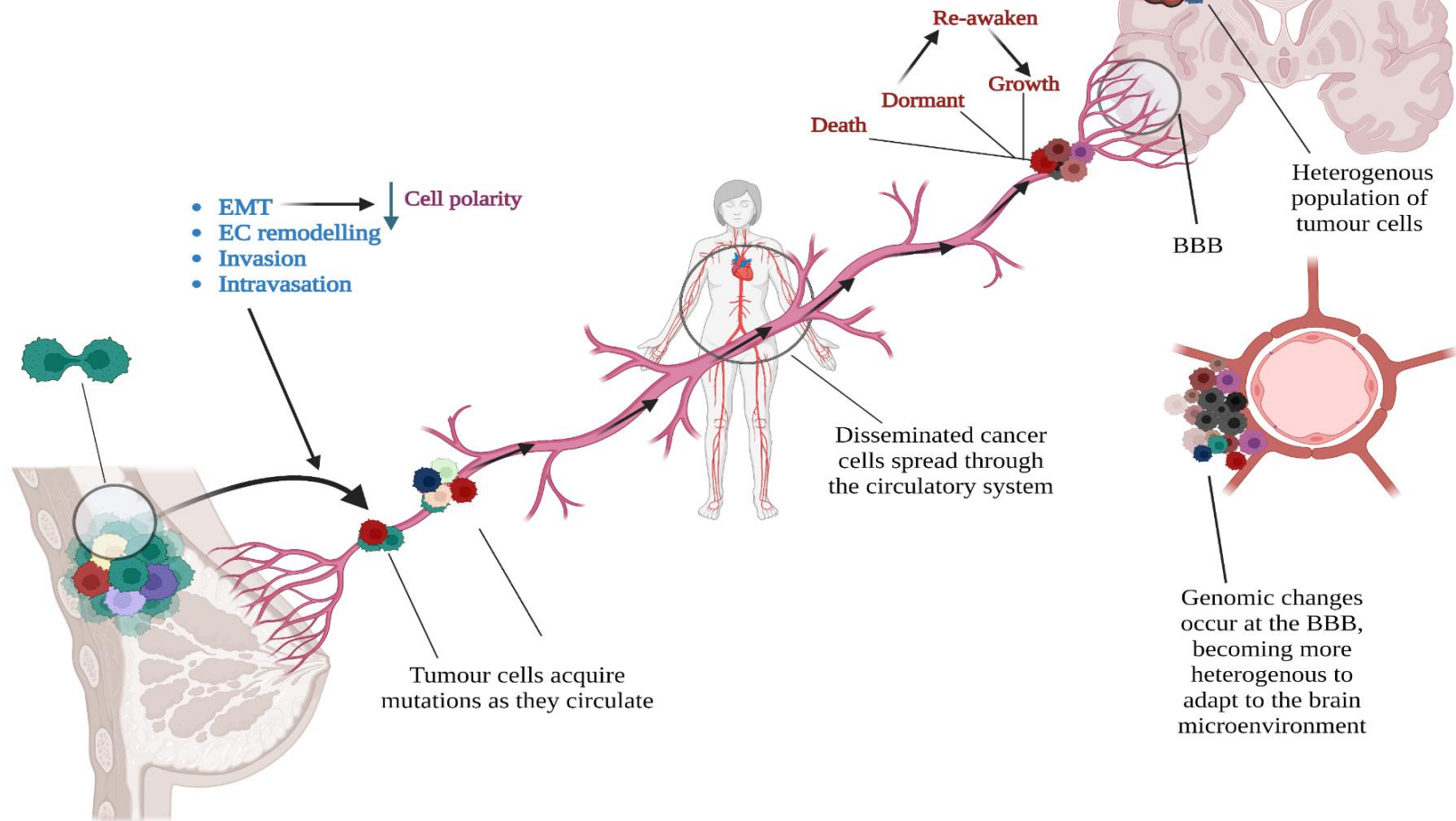
Figure 1.8: Schematic representation of breast to brain metastases development. A small population of cells with an invasive phenotype leave the primary tissue site. Basement membrane disruption through remodelling of the extracellular (EC) matrix allow the invasive cells to interact with endothelial cells and enter the circulatory system (intravasation). Circulating tumour cell tight junction interactions at the blood brain barrier (BBB) with the brain stromal cells allow them to the brain. After crossing the BBB most cells will die, some will go dormant and very few or none will grow and colonise the brain.

Recent studies of the genomic landscape of breast to brain metastases have indicated that the most frequently mutated genes in breast to brain metastases are TP53, PIK3CA, ESR1, BRCA2, Notch1, and amplifications of ASXL1 and PTEN deletions (Morgan, Giannoudis and Palmieri, 2021, Giannoudis et al., 2021; Rinaldi et al., 2020). Alterations of TP53 and PIK3CA genes remain the most common mutations identified in brain metastatic tumours, thus making them the most appropriate targets for therapeutic approach of these malignancies (Morgan, Giannoudis and Palmieri, 2021). However, these mutations are also commonly found in tumours that do not metastasise to the brain so they may not specifically drive this organotropism (Brastianos et al., 2015).

Several of the identified genes by genomic profiling of breast to brain metastases may represent plausible therapeutic targets (Morgan, Giannoudis and Palmieri, 2021, Giannoudis et al., 2021; Rinaldi et al., 2020; Zimmer, Van Swearingen and Anders, 2020). However, despite the efforts in characterising the genomic profiling of breast to brain metastases, finding successful courses of treatment remains an unmet need. A more holistic understanding of metastatic tumour development is necessary to create a rational approach to treat these deadly malignancies. A possible successful treatment would have to account not only for the alterations that promote predisposal to metastatic tumour development and the molecules that allow for the extravasation and invasion of the brain, but also for the modifications acquired in the brain that ensure the colonisation and survival of the invasive tumour's cells.

The data that will be presented in the following results chapters represent a preliminary contribution to identifying genomic alterations that may be implicated in tumour adaptation and survival in the brain microenvironment. Additionally, functional analysis

data of an ARFGEF3 protein coding gene revealed promising regulatory functions for breast to brain metastases adaptation and colonisation of the brain.

1.6.3 Brain tumour microenvironment

The brain microenvironment is anatomically and chemically different than the microenvironment of the extracranial malignancy (Álvaro-Espinosa et al., 2021). Bidirectional interaction between tumour cells and their microenvironment (which is composed of malignant and non-malignant cells) is required to ensure colonisation of the brain (Abbott, Rönnbäck and Hansson, 2006; Tang et al., 2020). These interactions give tumour cells the ability to modulate immune response, cell survival signalling pathways, as well as control of the metabolic and catabolic process to fulfil their nutrient requirements (Zheng et al., 2019; Neman et al., 2014).

Development of brain metastases was once considered to be a random event, however the expression of specific molecules by primary tumour cells that aid disseminated cells to cross the non-fenestrated capillaries that form the BBB have modified this concept (Lowery and Yu, 2017). Molecules such as ST6GALNAC5 and EGFR ligand HBEGF are implicated in the metastatic invasion of the brain by enhancing tumour cell adhesion to the brain epithelial cells (García-Gómez et al., 2019). However, crossing the BBB does not imply the colonisation of the brain by tumour cells; interplay between the new tumour microenvironment and the invasive cells determines colonisation (Cacho-Diaz et al., 2020). Interestingly, the process of adaptation to the brain microenvironment seems to start before the tumour cells even cross the BBB (Schulz et al., 2019). It has been determined that breast tumours express neuronal growth factors and GABAergic signalling which is believed to confer adaptative advantage to the metabolic conditions in the brain (Venkatesh and Monje,

2017). The PI3K/AKT signalling pathway has also been implicated in tumour cell extravasation into the brain by promoting the activation of central nervous system macrophages (microglia) to promote brain invasion and colonization by breast cancer cells via secretion of proteases (e.g., Ctss, Mmp3, and Mmp9) and chemokines (e.g., Cxcl12) (Basques et al., 2018; Li et al., 2018).

Recent studies have shown that that tumour cell colonization of the brain require tumour cells to adopt brain like properties that allow them to survive (Schulz et al 2019). Intriguingly, it has been shown that tumour cells co-opt the neuronal signalling network by expression of GABA receptors and by using GABA as an oncometabolite (Neman et al., 2014). Additionally, supplanting of the astrocytes at the synaptic site and the expression of NMDA receptors have been described as a major mechanism of brain colonisation by tumour cells (Zheng et al., 2019).

After crossing the BBB tumour cells release pro-survival molecules such as cytokines and chemokines, these include: CCL2, COX2, CXCL8, IFN-γ, IL-1β, and TNF-α, that allow the modification of the tumour microenvironment in support of tumour growth and proliferation by evading the microenvironment's immune attacks (Cacho-Diaz et al., 2020). Tumour cells also co-opt the astrocytes to modulate the immune metabolic response in the microenvironment to support and promote cancer cell survival (Perelroizen et al., 2022). The release of inflammation reducing cytokines such as TGFβ, IFNγ and IL-10 in the tumour's microenvironment seems to be mediated by interactions between the tumour-associated astrocytes and the microglial cells, which creates the immunosuppressive microenvironment required to protect tumour cells and promote their survival (Henrik Heiland et al., 2019). Experimental results have shown that increased secretion of TGFβ and IFNγ in the tumour microenvironment results in initiation of the JAK/STAT (The Janus kinase/signal transducer and activator

of transcription) pathway (Nicolas et al., 2013; Hu et al., 2021). Activation of the JAK/STAT pathway has been described in several tumour types, playing a major role in the regulation of pro-angiogenic and anti-apoptotic molecules, stem cell maintenance and suppression of immune response, all of which are necessary for tumour progression and cancer cell survival (Brooks and Putoczki, 2020; Ou et al., 2021). Furthermore, it has been demonstrated that the resistance to therapeutic agents acquired by metastatic brain tumours is mediated by the physical gap junction interaction between the astrocytes and tumour cells (Izraely and Witz, 2020).

Another molecule that supports the brain's colonisation by metastatic brain tumours is neurotrophin-3 (NT-3), where increases have been associated with decreased activation of the brain immune response by decreasing the activation of a cytotoxic microglial (Louie et al., 2013, Anaya-Ruiz et al., 2018)

All these findings show that the metastatic brain tumour cells' and non-tumour cells' interaction within the brain tumour microenvironment provides a protective pro-survival sanctuary which suggests that the brain secretes factors that promote tumour development. Identifying and understanding the mechanisms of action elicited by genomics in the regulation of the brain's resident cells and invasive tumour cells interactions may provide the necessary knowledge to develop an effective therapeutic approach as it will consider not only the anatomical restriction imposed by the blood brain barrier but also the particularities of the tumour microenvironment.

1.7 Aims and Objectives

The aim of this project was to identify the mechanisms involved in angiogenesis and breast to brain metastasis.

A) To identify the role of transcription factor ATF2 (molecule common to major pro-angiogenic growth factors) in the process of angiogenesis. ATF2 could be used a new therapeutic target to stop tumour angiogenesis.

• Investigate the role of ATF2 in Notch signalling pathway regulation of angiogenesis using Notch signalling pathway RT2-PCR array.

• Determine the expression of Notch pathway ligands in HUVECs lacking functional ATF2 using RT-PCR expression analysis.

• Assess the role of AFT2 in blood vessel development using organotypic co-culture assay.

B) Identify genomic alterations involved in the development, progression, and survival of breast to brain metastasis. This alteration may not have been involved in the early stages of the primary tumour but may have been essential for metastatic breast tumour cells to invade and colonise the brain.

• Identification of recurrent genomic alterations using Whole Exome Sequencing and bioinformatic filtering in 26 breasts to brain metastases that are infrequently found in primary breast tumours.

• Establishing the pathogenic potential of missense mutations using SIFT and Polyphen-2 protein consequence Insilico prediction tools.

• Investigating the role of BIG3 in breast to brain metastasis.

Chapter 2 Materials and Methods

2.1 Patients and samples

All tissue was provided by The Walton Research and Brain Tumour North West (BTNW) combined tissue bank. Ethical approval for tissue used in this project is covered by the research tissue bank's approval and local ethical approval has been granted (Ref: 14/EE/1270, North Wales REC: 11/WNo03/2). Informed consent was obtained from each patient. This study was conducted according to the principles expressed in the Declaration of Helsinki.

2.1.1 Breast to brain metastases (BBM) samples

Fresh frozen metastatic brain tumour samples that were derived from the breast were supplied by Brain Tumour North West in Preston and The Walton Research Tissue Bank (WRTB) in Liverpool. Extraction of genomic DNA was carried out for all tumours. The samples from breast to brain metastasis were designated as BBM. The list of BBM samples is shown in Table 2.1.

2.1.2 Matched primary breast tumours

The BTNW tissue bank in Preston provided eleven Formalin-Fixed Paraffin-Embedded (FFPE) matched primary breast tumours corresponding to metastatic brain tumours. The list of FFPE samples provided is given shown in Table 2.1.

Breast to brain metastasis (BBM)	Matched Primary	Matched blood
BM1	PB1	BL1
BM2	PB2	BL2
BM3	PB3	
BM5	PB5	BL5
BM7		BL7
BM8	PB8	BL8
BM12	PB12	BL12
BM14		
BM15	PB15	
BM16		
BM9		
BM20		
BM21		
BM22		
BM23		
BM24		
BM26		
BM27		
BM28		
BM29		
BM30		
BM31		BL31
BM32		
BM33		BL33
BM34		BL34
BM35		BL35

Table 2.1: BBM samples. 26 BBM samples that were Whole Exome Sequenced and their matched primary (PB), and patients' DNA from Blood. Only four samples had matched primary and blood DNA.

2.2 Tissue culture

A class II microbiology safety cabinet (Bio MAT 2, CAS, UK) was utilized to carry out all cell culture procedures. To prevent cell culture contamination, the safety cabinet was disinfected before each used with 70% ethanol and 1% Trigine solution purchased from Sigma-Aldrich UK. All the reagents needed to carry out tissue culture were first left to thaw at room temperature before using.

2.2.1 Breast cancer cell lines

Five breast cancer cell lines (MDA-MB231, MCF7, T47-D, ZR-75 and BT-549) were used for tissue culture practice and to perform experiments. All cell lines were provided by Prof. Weiguang Wang's group at the Research Institute of Healthcare Sciences, the University of Wolverhampton. Cells were maintained in DMEN (SLS, UK) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin from Sigma Aldrich at 37°C and 5% CO_2.

Cell line	Primary tumour	Cell origin
MDA-MB-231	Metastatic Adenocarcinoma	Metastasis (Pleural effusion)
MCF7	Adenocarcinoma	Metastasis (Pleural effusion)
T47-D	Ductal carcinoma	Metastasis (Pleural effusion)
ZR-75	Invasive ductal carcinoma	Malignant ascitic effusion
BT-549	Invasive ductal carcinoma	Metastasis (regional lymph nodes)

Table 2.2: Breast cancer cell lines. Breast cancer cell lines used to carry out tissue culture practice and functional analysis.

2.2.2 Human Umbilical Vein Endothelial Cells (HUVECs)

HUVECs were purchased from TCS cellworks. The cells were grown in tissue culture flasks pre-coated with 0.1% gelatine containing Endothelial cell growth media (ECGM, PromoCell UK) supplemented with 1% penicillin/streptomycin/amphotericin B from Sigma Aldrich and ECGM supplement mix (PromoCell UK) containing 2% FBS, 0.4% endothelial cell growth supplement, 0.1 ng/ml of recombinant human epidermal growth factor, 1 ng/ml recombinant human basic fibroblast growth factor, 90 μg/ml Heparin, and 1 μl/ml hydrocortisone. The cells with passages between 5 and 8 were used for experiments.

2.2.3 Human Dermal Fibroblast adult cells (HDFa)

HDFa cells were grown in tissue culture grade T75 flasks, previously coated with 0.1% gelatine using Dulbecco's Modified Eagle medium (DMEM) (Sigma-Aldrich, UK) supplemented 10% FBS, 1.73 mM L-glutamine, 1% penicillin/streptomycin/amphotericin B (Sigma-Aldrich, UK) HDFa were used to carried out the organotypic co-culture experiment.

2.2.4 Cell culture maintenance

Cells were consistently monitored for any changes in the cultured media and the cell population. Media was changed every 1-2 days or when there was colour alteration which is an indication of a drop in the media pH. Cells were cultured at 37ºC, 5% CO_2 until reaching 75-90% confluency.

2.2.5 Cell counting

The number of cells contained in a cell suspension was determined by placing 10 µl of cell suspension in a haemocytometer. Cells were counted under the microscope. The number of cells contained in 1 ml of cell suspension was determined by multiplying the number of cells counted under the microscope by 104.

2.2.6 Freezing cells

Cells were first trypsinized and centrifuged at 1200 rpm for 5 minutes. The supernatant was removed, and the pellet re-suspended in 1 ml of freezing medium solution (90% FBS, 10% DMSO). Cells were transferred into a labelled cryovial (Nalgene CryowareTM Labware, Roskilde, Denmark) and placed at -80ºC for a week and subsequently transferred into a liquid nitrogen tank (-196ºC) for long term storage.

2.2.7 Recovering cells from liquid nitrogen

Cells were stored in a cryovial in liquid nitrogen in a solution consisting of 10% dimethyl sulfoxide (DMSO) in foetal bovine calf serum (FBS). To recover the cells, these were removed from liquid nitrogen and left to thaw at 37°C. Cells were resuspended in a 1:10 cell to tissue culture medium ratio to avoid cell damage. Cells were centrifuged at 1200 rpm for 5 minutes, and the supernatant was cautiously removed. The cell pellet was re-suspended in fresh tissue culture media, transferred to a flask of suitable size, and cultured at 37°C, 5% CO_2.

2.2.8 Cell stimulation

For some experiments, cell stimulation was required. Prior to stimulation cells were washed with PBS 1x to entirely remove any remaining medium containing serum; cells

were then incubated in serum-free medium without any supplementary growth factors for the required time.

HUVECs were stimulated with VEGF 25 ng/ml (Peprotech). To prepare a VEGF stock solution, 10 µg lyophilized form of VEGF was dissolved in 400 µl of PBS 1x (Phosphate Buffer Saline), and then, 5µl of this stock solution was added into 5 ml of culture medium to obtain a final concentration of VEGF 25 ng/ml.

HUVECs were stimulated with bFGF 50 ng/ml (Peprotech). To prepare this stock solution, 10 µg of bFGF was added to PBS 1x (400 µl). A final working concentration of 50 ng/ml of bFGF was achieved by mixing 10 µl of stock bFGF solution into 5 ml of culture medium.

2.2.9 Transfection of HUVECs with small non-interfering RNA (siRNA)

HUVECs were cultured in a six-well plate pre-coated with 0.01% gelatin at a ratio of 2.5×10^5 cells per well and incubated overnight at 37ºC, 5% CO2. The following day, medium was removed, and wells were washed twice with 1% PBS. Cells were then incubated for 1 hour at 37ºC, 5% CO2 with 5 ml/well of opti-MEM serum-free medium. Transfection precipitates were prepared as follows: two solutions were prepared independently. Solution A containing 4.5 µl (20 µmol) of the corresponding si-RNA (non-target control or targeting human ATF2) plus 245.5 µl of opti-MEM, and solution B consisting of 5 µl Lipofectamine transfection reagent (Thermo Fisher Scientific) plus 245 µl of opti-MEM. Following this, Solution B was mixed dropwise with the corresponding solution A, and the mixture incubated in a dark tissue culture cabinet at room temperature for 20 minutes. Following incubation, HUVECs were transfected with the corresponding precipitate and incubated for 6 hours at 37°C, 5% CO2. After incubation opti-MEM plus transfection solution was removed from the wells. 5 ml of

ECGM was added to each well, and cells were incubated at 37°C, 5% CO2 for 48 hours. After 48 hours, ECGM medium was removed, cells were washed with 1x PBS, serum-starved, and stimulated with pro-angiogenic molecules.

2.2.10 Transfection of endothelial cells using replication-deficient adenoviral vectors

2.2.10.1 Adenoviruses used in this study

To suppress the functionality of ATF2 in primary endothelial cells, HUVECs were infected with adenovirus Ad-ATF2(AA). This adenovirus encodes a functionally inactive mutant version of the ATF2 protein where potential phosphorylation of amino acids threonine 69 and threonine 71 has been disrupted by mutation to alanine (Figure 3.1). Adenovirus Ad-GFP encodes the green fluorescence protein and was used as a control (Figure 2.1).

Both adenoviruses were provided by Dr Wolfgang Breitwieser (CRUK-Manchester Institute). Generation and use of these adenoviruses have been previously described by Gozdecka et al. (2014).
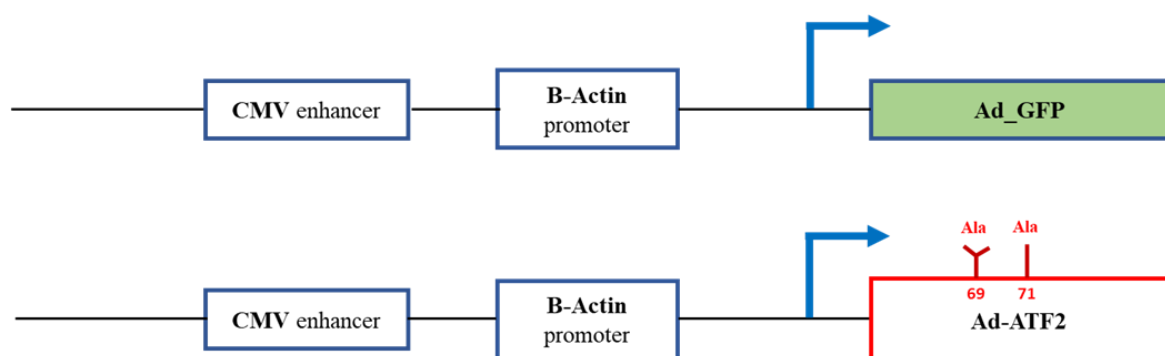
Figure 2.1: Schematic representation of the main features of the Adenoviruses used for this study. AD_ATF7_AA amino acid residues 69 and 71 have been mutated from threonine to alanine.

2.3 Genomic DNA isolation

Genomic DNA was extracted from fresh frozen breast to brain metastasis, from FFEP tissue and from patient's blood that had been collected at the time of the brain tumour surgery.

2.3.1 Genomic DNA isolation from frozen breast to brain metastatic tumours

Genomic DNA was isolated from fresh frozen metastatic brain tumours using the DNA isolation kit for cells and tissues version 08 (Roche, Germany). All centrifugation steps were carried out at room temperature (between 15-25°C). 80 mg of fresh frozen tissue was ground using a mortar and pestle in dry ice. Subsequently, 2 ml of cellular lysis buffer was added to a sterile centrifuge tube. Samples were homogenized using a syringe until a fine suspension was obtained. 2 μl of proteinase K solution was added to each sample, vortexed, and incubated at 65°C for 1 hour to ensure degradation of proteins. 200 μl of RNase solution was added to each sample, vortexed to mix the solution into suspension, and incubated at 37°C for 15 minutes to degrade the samples RNA. Protein precipitation was achieved by adding 840 μl of protein precipitation solution to each sample. Samples were vortexed thoroughly for 5-10 seconds to precipitate the proteins, placed on ice for 5 minutes to help protein precipitation, and subsequently centrifuged at 26,900 x g for 20 minutes at room temperature. The supernatant containing the DNA was transferred into a new sterile centrifuge tube, 0.7 volumes of isopropanol were added to the samples to help DNA precipitation. Samples were centrifuged a 1,370 x g for 10 minutes. 5 ml of ice-cold 70% ethanol was added to the samples to wash precipitated DNA. Samples were centrifuged at 1,370 x g for 5 minutes and the supernatant was discarded. The DNA pellet was air-dried until the ethanol was fully evaporated. An appropriate amount of water (200-400 μl) was added

to each sample to resuspend the DNA and left overnight at +4°C. Nanodrop 2000 spectrophotometer (Thermo Scientific, USA) was used to quantify the gDNA.

2.3.2 Genomic DNA isolation from blood

DNA was isolated from fresh-frozen blood samples using the DNA isolation kit for mammalian blood version 08 (Roche, Germany). All centrifugation steps were carried out at room temperature (15-250C). 2 ml of each defrosted blood sample was added into 20 ml centrifuge tube containing 6 ml of red blood lysis buffer and mixed gently by inversion. The tube containing the blood-red blood cells lysis buffer solution was placed in a rocking plate for 10 minutes and then centrifuged at 875 x g for 10 minutes resulting in two layers: white cells pellet in the bottom of the tube and a clear, red supernatant at the top of the tube. The clear, red supernatant was disposed of, and the white cell pellet was thoroughly vortexed. 2 ml of white cell lysis buffer was added and vortexed to completely lyse the white cells. To ensure successful lysis of the leukocytes, 30 minutes incubation at 37°C was carried out; resulting in a clear dark red/brown solution with no particle material present. The samples were transferred to a clean sterile centrifuge tube and 520 µl of protein precipitation solution was added and vortexed thoroughly for 25 seconds, then centrifuged at 12,000 x g for 10 minutes. The resulting supernatant, which contained the DNA, was carefully poured into a new clean sterile centrifuge tube. For DNA precipitation, 2 volumes of +15-25°C 100% ethanol were added to the supernatant. The solution was gently mixed by inversion until DNA strands precipitated out of the solution and the remaining liquid was no longer cloudy. Samples were centrifuged at 875 x g for 10 minutes and the resulting supernatant was discarded. DNA wash was carried out by adding 1.5 ml of cold 70% ethanol to the DNA pellet, samples were mixed several times by gentle inversion and then centrifuged at 875 x g for 5 minutes; the supernatant was discarded. The DNA

pellet was left to air dry under a tissue culture cabinet until the ethanol was completely evaporated. DNA pellets were resuspended in 250 µl of RNase/nuclease-free water, vortexed thoroughly and then placed in a heating block at 65°C for 30 minutes. Samples were stored at -20°C.

2.3.3 Purification of genomic DNA from formalin-fixed, paraffin-embedded tissues (FFPE)

Genomic DNA was isolated from FFPE samples using the QIAamp DNA FFPE tissue kit (QIAGEN, UK). All centrifugation steps were carried out at room temperature (15-25°C). Using a scalpel, excess paraffin was cut off the samples. Each sample was placed into 1.5 ml microcentrifuge tube and 1 ml of xylene was added; samples were vortexed vigorously for 10 seconds and centrifuged at 13,000 rpm for 2 minutes. The resulting supernatant was removed by pipetting without removing any of the pellet. 1 ml of 100% ethanol was added to the pellet and mixed by vortexing. Samples were subsequently centrifuged at 13,000 rpm for 2 minutes. The supernatant was removed, conserving the pellet. With the tube open the samples were incubated at 37°C for 10 minutes or until all residual ethanol had evaporated. The pellets were resuspended in 180 µl of buffer ATL, 20 µl of proteinase K was added to each sample and mixed by vortexing. Samples were incubated at 56°C for 1 hour (or until the sample had been completely lysed) and after that incubated at 96°C for another hour. Afterwards, samples were briefly centrifuged and left to cool down before adding 2 µl of RNase A and incubated for 2 minutes at room temperature. 200 µl of buffer AL and 200 µl of 100% ethanol per sample were premixed; 400 µl of the mix was added to each sample and mixed immediately by vortexing. The entire lysate was transferred to QIAamp MinElute column and centrifuged at 6000 x g for 1 minute. The columns were then transferred to a clean 2 ml collection tube and discarded the second collection tube

containing the flow-through. Then, 500 µl of AW1 buffer was added to each sample and centrifuged at 6000 x g for 1 minute. QIAamp columns were transferred to clean 2 ml collection tubes and flow-through discarded. 500 µl of buffer AW2 was added to each sample and centrifuged for 1 minute at 6000 x g; flow-through was discarded and columns were placed in clean 2 ml collection tubes. An extra centrifugation step at 13,000 rpm for 3 minutes was carried out to dry the membrane completely. To elute the DNA each QIAamp column was placed into a clean 1.5 ml microcentrifuge tube, an adequate amount of RNase/Nuclease free water was added to the centre of each column and centrifuged at 13,000 rpm for 1 minute. DNA was stored at -20°C for further use.

2.4 Whole Exome Sequencing

2.4.1 Validation of DNA quality

Genomic DNA from 26 fresh frozen tissues or FFPE BBMs was sent to the Institute of Cancer and Genomics at Birmingham University. The quality and concentration of the DNA were determined using the Genomic DNA Screen Tape Analysis (Agilent UK) following the manufacturer's specifications. The test generates a DNA Integrity Number (DIN) which is a representation of quality. Prepared samples were loaded in the 2200 TapeStation instrument to be run and analysed. DIN calculation was performed using the Agilent 2200 TapeStation analysis software (Software version A.02.02). The software assessed the signal distribution across the size range to determine the grade of DNA fragmentation. The algorithm generated a number as a measure of quality. The DIN mathematical assessment applied values from 1 to 10 where a high value indicated highly intact DNA and a low value implied that the sample was degraded. All BBM samples used for WES had DIN ≥ 6.4.

2.4.2 Library preparation and Whole Exome Sequencing

Samples' exons were captured using the TruSeq Exome Library prep kit (Illumina) using the XGen exome probes from Integra DAN Technologies. Sequencing was carried out on a high output V2 150 cycle flowcell (Illumina) as a 75 paired (2 x 75) and run on the NextSeq 500 sequencer from Illumina with 1% PhiX spike-in (PhiX Control v3 Library, Illumina).

2.4.3 WES data analysis

The data generated by WES alignment, variant calling and cohort analysis was performed by Dr. Andrew Beggs at the Beggs Laboratory, Institute of Cancer and Genomic Science at the University of Birmingham, UK, using the Illumina BaseSapce Hub. Sequenced reads from all samples were aligned to the human genome reference sequence (reference number GRCh38) using BWA 0.7.x bioinformatic tool. Variant calling of somatic single nucleotide variants (SNVs) was done using VarScan2 applying the automatic settings. Variant Alle Frequency (VAF) was designated for all tumours. Following variant calling, the WES data file was converted into an excel text file and a strict filtering method was followed to identify somatic and pathogenic variants retaining only nonsynonymous SNVs (Missense, nonsense, and splice site) and indels were selected. To avoid potential germline polymorphism, variants with a minor allele frequency (MAF) > 0.1% and reported in the Exome Aggregation consortium (ExAC) (currently available at https://gnomad.broadinstitute.org/) or the dbSNP (https://www.ncbi.nlm.nih.gov/snp/) were eliminated. Variants with a quality score below zero were eliminated. Variants with a MAF ≤ 0.1% were retained. Variants not reported on the formerly mentioned databases were also retained. The pathogenic

impact of missense mutations was predicted using the Polyphen-2 tool (http://genetics.bwh.harvard.edu/pph2/index.shtml).

2.5 PCR

2.5.1 PCR Primer design

PCR primers for Sanger sequencing validation of mutations identified by Whole Exome Sequencing, for expression analysis of the genes of interest were designed from the genes' primary transcript. The exon sequenced of candidate genes was retrieve form Ensembl Genome Browser (http://www.ensembl.org/index.html). For Sanger sequencing validation primers were designed to amplify the specific exome region containing the mutation of interest. Primers are listed in Appendix A1

2.5.2 RT-PCR primer design

For RT-PCR expression analysis primers were designed encompassing more than one exome to make sure that the amplification product did not result from any contaminating gDNA. All primers were designed manually containing between 18 to 25 base pairs. Even distribution of G/C content of 40 to 60% for optimal primer template annealing. The annealing temperature of each primer was calculated using the equation $[(\% \, CG \times 0.41) + 64.9 - (600/N)]$ where the %CG is calculated by dividing the total number of C and G content in the primer by the total primer length (N). All designed primers are listed in Appendix A.2.

2.5.3 PCR sample amplification

Following primer design, PCR amplification of relevant BBMs' exons was performed. The samples were run for 35 cycles in a thermal cycler, using primer dependent annealing temperature between 57°C and 60°C, and touchdown. Then, 5 µl of PCR

product was cleaned up using 2 µl of ExoSAP-IT (Thermo Fisher); a PCR enzymatic cleaning reagent used to remove the excess of primers and nucleotides. The reaction was incubated for 15 minutes at 37°C to degrade the remaining primers and nucleotides, and 5 minutes at 80°C to inactivate ExoSAP-IT activity. ExoSAP-IT purified samples are prepared to use in Sanger sequencing.

2.5.4 Sanger sequencing validation of WES results

Following PCR amplification and purification of breast to brain metastasis tumour DNA samples, validation of gene mutations identified by WES using conventional dye-termination Sanger sequencing the BigDye™ Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems) was carried out. 2 µl of purified DNA was used in combination with 0.5 µl of BigDye, 2µl of 5x BigDye buffer, 0.5 µl of forward or reverse primers in two separate reactions. Samples were placed in the thermal cycler for 2 minutes held at 96°C and 25 cycles of 10 seconds at 96°C, 15 seconds at 52°C, and 3 minutes at 60°C. Cycle sequencing products were cleaned up using the BigDye XTerminator™ Purification Kit (Applied Biosystems). 22.5 µl/sample of SAM solution and 5 µl/sample of XTerminator™ solution were added to the cycle sequencing DNA solution; samples were vortexed for 10 minutes and centrifuged for 2 minutes at 1000g.

2.6 Gene knockout using CRISPR

2.6.1 Recovering cells from liquid Nitrogen

Breast cancer cell lines (MDA-MB231, MCF7, T47-D, ZR-75 and BT-549) were collected from liquid nitrogen, left to thaw at 37°C and then quickly transferred to T-75 vented flask containing 19 ml DMEM (serum-containing medium). The cells were incubated overnight at 37°C. This was followed by changing media the next day to

ensure that the trace amounts of DMSO were removed and no unattached or dead cells were remaining.

2.6.2 Trypsinization of healthy adherent cells

For cell subculturing, the consumed media was aspirated, followed by washing with 5 ml sterile PBS. This was followed by addition of 2 ml trypsin from working stock and the cells were incubated at 37°C for 5 minutes. The flask was checked regularly and lightly tapped from the sides to gently detach the cells. After 5 minutes the cells were examined under the microscope. To confirm detachment from the flask. Trypsin solution was neutralised using 2 ml serum-containing medium which was added to the flask. The solution was mixed thoroughly in the flask to break any clumps left after trypsinization. The mixture was then transferred to a sterile tube and centrifuged at 1.2 rpm for 5 minutes; the pellet was collected and resuspended into fresh medium for subculturing.

2.6.3 Maintenance of cell lines

The cells were observed under the microscope to determine if they were confluent. The cells were subcultured when confluence of 75% to 90% was reached. The cells were again trypsinized, collected as a pellet, counted, and added accordingly to the flask.

2.6.4 CRISPR lentivector set

All-in-One lentivector sets of sgRNA CRISPR/Cas9 for each our genes of interest KIAA1244/BIG3 were purchased from abm goods (Canada). Each lentiviral set consists of multi-guide sgRNA that target the same gene but will introduce three fragment deletion in the DNA of the target gene.

Genes of interest	Vector	Target Sequences
BIG3/ ARFGEF3	pLenti-U6-sgRNA-SFFV-Cas9-2A-Puro	**T1: 306** AGTGACGCCTTCGCTCAACG **T2: 399** TGCCGTGCTGAAGATCGCGG **T3: 567** CTTGATTCCCGAAATCCTGT

Table 2.3: Lentiviral vector used for CRISPR cas9 gene knockout of BIG3 and three targeted DNA sequences**.**

2.7 Bacterial transformation

2.7.1 Preparation of ampicillin (Amp+) containing agar plates and LB broth

50 g of LB (Luria-Bertani) were dissolved in 2 litres of deionised H2O. 5 aliquots of 400 ml of LB were made; 6 g of agar (Fluka Biochemika) were added into aliquots and then autoclaved. Ampicillin was added to the LB broth-agar mix to a final concentration of 100 µg/ml. 25 ml of Amp+ LB broth-agar solution was added to sterile Petri dishes, left to cool down and set, and then stored at +4°C for further use.

2.7.2 Stable Transformation

20 ng of each plasmid (T1, T2, T3) DNA was mixed with 50 µl of JM109 competent cells (Promega, USA). The mix was placed on ice for 20 minutes followed by 1 and a half minutes incubation at 42°C and then placed back into ice for another 2 minutes. Subsequently, 950 µl of LB broth without ampicillin was added to the mix and left in a shaking incubator at 37°C for 1 hour. After incubation was finished, 150 µl of each target was spread in Amp+ LB broth-agar plates; the plate was left overnight at 37°C.

2.7.3 Plasmid purification

From the LB broth-agar-amp+ Petri plates, one colony per target was picked. Each colony was placed into a sterile 20 ml tube containing 3 ml of LB broth and ampicillin to a final concentration of 100 µg/ml. The tubes were left overnight in the shaking incubator at 37°C. The following day the solution in each tube was divided and stored at -20°C for further use.

2.7.4 Mini-Prep

Mini-prep was performed using the QIAprep Spin Miniprep Kit (QIAGEN, UK). The tubes containing 1 ml of the transformed plasmid were centrifuged at 8000 rpms for 3 minutes at room temperature. The supernatant was discarded, and the bacterial pellet was resuspended into 250 µl of buffer P1. The mix was then transferred into a microcentrifuge tube and 250 µl of buffer P2 was added, the sample was mixed 4-6 times by inversion until the solution became clear. After that, 350 µl of buffer N3 was added and mixed by inverting the tube 4-6 times. The solution was centrifuged at 13,000 rpm for 10 minutes. The supernatant was transferred into a QIAprep spin column by pipetting, after one minute centrifugation flow-through was discarded. QIAprep column was washed with buffer PB, centrifuged for one minute and flow-

through was discarded. The column was washed a second time using 750 µl of buffer PE, centrifuged for 1 minute and flow-through discarded. An additional 1 minute centrifugation step was performed to remove any residual wash buffer. The column was placed into a sterile 1.5 microcentrifuge tube, to elute the DNA 50 µl of buffer EB (PH 8.5) was added to the centre of the column, left to stand for 1 minute and then centrifuged for 1 minute. Eluted DNA was quantified by nanodrop and immediately after stored at -20°C for further use.

2.7.5 Maxi-Prep

1 ml from each purified plasmid was inoculated into 9 ml of LB-ampicillin (100 µg/ml) medium and incubated overnight at 37°C in a rotary shaker incubator. The next day 1;10 dilution was made in a fresh LB-ampicillin medium and incubated for 6 hours at 37°C in the rotary shaker incubator. After incubation, the culture was inoculated into 400 ml of LB-ampicillin and incubated overnight at 37°C in the rotary shaker incubator. The following day, the cultures were transferred into sterile plastic 500 ml maxi-prep bottles and centrifuged at 4,500 rpm for 15 minutes; the supernatant was discarded, and the pellet was resuspended in 10 ml of buffer P1 (QIAGEN Plasmid Maxi Kit). The solution was transferred into 50 ml maxi-prep tubes, and 10 ml of buffer P2 was added and mixed by inversion. Subsequently, 10 ml of buffer P3 was added and mixed by inversion. The solution was immediately incubated on ice for 20 minutes and after that centrifuged at 9,600 rpm for 30 minutes. In the meantime, QIAamp maxi-prep columns were equilibrated with 10 ml of buffer QBT. After centrifugation, the resulting supernatant was transferred into the maxi-prep columns, allowing flow-through. After that, columns were washed twice with 30 ml of buffer QC. To elute the DNA, the columns were transferred into new 50 ml centrifuge tubes; 15 ml of buffer QF was added to each column and allow flow-through. DNA was precipitated by adding 10.3

ml of isopropanol at room temperature, the solution was mixed by inversion and placed in the centrifuge at 9,600 rpm for 1 hour. The supernatant was carefully poured off and 1.5 ml of 70% ethanol at room temperature was added to resuspend the pellet, then centrifuged at 13,000 rpm for 10 minutes. The supernatant was removed, and tubes were centrifuged one more time for 1 minute to remove residual ethanol. The pellets were left to air dry on a tissue culture hood for 5-10 minutes or until the ethanol has completely evaporated. After being completely dried, pellets were resuspended in an appropriate amount of water. DNA was quantified by nanodrop, and the 3 targets were aliquoted to a concentration of 3 µg/µl and immediately after stored at -20°C for further uses.

2.7.6 Restriction digest

The obtained plasmid DNA was digested using plasmid-specific restriction enzymes (Promega, USA) and incubated overnight. 6 µl of digested plasmid was mixed with 6 µl of 2X loading buffer and run in a 0.7% agarose gel.

Reagents	Concentration	Volume used
Transformed Plasmid	100ng/ul	10ul
10x NE buffer H	10X	3ul
EcoR1	12 u/µl	2ul
NOT1	10 u/µl	2ul
H_2O	-	10ul
Total	-	30ul

Table 2.4: Components of enzymatic restriction digestion.

2.7.7 Stable transfection by lipofection of MCF7cell line

Before transfection was carried out, mRNA expression of the BIG3 gene in MCF7 breast cancer cell line was carried out using RT-PCR. RT. The sequence of the RT-PCR primers used are shown in Appendix A.1.

2.7.8 Plating cells

400,000 MCF-7 per well were seeded in tissue culture grade six-well plates containing 3 ml of DMEM supplemented with 10% FBS and left overnight at 37°C.

2.7.9 Preparing transfection solutions

Each target (T1, T2, T3) was diluted to a concentration of 3 µg of plasmid DNA per microliter and 10 µl of each was added in a 1.5 ml microcentrifuge tube. Transfection solutions were prepared as follows: solution A; 250 µl/well of serum-free media (SFM) was added into 3 sterile 15 ml universal tubes and mixed with 10 µl/well of Lipofectamine 2000 transfection reagent was added to each tube, mixed thoroughly, and incubated for 15 minutes. Solution B: in 3 separate 15 ml sterile universal tubes 250 µl/well of SFM were added, into one tube 3 µl/well of plasmid DNA at a concentration of 3 µg/µl were mixed with the media. In another tube 6 µl/well of plasmid DNA at a concentration of 3 µg/µl were mixed with the media. Into the third tube 3 µl/well of a scrambled plasmid (empty plasmid) at a concentration of 2 µg/µl was mixed with DMEM media. The negative control was prepared by mixing 250 µl/well of SFM with 10 µl/well of Lipofectamine 2000. Solution A was mixed with its corresponding solution B to create a DNA-lipid complex. Complexes were incubated at room temperature in a tissue culture cabinet for 20 minutes. After incubation, the mixture was transferred into six-well plates containing cells and incubated at 37°C for 48 hours.

After the incubation period, cells were trypsinised with 500 µl of trypsin and neutralised with 1.5 ml of serum-containing media. Cells were transferred into 21 ml tissue culture grade Petri dishes containing 15 ml of DMEN media plus FBS. Puromycin, the selective antibiotic for the plasmid was added at a final concentration of 200 ng/ml. Media was changed every 4 to 5 days; cells were kept in the Petri dishes until individual colonies started to form. Once the colonies were formed, they were picked and individually transferred into T-25 tissue culture flasks. Colonies were kept in the flasks until they were 70 to 90 per cent confluent and then split into two tissue culture flasks where they were left to grow until they were 70 to 90 per cent confluent. Cells were maintained, changing media every 4 to 5 days until they reached 70 to 90 per cent confluence, to subsequently subculture for RNA and protein extraction for further downstream applications.

2.8 Quantification of gene expression

2.8.1 RNA isolation

RNA was extracted from cultured cells using the total RNA purification kit from NORGEN BIOTEK CORP. After growing the CRISPR-modified cells in a T-75 flask until they were 75 to 90 per cent confluent, cells were trypsinized using 1x 500 µl trypsin to detach them from the flask and mixed with 1 ml DMEM immediately to inactivate the trypsin. The cell suspension was centrifuged at 1200 rpm for 5 minutes. The cell pellet was resuspended in PBS and re-centrifuged. To lysate, the cells 350 µl of Buffer RL was added to the pellet and let it stand for 5 minutes at room temperature. The lysate was transferred into a microcentrifuge Eppendorf tube to which 200 µl of 100% ethanol was added, then mixed by vortexing for 10 seconds.

The lysate was transferred into the spin columns assembled with the collection tubes and centrifuged for 1 minute at 3,500 x g. After making sure the entire lysate has passed through flow-through was discarded. To wash the column 400 μl of wash solution A was applied to the column containing the RNA. The column was centrifuged for 1 minute at 3,500 x g and flow-through was discarded. This step was repeated 2 additional times. To elute the RNA the column was placed into a 1.7 sterile elution tube and an appropriate amount of elution solution A was added to the column and centrifuged for 2 minutes at 200 x g, followed by 1 minute at 13,000 x g. Total RNA concentration was measured using a Nanodrop2000 Spectrophotometer (Thermo Scientific, USA). Purified RNA was stored at -80°C.

For HUVECs, RNA isolation was carried using a "Total Plus RNA Purification Kit" (Norgen) according to the manufacturer protocol. HUVEC cells cultured on a six-well plate were washed twice with 1x PBS. PBS was removed and 300 μl of lysis solution was added to each well. The plates were tilted from side to side to cover the entire well and the lysate containing the components of the cell along with the cell debris was completely transferred to an Eppendorf tube. Samples were stored at -80°C until the RNA purification protocol was carried out. Before RNA purification, DNA was removed, samples were placed in gDNA removal columns and centrifuged at 10,000 rpm for 1 minute. Flow-through was retained for RNA extraction. For RNA purification, samples were treated with 200 μl of 100% ethanol and vortexed briefly for approximately 10 seconds at room temperature. The solution was passed through a spin-column by centrifugation at 13000 rpm for 60 seconds at room temperature, flow-through was discarded from the collection tubes. 400 μl of washed solution was added to the column-containing sample and centrifuged for 60 seconds at 13,000 rpm, flow-through was discarded. This step was repeated three times, the last centrifugation was carried

out for 2 minutes to thoroughly dry the membrane. To elute the RNA, the collection tube was discarded, and the column was placed into a 1.5 ml elution tube provided with the kit. 40 µl of elution solution was added to the column left to rest at room temperature for 60 seconds. After incubation, the sample was centrifuged at 2,000 rpm for 2 minutes followed by 60 seconds at 10,000 rpm. Eluted RNA was placed at -80ºC for long term storage.

2.8.2 RNA quantification

Concentration and quality of each sample were determined using a Nanodrop 2000 spectrophotometer (Thermo Scientific UK). RNA samples' purity was assessed by the ratio of absorbance at 260 nm and 280 nm.

2.8.3 cDNA synthesis

All cDNA was synthesised after RNA quantification using High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher). Samples were diluted to 500 ng of total RNA in a final volume of 10 µl with nuclease-free water (Invitrogen, UK). 10 µl of total RNA is mixed with 10 µl of a Reverse Transcription (RT) mixture containing 10x Transcription Buffer, 25x dNTP mix (100 mM, 0.2 ml), 10x Random Primers, MultiScribe reverse transcriptase (0.2 ml at 50 U/µl), RNase Inhibitor (2 × 0.1 ml at 20 U/µl). This RT master mix was mixed with each sample to make a final volume of 20 µl. Reverse transcription was carried out in a thermal cycler using the following conditions: 25°C for 10 minutes, 37°C for 120 minutes, 85°C for 5 minutes. After this time, 80 µl of nuclease-free water was added to the reaction. The cDNA was placed on ice for immediate use or stored at -20ºC for long term storage. Reaction components and volumes are summarised in Table 2.5.

Components	Volume/µl
10x Buffer	2.0
25x dNTPS	0.8
10x Random Primers	2.0
MultiScribe Reverse Transcriptase	1.0
RNase Inhibitor	1.0
Nuclease Free Water	3.2

Table 2.5: Components of the "High-Capacity cDNA Reverse Transcription Kit" and volumes per sample used to prepare the RT master mix solution.

2.8.4 Reverse transcription (RT) PCR

Gene expression was determined by RT-PCR. RT-PCR mix was prepared by mixing 1 µl of the desired prepared cDNA with a 25 µl volume RT-PCR reaction containing 2.5µl 10X PCR buffer containing $MgCl2$ (pH 8.3), 2.5mM dNTPs mix, 1.25 µM forward primer, 1.25 µl reverse primer (Primer design previously described in section 3.5.2), 0.5U Fast Start Taq DNA polymerase (Roche, Germany), and 16.3 µl of deionised water. RT-PCR reactions were carried out using a touchdown PCR. The β-actin and GAPDH genes were used as a positive housekeeping control. The same thermocycler conditions were used to assess gene expression levels in the control and the samples being used, except for the number of cycles (20 for control, 28 for samples). RT-PCR primers used are shown in Appendix A.2.

2.8.5 Notch signalling gene array

2.8.5.1 RT2 PCR Array

Expression of genes related to the Notch signalling pathway in endothelial cells lacking functional ATF2 was determined by PCR-based screening of a Notch Signalling Plus gene array (Reference PAHS-059Y, QIAGEN). PCR components mix was prepared as indicated in Table 2.6, and 25 µl of this mixture was dispensed into each well of the array plate. The plate was carefully sealed with MicroAmp® Optical Adhesive Film (Thermo Fisher) centrifuged for 2 minutes at 900 rpm at room temperature. The array plate was then placed into a 7500 Fast real-time PCR cycler (Applied Biosystems UK). The conditions at which the cycler was set are shown in Table 2.7.

Components	96 well array plate
2x RT^2 SYBR green master mix	1350 µl
cDNA synthesis reaction	102 µl
RNase free water	1248 µl
Total volume	2700 µl

Table 2.6: Components and volumes used to prepare- PCR mixture used for screening gene expression in RT2 PCR Array plates.

Components	Duration	Temperature	Description
1	10 minutes	95 ^{0}C	Hot start DNA taq is activated in this step
40	15 seconds 60 seconds	95 ^{0}C 60 ^{0}C	Perform fluorescen data collection

Table 2.7: Real-time PCR cycler settings used to screen RT2 Notch Signalling Plus gene array. Excess volume of 300 µl, including 9 µl of cDNA reaction to perform quality control analysis.

2.8.6 Real-time PCR

PCR reaction was prepared for each primer according to the instructions shown in Table 2.8. 14.4 µl of TaqMan reaction master mix and 5.6 µl of cDNA (20 µl total volume per well) were added to a fast optical 96-well reaction plate (Applied Biosystems UK). An optical adhesive film was used to seal the plate. Subsequently, the plate was centrifuged for 1 minute at 900 rpm and loaded into a 7500 Fast real-time PCR cycler (Applied Biosystems UK). The machine was set up for amplification using a program consisting of 1 holding stage at 95°C for 10 minutes for enzyme activation, followed by 40 cycles of denaturation at 95°C for 15 seconds and annealing/extension at 60°C for 60 seconds. Following PCR amplification, raw data was exported into an excel document to be analysed. To determine fold change, the Ct value of each sample was normalised according to the Ct value of Hprt-1 (housekeeping gene). The TaqMan primers used for this study were purchased for Applied Biosystems UK and are shown in Table 2.9.

Reagents	1x
TaqMan primer	1.0µl
TaqMan Universal PCR master mix (2X)	10µl
Nuclease free water	3.4µl
Total	14.4µl

Table 2.8: Components and volumes used to prepare PCR master mix to carried out TaqMan real-time PCR assays.

TaqMan Primers	ID Assay
Delta like ligand 1 (DLL1)	Hs00194509_m1
Delta like ligand 4 (DLL4)	Hs00184092_m1
Peroxisome Proliferator-Activated Receptor Gamma (PPARG)	Hs00234592_m1
Snail Family Transcription Repressor 2 (SNAI2)	Hs00161904_m1
Hypoxanthine Phosphoribosyltransferase 1 (HPRT1)	Hs99999909_m1

Table 2.9: TaqMan primers and their reference code used to carry out real-time PCR in this study.

2.9 Tubule formation assay

2.9.1 Organotypic co-cultures of adult Human Dermal Fibroblasts (HDFα) and GFP-HUVECs

HDFα purchased from TC Cell Works were cultured in tissue culture flasks containing Dulbecco's Modified Eagles Medium supplemented with 10% FBS, 5 ml of L-Glutamine and 5 ml 100x Penicillin/Streptomycin/Amphotericin (Sigma-Aldrich) (DMEM complete) until reaching the desirable confluency. Cells were then subcultured and 30,000 cells per well plated into a 12 well plate containing DMEM complete medium and incubated for 72 hours so they reach confluency. Then, 30,000 GFP-HUVECs that had been previously infected with Ad-ATF2AA (or control Ad-GFP) were plated on top of the HDFα in 1 ml of "Angiogenesis Seeding Medium" (Calteq Ltd). Cultures were incubated overnight and the following day the culture medium was replaced by "Angiogenesis Growth Medium" (Calteq Ltd) with or without stimulus and cells incubated at 37°C, 5% CO_2 for 13 days, with the medium (with or without stimulus) being replaced every 2-3 days.

2.9.2 Tube fixing and staining with CD31 antibody

To visualize tube formation, on day 13 of the experiment, cells were carefully washed three times with 0.5 ml of 1x PBS. Following, 0.5 ml of ice-cold fixative (70% ethanol) was added to each plate. Plates were incubated at room temperature for 30 minutes. After incubation, a fixative solution was discarded and wells were washed three times with blocking buffer (1x PBS, 1% BSA).

For staining the endothelial cells forming tubular structures, a mouse anti-human CD31 antibody (ab24590, Abcam) was diluted 1:400 in blocking buffer. 0.5 ml of diluted antibody solution was then added to each well and incubated for 60 minutes at

37°C, 5% CO_2. After incubation, the primary antibody was removed, each well was washed with 0.5 ml of blocking buffer and incubated at room temperature for 5 minutes (repeated three times). Before the final wash, a solution of secondary antibody was prepared by diluting 1:500 a goat anti-mouse IgG alkaline phosphatase-conjugated antibody (ad97237, Abcam) in blocking buffer. After the final wash, cells were washed three times with 0.5 ml of dH2O (deionised water) following the procedure for washing off the primary antibody.

For staining, an insoluble substrate (BCIP/NBT) was prepared by dissolving two BCIP/NBT ready to use tablets (Sigma Aldrich) into 20 ml of dH2O. The substrate was filtered using a syringe and a 0.2 µm filter disc. 0.5 ml of the substrate solution was added to each well and plates were incubated at room temperature until tubules developed a dark purple colour (within 3-10 minutes). Wells were carefully washed once with dH2O to completely remove the substrate. Pictures were taken immediately after this step. Plates can be stored in the dark indefinitely, but the colour intensity will fade over time. Images were taken with an EVOS microscope at 4x magnification. Tubular structure quantification was carried out using imageJ software with the macros plugin angiogenesis analyser (Carpentier et al., 2012). Quantification was performed by analysing 2 random experiment fields in each independent experiments. Number of tubes in each independent experiment was calculated as the average of two random fields. Six independent experiments were quantified to determine statistical significance differences between the groups.

2.10 Statistical analysis

Angiogenesis experiment results are expressed as mean ± SEM. Statistical differences between the two groups were analysed by unpaired two-tailed Student's t-test. One-Way ANOVA with Post hoc Turkey's comparison test was used to analyse differences between more than two groups. Statistical significance was set at $p \leq 0.05$.

Chapter 3 Role of the transcription factor ATF2 in the regulation of angiogenic gene expression in VEGF-Stimulated endothelial cells

3.1 Introduction

Angiogenesis takes place throughout the body during physiological and pathological states to fulfil the organs' demand for oxygen and nutrients (Lugano et al., 2020). Several pro-angiogenic molecules are overexpressed during tumour angiogenesis, including the VEGF family of growth factors (Holmes et.al., 2010). Because upregulation of VEGF ligands has been found in most solid tumours (Vasudev and Reynolds, 2014), inhibition of VEGF is believed to be effective to counteract tumour angiogenesis (Carmeliet, 2005). Several therapeutic agents have been developed to target VEGF signalling to decrease angiogenesis in cancer and other inflammatory diseases (Pożarowska and Pożarowski, 2016; Khanna et al., 2019). However, this approach is not always efficient for tumour treatment as the development of resistance to anti-VEGF therapy is a frequent outcome (Kim et.al., 2018). This inability of achieving sustained inhibition of tumour angiogenesis has created the necessity to explore new ways of targeting the formation of new blood vessels in tumours.

VEGF-A downstream signalling that results in the stimulation of new blood vessels is still not entirely understood. Characterization of the VEGF-A/VEGFR2 axis in angiogenesis has been carried out by several research groups. These studies have found that ATF2 transcriptional activity is required for VEGF-A- dependent stimulated angiogenesis (Chung and Ferrara, 2011, Koch et al., 2011, Fearnley et al., 2014). Previously, it has been reported that ATF2 phosphorylation is regulated by VEGF in cardiac myocytes and endothelial cells (Salameh et.al., 2010). However, the mechanism behind the interaction between these two molecules has not been

completely characterized. It has been suggested that different VEGF-A isoforms will have different stimulatory effects on VEGFR2 causing a collective nuclear shift that involves ATF2 transcriptional regulation (Fearnley et al., 2014). The aim of this project is to characterise the role played by the transcription ATF2 in the angiogenic process.

3.1.1 Expression of Notch signalling pathway ligands in HUVEC cells lacking functional ATF2

To evaluate the role of ATF2 in angiogenic regulation, HUVECs were infected with Adenovirus expressing ATF2_AA as described in the materials and methods section. ATF2_AA is a dominant-negative version of ATF2 where ATF2 phosphorylation residues Thr69 and Thr71 have been mutated to Alanine (Ala) consequently lacking ATF2-dependent transcriptional activity (Gozdecka et al., 2014). HUVECs infected with an adenovirus encoding for GFP (Ad-GFP) were used as control. The expression of genes related to the Notch signalling pathway was determined by qPCR-based screening of a "Notch Signalling Plus gene array" (QIAGEN).

The Notch pathway is a canonical signalling mechanism that controls cell fate and differentiation which plays a vital role during vascular development, and the angiogenic process (Hasan et.al., 2017). Notch ligand-receptor mechanism is activated via cell-cell interactions when the receptor's extracellular domain interacts with the corresponding ligand found on the adjacent cells (Dufrain et.al., 2008). The Notch pathway plays a key role during vascular development. Complete knockout of Notch ligands Notch1 and Notch4 was found to cause lethal vascular defects (Krebs et.al., 2000).
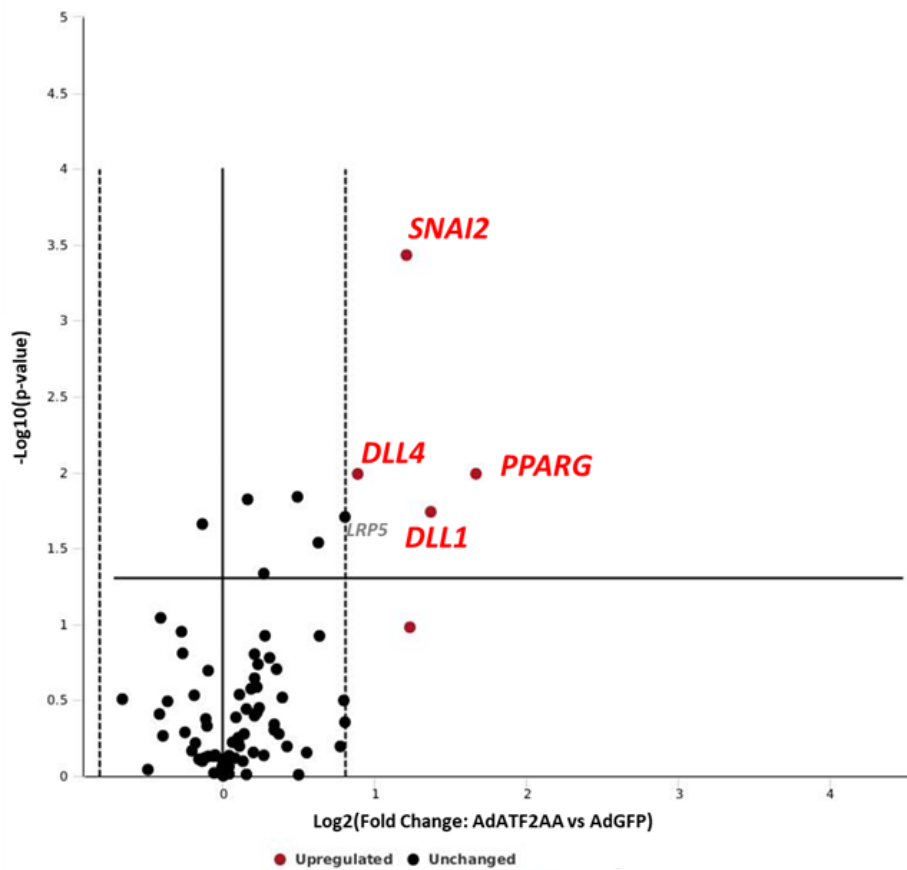
Figure 3.1: Volcano plot of Notch signalling gene expression in endothelial cells lacking functional ATF2. HUVECs were infected with an adenovirus expressing ATF2-AA (a dominant-negative version of ATF2) or an adenovirus expressing GFP as control. Four genes were found to be upregulated. Two major Notch pathway ligands Delta-like Protein 4 (DLL4) and Delta-like Protein 1 (DLL1) were upregulated in endothelial cells lacking functional ATF2, along with SNAI2, a transcription factor, and PPARG; a nuclear receptor from peroxisome proliferator-activated receptor (PPAR) subfamily.

3.1.2 Expression of DLL4 Notch signalling ligand in HUVECs infected with Ad_ATF2

It has been demonstrated in experimental settings that DLL4 expression on tip endothelial cells is regulated by VEGF (Siekmann et.al., 2007, Hasan et.al., 2017). The question here is what role could ATF2 be playing in the regulation of this interaction? To approach this matter HUVECs infected with Ad-ATF2_AA, and Ad-GFP (control) were stimulated with VEGF at a concentration of 25ng/ml in a course-dependent manner (0.5h, 1h and 3h). Before stimulation, Ad-GFP and Ad-ATF2_AA expression in HUVECs was determined by western blot analysis in Dr. Angel Armesilla's group (Figure 3.2, Suhail Ahmed and Prof. Angel Armesilla, internal communications). When stimulation with VEGF was completed DLL4 expression was determined by qPCR using specific TaqMan primers. The result showed (Figure 3.3) enhanced VEGF-dependent upregulation of DLL4 in endothelial cells infected with Ad_ATF2_AA when compared with the control after half an hour and one hour. The increase in basal (0 hours) and VEGF- dependent stimulation at 3 hours of DLL4 expression was not statistically significant.
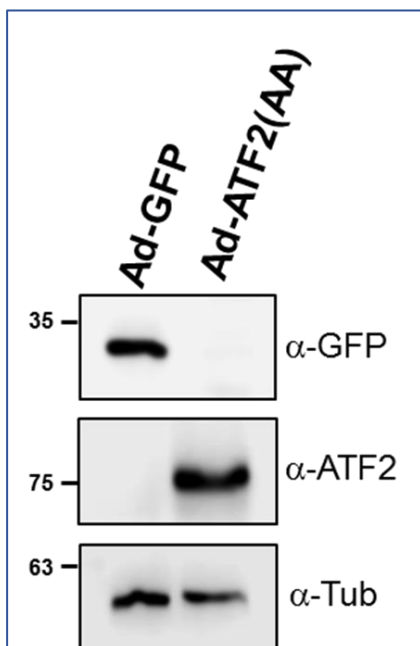
Figure 3.2: Western blot results for Ad_GFP and Ad_ATF2 protein expression in HuVECs. Infection efficiency of the selected adenoviruses, showed by significant expression of recombinant Ad_ATF2_AA when compared with cells infected with Ad_GFP. A-Tubulin was used to confirm equal loading.
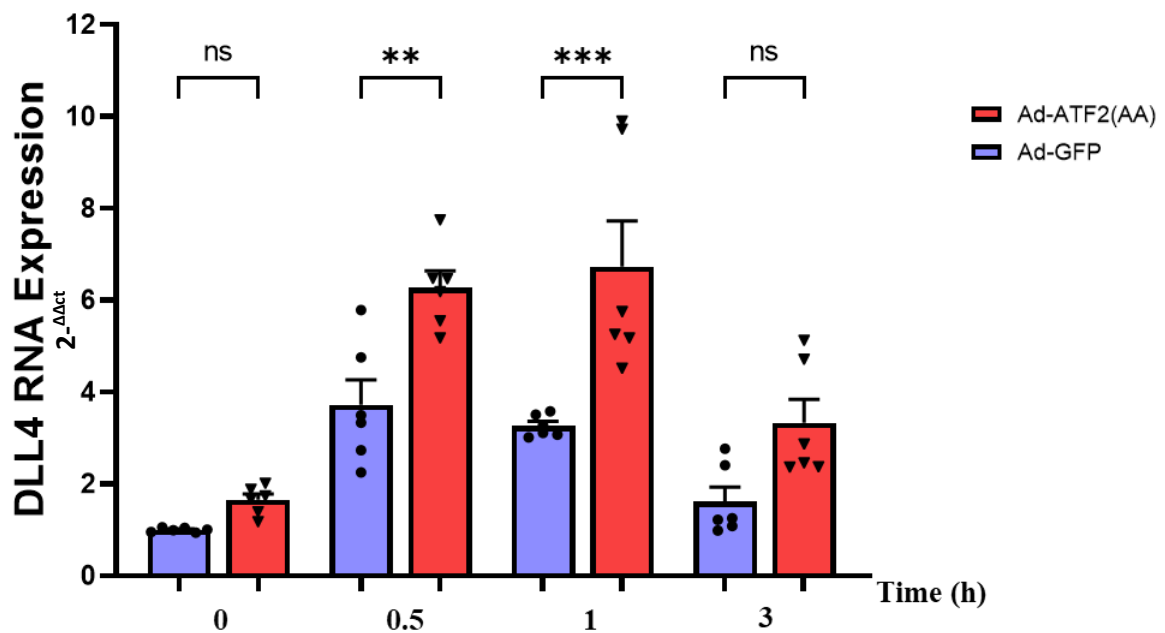
Figure 3.3: DLL4 expression in HUVEC cells infected with ad-GFP and AD_ATF2_AA. Suppression of ATF2 functional activity enhances VEGF-dependent upregulation of DLL4. HUVEC infected with Ad-GFP (control) or Ad-ATF2 (AA) were stimulated with VEGF by the time indicated and the expression of DLL4 was determined by qPCR using a specific TaqMan gene expression assay. Data are expressed as $2-\Delta\Delta ct$ of unstimulated Ad-GFP infected cells. Hprt-1 expression was used for control. Experiments were done in biological triplicates.

3.1.3 Expression of DLL1 Notch signalling ligand in HUVECs infected with Ad_ATF2

The expression of DLL1 in HUVECs after adenoviral infection was also determined by RT-PCR. Cells were stimulated with VEGF (25 ng/ml) in a time-dependent manner. The results showed a VEGF-dependent upregulation of DLL1 at half an hour and one hour post-stimulation (Figure 3.4). The role of DLL1 during angiogenesis has not been characterized. However, studies in mice have shown that DLL1 is required for Notch activation in vascular endothelial cells to maintain endothelial identity via regulation of VEGFR2 expression in foetal arteries (Sörensen et.al., 2009). Also, it has been reported that DLL1 is essential for the regulation of postnatal arteriogenesis (Limbourg et al., 2007). In a study carried out in 2019 it was found that DLL1 has a vital role in promoting angiogenesis and tumour cell proliferation (Kumar et al., 2019).
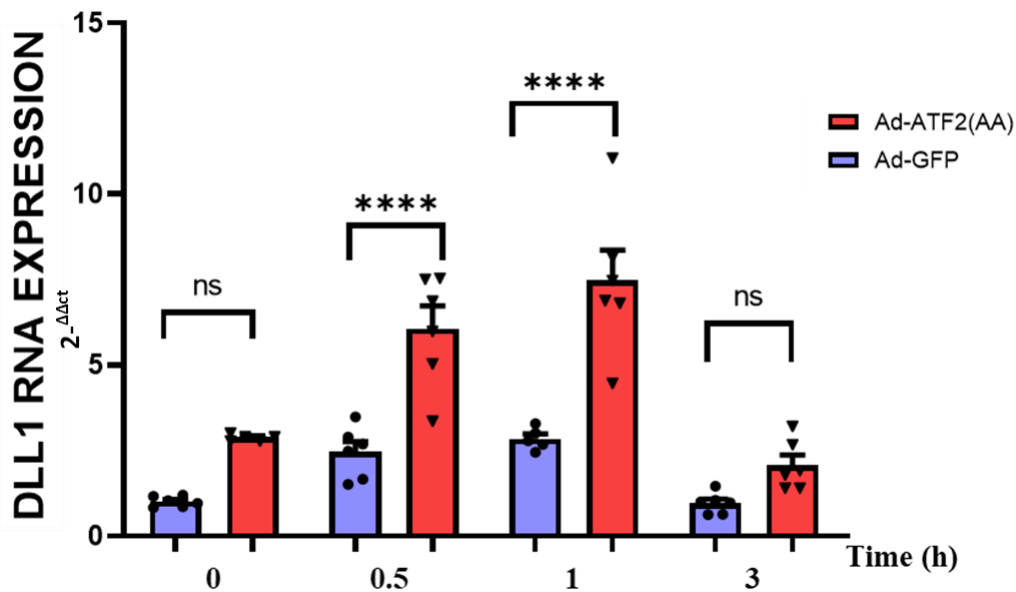
Figure 3.4: DLL1 gene expression in HUVECs infected with Ad-GFP and Ad_ATF2_AA. Depression of ATF2 functional activity enhances VEGF-dependent upregulation of DLL1. HUVEC infected with Ad-GFP (control) or Ad-ATF2 (AA) were stimulated with VEGF by the time indicated and the expression of DLL1 was determined by qPCR using a specific TaqMan gene expression assay. Data are expressed as 2-ΔΔct of unstimulated Ad-GFP infected cells. Hprt-1 expression was used for control. Experiments were done in biological triplicate.

Figure 3.5: SNAI2 and PPARG expression in HUVECs infected with Ad-GFP and Ad_ATF2_AA. Depression of ATF2 functional activity resulted in no statistically significant regulation of SNAI2 and PPARG genes. HUVECs infected with Ad-GFP (control) or Ad-ATF2 (AA) were stimulated with VEGF by the time indicated and SNAI2 (A) and PPARG (B) gene expression was determined by qPCR using a specific TaqMan gene expression assay. Data are expressed as 2-ΔΔct of unstimulated Ad-GFP infected cells. Hprt-1 expression was used for control. Experiments were done in biological triplicates.

3.1.4 Expression of Notch ligands DLL1 and DLL4 in HUVECs transfected with siRNA-ATF2

To further test ATF2 effects on endothelial cells, ATF2 gene knockout was performed in HUVECs using small interfering RNA (siRNA). ATF2 expression in HUVECs transfected with either siRNA-non-Target or siRNA-ATF2 was determined by qPCR (Figure 3.6 A) The siRNA transfected cells were stimulated with 25ng/ml of VEGF in a time-dependent manner (0.5h, 1h and 3h). RNA expression of Notch ligands DLL1 and DLL4 in HUVECs transfected with siRNA non-target (siRNA-NT) or siRNA-ATF2 was determined by qPCR. Results showed (Figure 3.6 B) that ATF2 silencing in HUVECs resulted in VEGF-dependent upregulation of Notch pathway ligand DLL4 at 0.5 and 1-hour post-stimulation. An increase in DLL1 expression was observed however it was not statistically significant (refer to Figure 3.7).
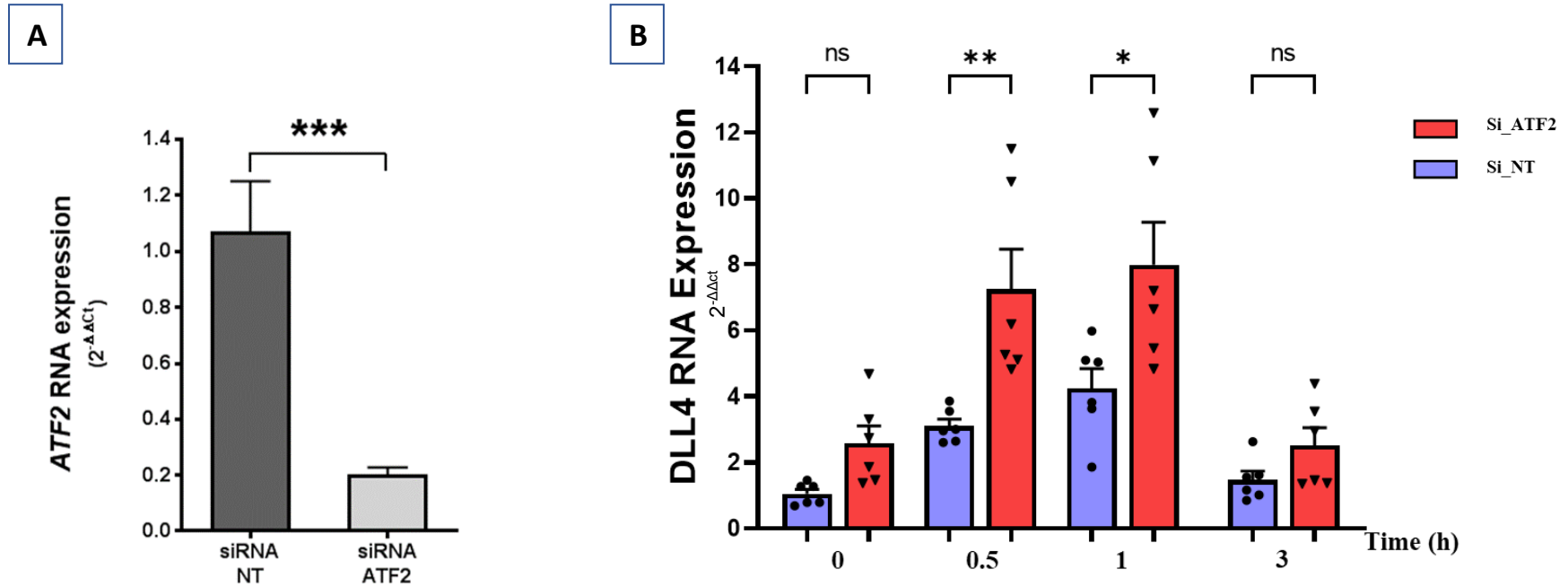
Figure 3.6: DLL4 expression in HUVECs transfected with si_ATF2 or si_NT. A) ATF2 RNA expression in HUVECs transfected with siRNA non-target (siRNA-NT) or si RNA-ATF2 was determined by qPCR. B) TaqMan validation of DLL4 expression after transfection of HUVECs with Si-Non-Target (Si-NT) and Si-ATF2. ATF2 silencing results in an increase in basal and VEGF-dependent DLL4 expression. Data are expressed as 2-ΔΔct of unstimulated Si-NT transfected cells. Hprt-1 expression was used for normalization. Experiments were done in biological triplicates.
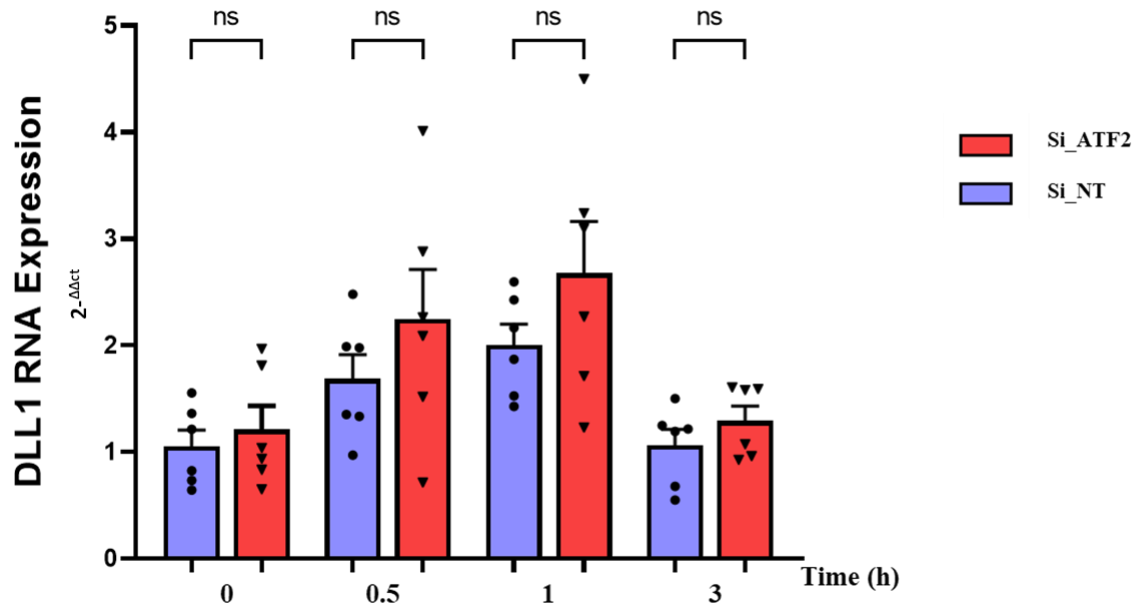
Figure 3.7: DLL1 expression in HUVECs transfected with si_ATF2 or si_NT. TaqMan validation of DLL1 expression after transfection of HUVECs with siRNA-Non-Target (Si-NT) and Si-ATF2. ATF2 silencing results in a non-significant (ns) increase of DLL1 expression in basal and VEGF stimulated con. Data are expressed as 2-ΔΔct of unstimulated Si-NT transfected cells. HPRT1 expression was used for normalization. Experiments were done in biological triplicates.

3.2 Assessing the role of ATF2 activity in endothelial tubule formation

3.2.1 Organotypic coculture

The mechanism underlying tubule formation and development involves the signalling of several molecules such as growth factors and their receptors (Iruela-Arispe and Beite, 2013). RTKs are major regulators of cell proliferation, differentiation, and apoptosis (Fearnley et.al. 2020). RTKS regulatory functions are ubiquitous or cell-specific, and their signal transduction regulation is required to modulate cell behaviour and responses to extracellular signals (Lemmon and Schlessinger, 2010).

VEGF signalling mechanism is carried out throughout VEGFR which are class III RTKs. The function of the VEGF/VEGFR axis in cellular signal transduction that regulates cell proliferation and tubule formation has been well described (Smith et.al., 2015). However, how VEFG modulates the biochemical signals responsible for cell fate is not well characterized (Parks et.al., 2018). To understand how biochemical regulation takes place Fearnley and colleagues (2020) studied the impact of ATF2 knockout on endothelial cells. The result was a decrease in VEGF-A-dependent cell proliferation along with a decrease in tubule formation due to p53 overexpression.

To assess the role of ATF2 in tubulogenic organotypic co-culture was carried out in adenovirus-infected GFP-HUVECs lacking functional AFT2. Cells were seeded on top of a monolayer of human derived fibroblast (HDFα), stimulated with bFGF, and quantification of tubule formation determined by anti-CD31 antibody staining. Pairwise comparison between the groups showed no statistical significance in tubule formation between the Ad-GFP and Ad-AFT2 controls. However, there was a significant increase in tubule formation when comparing HUVECs Ad_GFP control and Ad_GFP stimulated with bFGF (Figure 3.8). Tubule formation in Ad_ATF2 HUVECs stimulated

with bFGF showed a slight increase in tubule formation when compared with Ad_GFP control. However, when compared with Ad_GFP cells stimulated with bFGF tubule formation was higher in Ad_GFP transfected cells (Figure 3.8).
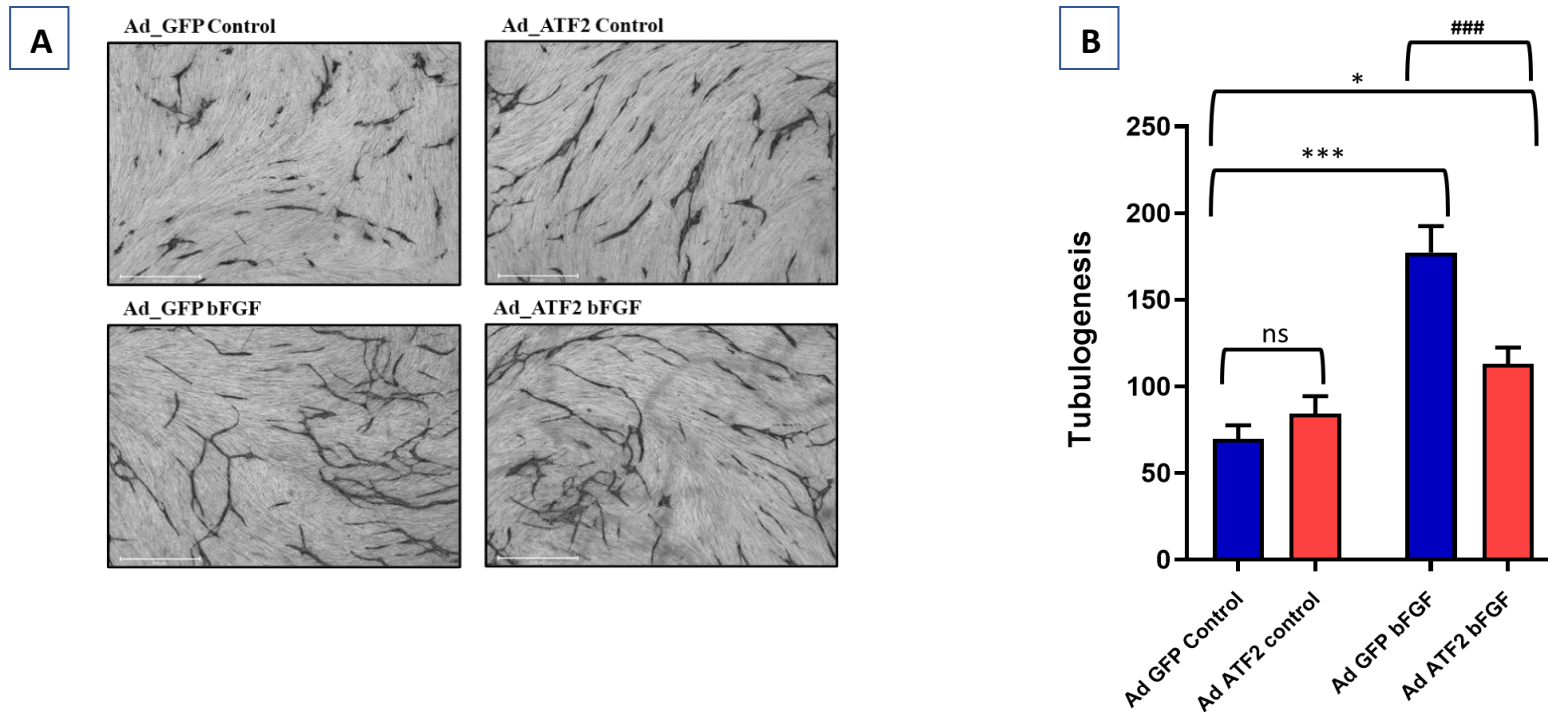
Figure 3.8: Organotypic coculture of adenovirus-infected GFP-HUVECs lacking functional AFT2. GFP-HUVECs infected with the indicated adenoviruses were seeded on top of a monolayer of HDFα to analyse tubule formation in organotypic co-culture angiogenesis assay. Tubule formation was quantified by staining with an anti-CD31 antibody. The total tubule length was quantified using imaging processing NIH ImageJ 1.32 software (N=6).

3.3 Discussion

VEGF is a well-characterised endogenous stimulator during the physiological angiogenic process and a major regulator of pathological angiogenesis (Goradel et al, 2018). Anti-VEGF therapy has been widely used for the treatment of human pathologies characterised by increased blood vessel formation such as cancer. However, the currently available anti-VEGF drugs are not completely effective when targeting angiogenesis, as tumours develop resistance and other pro-angiogenic molecules are released by the tumours as a consequence of anti-VEGF tumour treatment (Zahra et al., 2021). Thus, it is essential to explore a new molecular target that is commonly activated by all the pro-angiogenic factors and that will lead to decreased tumour angiogenesis.

The proposed hypothesis suggests that ATF2 is a downstream transcription factor common to the major pro-angiogenic molecules such as VEGF, bFGF, EGF, and HGF and preliminary unpublished results obtained by Prof. Armesilla's laboratory group have shown that ATF2 is activated by all the pro-angiogenic molecules (Prof. A Armesilla, personal communication). However, the molecular mechanisms involved in the regulation of the angiogenic process by ATF2 remain largely unknown.

ATF2 is a phosphorylation-dependent regulator of cellular processes such as the DNA damage response elicited by ionizing radiation or ultraviolet light (Bohoumik et al.,2005, Lopez-Bergami, Lau and Ronai, 2010). The link between ATF2 and VEGF-induced angiogenesis has been supported by in vitro and in vivo studies (Fearnley et al., 2014; Bus et al., 2018; Fearnley et al., 2019; Wang et al., 2021). VEGF regulates the angiogenic and inflammatory process via activation of specific extracellular signals leading to phosphorylation of ATF2 that modulates the expression of angiogenesis-

associated genes (Fearnley et al., 2014; Gozdecka et al., 2014). Also, as part of the pro-angiogenic transcription program in nasopharyngeal carcinomas, ATF2 transcriptional activity is stimulated by the Epstein–Barr nuclear antigen 1 (EBNA1) (Lau and Ronai, 2012). Consequently, it was decided to investigate the role of ATF2 in the regulation of the angiogenic process.

Preliminary Notch array assay showed that two major Notch pathway ligands Delta-like Protein 4 (DLL4) and Delta-like Protein 1 (DLL1) were upregulated in endothelial cells lacking functional ATF2, along with SNAI2, a transcription factor, and PPARG; a nuclear receptor from peroxisome proliferator-activated receptor (PPAR) subfamily. To further assess the role of ATF2 in the regulation of these 4 genes, RNA expression was quantified by real-time PCR in HUVECs lacking functional ATF2, this was done by either knockdown or overexpressing a mutant protein. Significant expression of the Notch ligand DLL4, but not significant expression of DLL1, SNAI2 and PPARG, were observed in HUVECs deficient in functional ATF2.

Notch-related genes have been reported to play a key role in blood vessel formation (Roca and Adams, 2007; Hofmann and Iruela-Arispe, 2007). DLL4 have been reported to be a negative regulator of the angiogenic process and several studies have established that DLL4 is essential for vascular development due to its regulatory role during the tip and stalk cell differentiation in the course of vascular development (Hellström et al., 2007; Suchting et al., 2007). I t was determined that DLL4 is crucial for formation of embryonic vasculature, an experiment in DLL4 knockout mice resulted anomalous vascular remodelling that cause the death of the mice embryos (Krebs, 2004).

During blood vessel sprouting, DLL4 expression in tip endothelial cells is stimulated by VEGF which results in the activation of Notch in the neighbouring tip cells which results in changes in DLL4 gene expression (Claxton and Fruttiger, 2004). This indicates that DLL4 regulates the angiogenic process downstream of VEGF activity.

It has been reported that loss of DLL4 in endothelial cells results in excessive blood vessel formation. In a study on tumour angiogenesis, conducted to study the effect of DLL4 in tumour angiogenesis, it was observed that inhibition of DLL4 resulted in markedly increased tumour vascularity. However, these vessels were characterised by poor perfusion and increased hypoxia (non-functional blood vessels) which resulted in inhibition of tumour growth. In the same study, when analysing the effect of VEGF-induced DLL4 expression, they noted that angiogenesis was negatively regulated by DLL4 during tumour development (Noguera-Troise et al., 2006). Importantly this suggests a novel anti-tumour therapy approach even for tumours resistant to anti-VEGF drugs.

Additionally, it has been described that loss of Notch signalling results in an increase in VEGF receptors (VEGFR-1 and VEGFR-3) signalling activity, which indicates Notch signalling negatively regulates the VEGF-VEGFR axis (Kuhnert, Kirshner and Thurston, 2011).

In this study, it was observed that ATF2 might negatively regulate the expression of DLL4. The molecular mechanisms by which ATF2 modulates DLL4 expression are still to be elucidated. It can be hypothesised that the negative regulation of DLL4 observed in this study could be the result of direct ATF2 binding to DLL4 gene regulatory regions or indirect activation by ATF2 of a DLL4 repressor that would result in decreased DLL4 expression. The negative regulatory effect of ATF2 in gene expression has been

demonstrated by several studies. A Chromatin immune precipitation assay (ChIP) determined that ATF2 has transcriptional repression activity on interferon-β1 (IFNβ1) by direct binding to the gene promoter region (Lau et al., 2015) and another study carried out in Macrobrachium Nipponese found that antimicrobial peptide genes are negatively regualted by ATF2 via regulating the expression of TNF (Zhang et al., 2020).

Another possible explanation for ATF2 negative regulation of DLL4 is that ATF2 might be regulating the expression of specific miRNAs that would target the DLL4's three prime untranslated region resulting in expression of the gene being reduced. In a study conducted to assess the role of microRNA-30b (miR-30B) in the regulation of cellular morphogenesis via transforming growth factor beta 2 (TGFβ2) regulation, it was found that cells depleted of ATF2 presented inhibited miR-30b- dependent TGFβ2 expression. This suggested that ATF2 activation (phosphorylation) was required for miR-30b up-regulation of TGFβ2. Conversely, the effects of miR-30b on ATF2 were carried out by JDP2, an ATF2 repressor molecule, this resulted in decreased JDP2 mRNA expression (Howe, Kazda and Addison, 2017).

As previously mentioned, when assessing for DLL1, SNAI2 and PPARG expression data showed no statistical significance. Many plausible explanations can be delivered to rationalize these results. It can be argued that this is due to the activity of another ATF family member such as ATF7 since it also targets the ATF2 DNA binding site (Liu et al., 2016). Preliminary analysis showed no significant alteration in ATF7 RNA expression in HUVECs with functionally suppressed ATF2 (Prof. A Armesilla, personal communication). Hence it can be argued that ATF7 might be compensating for suppressed ATF2 activity in the regulation of DLL1 in the experiments carried out using si-RNA mediated suppression where no significant changes in gene expression

were observed. Further analysis would be required to elucidate the role of ATF7 in the regulation of this gene.

ATF2/ATF7 redundant gene regulation was not observed in the case of DLL4, as significant upregulation of this gene was observed in both, ad_ATF2_AA infected cells and in si_ATF2 knockdown cells.

Furthermore, in the experiments, conducted using the mutant version of ATF2, ATF7 redundant action cannot occur because the ad-ATF2-AA (mutant) occupies the DNA binding site preventing both ATF2 and ATF7 from binding the domain and thus causing changes in gene expression. However, no significant changes were observed in DLL1, SNI2 and PPARG in HUVECs infected with ad_ATF2_AA. In this case, it could be debated that the difference in gene expression could be the result of ATF2 dual transcriptional activity that allows it to either inhibit or promote the expression of genes depending on its association with other transcription factors. It has been reported that ERK1/2/ AFT2 complex increased the expression of VCAM-1 which is essential for endothelial leukocyte interactions (Bus et al., 2018). On the other hand, the ATF2/JunD complex results in the downregulation of CDK4 (Xiao et al., 2010) a cyclin-dependent kinase involved in angiogenesis (Huang et al., 2019).

For endothelial cells to undergo angiogenesis, the interaction between pro-angiogenic molecules, stromal cells, and extracellular matrix components is essential (Allen and Mellor, 2014). Organotypic co-cultures are long-term 3D model assays that have been generally used to assess angiogenic regulatory molecules, signalling pathways, and for high-resolution imaging of cells undergoing angiogenesis (Paoli and Carrer, 2020; Shamir and Ewald, 2014). These 3D model assays recreate the angiogenic process

in the laboratory setting by using fibroblasts, thus recreating the stromal conditions of tissues (Hetheridge, Mavria and Mellor, 2011).

To further evaluate the role of ATF2 in the angiogenic process, organotypic co-culture experiments were carried out to quantify tubule formation in GFP-HUVEC. ATF2 suppression resulted in significantly decreased tubule development in GFP-HUVEC lacking functional ATF2 and stimulated with the pro-angiogenic molecule bFGF when compared to the control. This correlates with a study in live cell culture estimating tubule formation and wound healing, carried out in human retinal microvascular endothelial cells (HRECs) by Wang and colleagues, it was reported that formation of new vasculature in the retina was inhibited by H-KI20 (a small peptide), via the JNK/ATF2 signalling pathway and that ATF2 knockdown in retinal cells resulted in decreased VEGF-dependent tubule formation (Wang et al., 2021).

For future work, coculture assays could be used to characterise the breast to brain metastasis angiogenic process. Coculture studies have been used to determine the angiogenic potential of tumours and to assess their sensitivity to antiangiogenic therapy (Truelsen et al., 2021). Deregulated angiogenesis is one of the hallmarks of cancer cells that is required for tumour progression and metastasis development and maintenance (Hanahan, 2022). Truelsen and colleagues used a modified version of the organotypic coculture, the cancer coculture assay, to assess tubule formation and to quantify the sensitivity of these tumour derived tumoroids to antiangiogenic drugs. They cocultured fibroblasts, vascular cells and tumoroids from colorectal carcinoma and liver metastasis, thus recreating tumour angiogenesis. Their findings confirm that the angiogenic process is regulated by multiple pro angiogenic stimuli and that with their assay could be used to classify tumours as resistance or sensitive to antiangiogenic drugs in invitro experiments (Truelsen et al., 2021).

Additionally, coculture assays can be used to assess tumour and extracellular matrix interactions which could provide a better understanding of tumour interactions with their microenvironment (Paoli and Carrer, 2020). Coculture models have been largely used to assess functional interactions between tumours and their associated stroma (Schmeichel et al., 2003; Gaggioli et al., 2007; Froeling, Marshall and Kocher, 2010; Padmanaban et al., 2020). Studying these associations provides insight into the mechanisms involved in metastatic tumour development and progression, as well as the specific molecules, genes and their related pathways that could be implicated in drug resistance mechanisms (Dolznig et al., 2011; Padmanaban et al., 2020).

3.4 Conclusion

ATF2 can be considered as a novel molecular target for antiangiogenic therapy in endothelial cells. Here it was shown that this transcription factor may play a role in angiogenesis by negatively regulating DLL4 Notch signalling pathway ligand in endothelial cells. Future work would include establishing ATF2 binding sites in the regulatory region of DLL4 to determine the molecular mechanisms behind this interaction. Additionally, the effect of ATF2 suppression in DLL4 target genes could be determined to further understand ATF2 implication in the regulation of angiogenesis. Organotypic coculture assay showed that ATF2 participates in tubule formation in endothelial cells. Further analysis would be required to validate these results.

Chapter 4 – Whole-Exome Sequencing of 26 Breast to Brain Metastases

4.1 Introduction

One major problem for patients with systemic cancer is the development of Central Nervous System (CNS) metastases after receiving a successful course of treatment for the primary malignancy, as in the case with many patients with primary breast cancer (Bartsch et al., 2022). Several cancer studies have explored the factors implicated in brain metastasis development, however the mechanisms underlying their progression and why there is a preference for this site are not completely understood (Chen et al., 2018; (Chaffer and Weinberg, 2011; Zhen et al., 2019). Characterising molecular genomic changes that give rise to breast to brain metastases appears a step in the right direction to achieve a better understanding of the mechanisms behind the formation of these tumours and their evolution (Lv et al., 2021). Specific genomic alterations can be used as tumour biomarkers and therapeutic targets that might allow monitoring of disease progression and response to current treatment, as well as to support the development of target therapy that could be more efficient for treating these deadly tumours (In et al., 2020).

Most sequencing studies on breast cancer have been centred on determining the genomic evolution of the primary tumour. However, breast to brain metastasis (BBM) has become an ever-increasing medical challenge (Frisk et al., 2017; Kotecki et al.,2018). Despite the advances in treatment for brain metastases such as chemotherapy and target therapy, following local surgical resection of the tumour, patient survival rates remain low (Lee et al., 2015; Choong, Cullen and O'Sullivan, 2020). The scarcity of genomic characterization of brain tumours makes it difficult to

make improvements in systemic therapeutic approaches that will improve patient survival (Tyran et al., 2019).

Several studies have shown that breast to brain metastases are molecularly related to their primary tumour of origin. However, they also harbour additional molecular characteristics which suggest that they keep acquiring mutations during tumour evolution, characteristics that were not described in their ancestor (Ciriello et al., 2015; Yate et al., 2017; Brosnan and Anders, 2018). Despite these insights, the mechanisms by which primary tumours are capable of invading and adapting to the new tumour microenvironment have not yet been fully revealed (Ciriello et al.,2015; Rinaldi et al., 2020).

Next-generation sequencing (NGS) techniques are currently being employed to get a better understanding of the cancer molecular landscape by revealing differences in gene expression, copy number variations, and pathway alterations that can be used as predictors of cancer cell behavioural patterns, and as targets for personalising therapy (Zhang et al., 2018; Gambardella et al., 2020).

Whole-exome sequencing (WES) is an affordable, high-coverage NGS technique that allows massive parallel reads of DNA samples, resulting in the generation of a large amount of data within a brief time. Because exons comprise ~1% of the human genome and around 85% of disease-causing mutations have been identified within this region (Rabbani, Tekin and Mahdieh, 2013; Chang et al,.2020), WES might be a suitable approach for identifying novel gene mutations that could be contributing to cancer development and progression.

For this research, WES was carried out to determine the molecular composition of 26 breast to brain metastases. WES was carried out by Dr. Andrew Beggs at the Beggs Lab from the Institute of Cancer and Genomic Science at the University of Birmingham, UK. Tumour samples were sequenced on high output V2 flow cell as a 75 paired and run on the NextSeq 500 (Illumina). The data generated by WES was filtered using the bioinformatic approach previously described in the materials and methods chapter (Chapter Materials and Methods section 4.4.3). After variant calling, the information of 491,748 single nucleotide variants (SNVs) was converted into an excel text file for further Insilico analysis.

4.2 Preliminary WES data analysis

Sequencing data from metastatic breast cancer studies have reported that primary tumours share driver mutations with their resulting metastatic tumours (Bastianos et al., 2015; Yates et al., 2017; Bertucci et al., 2019). Following this model, the first step for WES data analysis in this study was to determine if any of the sequenced BBM samples harbour any of the genetic mutations commonly found in primary breast cancer. The frequency of the most mutated genes in primary breast tumours reported in the Catalog of Somatic Mutations in Cancer (CoSMIC) is represented in Figure 4.1.

Alterations in the 20 most commonly mutated genes in primary breast tumours were found across all of the 26 BBM samples of this cohort (Figure 4.2).
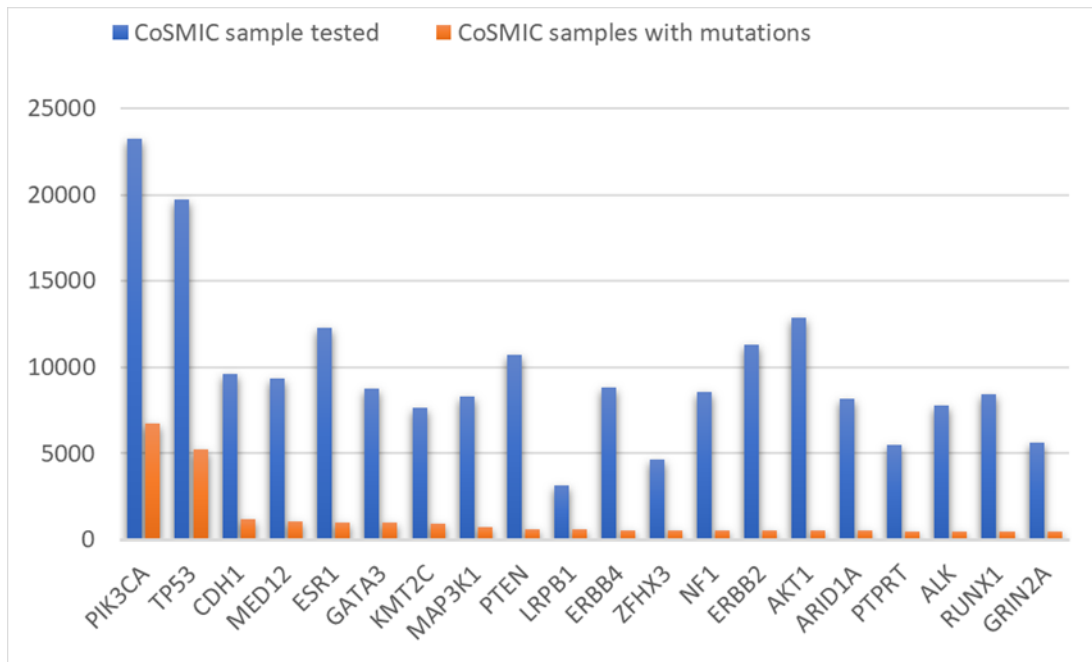
Figure 4.1: Most frequently mutated genes in primary breast tumours reported in CoSMIC. Image showing the most commonly mutated genes across primary breast cancer. The blue and orange bars represent the total samples tested for the specific gene and the number of samples with mutations respectively. Data obtained from CoSMIC (cancer.sanger.ac.uk, n.d.).
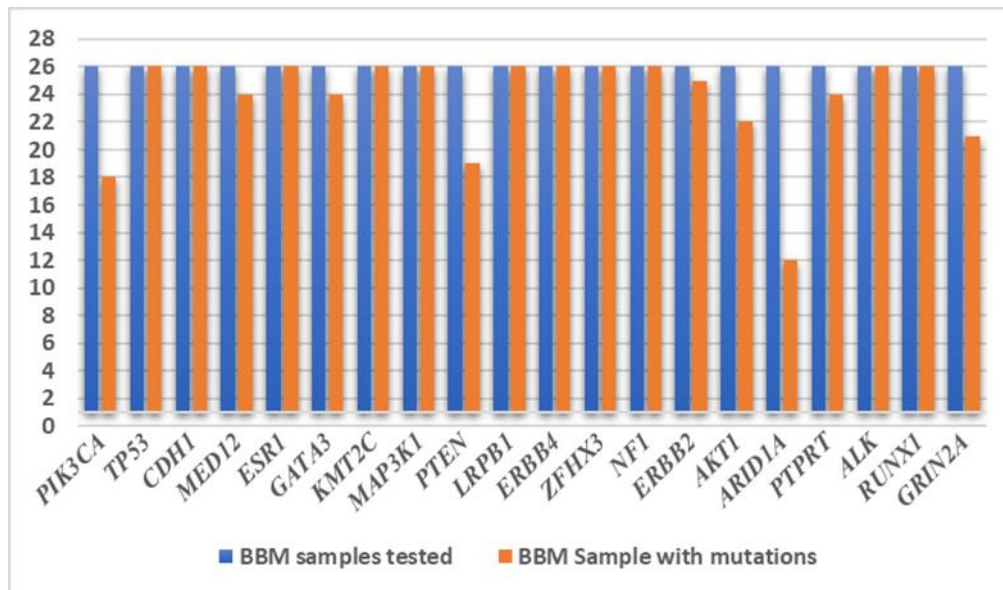
Figure 4.2: Distribution of most frequently muted genes in primary tumours found in the BBM samples. Genes reported to be frequently mutated in primary breast tumours in CoSMIC. Gene alterations in primary tumour-associated genes were found across all BBM samples (26 BBM) in this study.

PIK3CA is a well-characterised cancer driver gene (Pearson et al., 2018; Lai et al., 2020). Mutations in this gene have been found in several tumour types such as breast, bladder, cervical, colorectal, and head-and-neck cancers (Arafeh, Samuels, 2019, Jin et al., 2019a; Pergialiotis et al., 2020; Basu et al., 2020). Additionally, curated data from thousands of publications available in COSMIC (cancer.sanger.ac.uk, n.d.) described PIK3CA as the most common type of gene mutation (29%) found in primary breast cancer patients. Other well-documented tumour-associated genes, such as TP53 and CDH1, represent the second and third most common gene mutations in primary breast tumours (Niculescu, 2019; Gu et al., 2020; Zhang et al., 2020b).

Of the most frequently mutated 20 genes in primary breast cancer, a total of 1867 mutated variants were identified in the 26 BBM samples. Thirty-one PIK3CA mutated

variants were found in 18 of the BBM samples. Of the top 20 genes, a recurring set of eleven genes were found to be mutated in all of the samples. The variants identified were predominantly localised within the intronic genes regions shown in Figure 4.3 (A) therefore they were not considered to be damaging mutations for the translated protein. Sixty five of these mutations were SNVs and exon insertion-deletion (indels) mutations that can be damaging for the protein function. In the BBM the most frequently mutated gene from the 20 most commonly muted in primary breast cancer were KMT2C, followed by ZFHX3, LRPB1, and PIK3CA. The presence of these suggests that specific molecular characteristics of the primary tumour were maintained during primary tumour evolution to brain metastasis mutations (frequency of mutation per genein the BBM samples is shown in table 4.1).

Gene name	Missense	Nonsense	Exon Indels	Splice-site 3	Splice-site 5	Total variants/gene	Percentage
KMT2C	13	1	3	1	1	19	29
ZFHX3	8	1	4	0	0	13	20
LRPB1	4	0	1	1	0	6	9
PIK3CA	5	0	0	0	0	5	8
ALK	4	0	0	0	0	4	6
GATA3	0	0	3	0	0	3	5
PTPRT	3	0	0	0	0	3	5
MAP3K1	1	0	1	0	0	2	3
GRIN2A	2	0	0	0	0	2	3
NF1	1	0	1	1	0	3	5
MED12	1	0	0	0	0	1	1.5
ESR1	1	0	0	0	0	1	1.5
PTEN	1	0	0	0	0	1	1.5
ARID1A	0	0	1	0	0	1	1.5
ERBB2	1	0	0	0	0	1	1.5
TP53	0	0	0	0	0	0	0
CDH1	0	0	0	0	0	0	0
ERBB4	0	0	0	0	0	0	0
RUNX1	0	0	0	0	0	0	0
AKT1	0	0	0	0	0	0	0
Total	45	2	14	3	1	65	

Table 4.1: Frequency of mutations found in the BBMs of the most mutated tumours in primary breast tumours. Sixty-five variants that might be damaging to the translated protein were identified for this cohort of 20 genes in 26 BBM samples.

The top five mutated genes are highlighted in red. The highest number of missense variants were observed in KMT2C (13) and ZFHX3 (8).
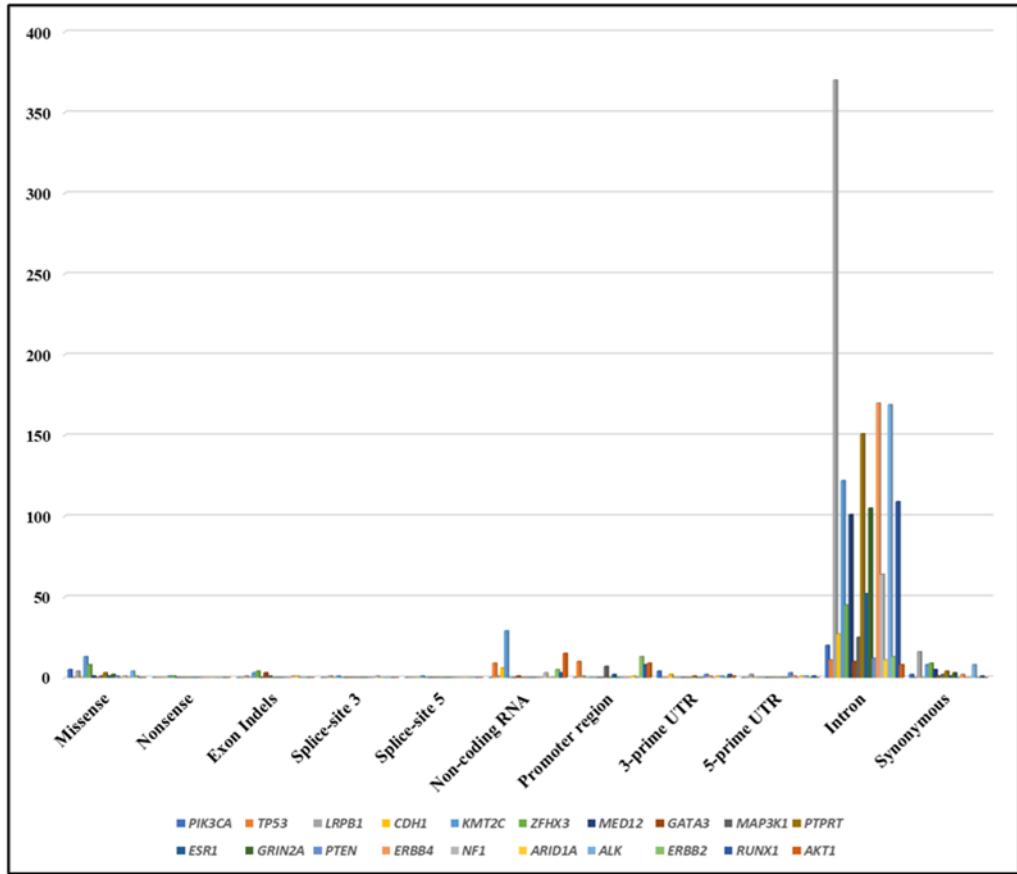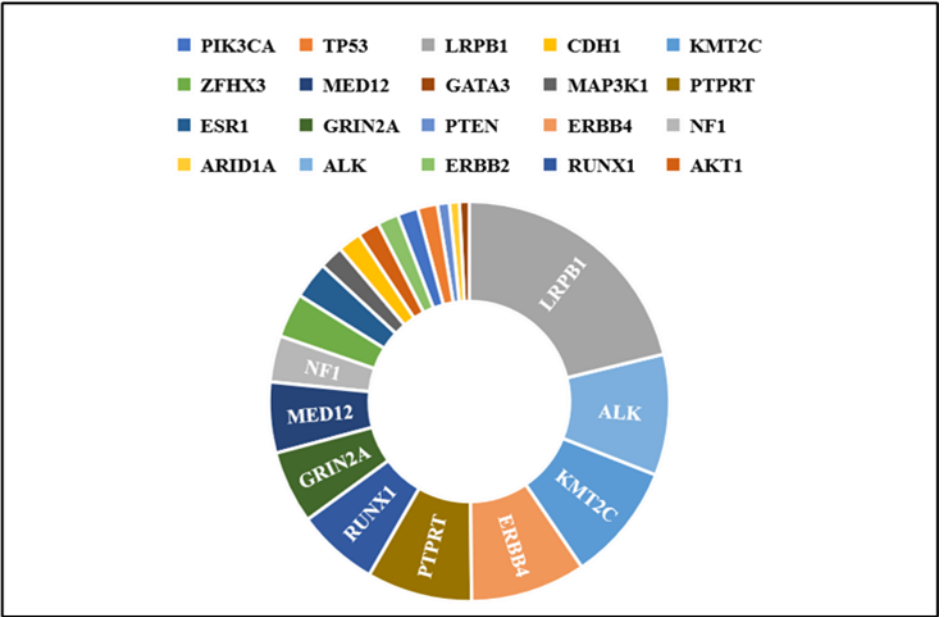
Figure 4.3: Mutations location in the BBM of the 20 most common mutated genes in primary breast tumours. Charts displaying

mutation distribution among twenty well-characterised primary breast cancer-associated genes found mutated in 26 BBM samples.

A) type of variants identified by WES in this cohort of 20 genes, where the majority of changes were found in the introns. B) Shows the most commonly mutated genes from the top 20 most frequently mutated genes found in primary breast tumours in CoSMIC; LRPB1 was the most mutated gene in the BBM for this group of 20 genes, with most mutation concentrated in the intronic regions of the genes Image A.

4.3 WES pipeline to identify potentially metastasis-associated genes

As previously described in the materials and methods section in chapter 2; WES and the initial bioinformatic analysis were carried out at the University of Birmingham, UK. In this study 491,748, single nucleotide variants (SNVs) and insertion or deletions (indels) were identified by WES in 26 BBM samples. Due to the high number of variants identified by WES studies, identifying disease-associated alterations can represent a challenge. Therefore, to reduce the total number of identified variants, a stepwise filtering process was carried out to identify metastasis-associated candidate genes (Figure 4.4). Firstly, synonymous alterations were excluded as they were considered to be changes that do not alter the protein structure (Sharma et al., 2019). Intronic and regulatory region alterations such as non-coding RNA, 5' and 3' UTRs, and promoter region variants were excluded from this analysis as they do not cause direct changes in the translated protein sequence (Martínez-Pizarro et al., 2018), 468,462 variants were removed in this step. Consequently, nonsynonymous variants, exon indels, splice site and nonsense alterations were prioritised as they have a direct effect on the protein function (Reva, Antipin and Sander, 2011). Nonsense mutations introduce a premature stop codon resulting in a shortened protein product, while splice site variants can change the exon-intron splicing during mRNA processing by altering the nucleotide sequence at the splice consensus sequence which results in either skipping or retention of the exon (Agnihotri et al., 2016; Makinen et al., 2016).

Following variant selection 23,286 variants were retained (Appendix D1), the Minor Allele Frequency (MAF) score was used to further filter potential interesting polymorphisms. MAF indicates the frequency at which the second most common allele appears at a specific location, specified by the 1000 Genomes project phase 3 (internationalgenome.org, 2012). MAF values help identify novel or rare single

nucleotide polymorphisms (SNPs) in population studies (Chandler, Bilgili and Merner, 2016; www.ensembl.org, 2020). A cut-off MAF of ≥1% was set to eliminate single-nucleotide variants (SNVs) and indels that are common in the general population as they were likely to be common polymorphisms. Variants with a MAF of ≤ 1% were retained as they are considered rare genetic variants in the general population, as indicated in other WES cancer studies and the 1000 Genomes Project (Panagiotou, Evangelou, and Ioannidis, 2010; Lasky-Su, 2016). A variant with low MAF value is associated with high gene conservation, and mutations at that locus might be detrimental to the protein function (Hosonaga, Saya and Arima, 2013). Variants without an annotated MAF value on the WES data set were also retained for further analysis. At this step of the filtering process 13,601 variants remained (Appendix D.2)

From this list of more than thirteen thousand variants, only frequently mutated genes (recurrent genes), i.e., those with 5 or more mutated variants affecting 5 or more patients were considered appropriate candidates (Variants list shown in Appendix D.3). They are mutated in 19% or more of the BBM samples, which could be an indication that they may be metastasis-associated genes. From the list the missing variants MAF values were manually retrieved from The Exome Aggregation Consortium (ExAC) database and manually annotated in WES data set.

4.4 The Exome Aggregation Consortium (ExAC)

The ExAC database is a catalogue of whole-exome sequencing data from over 60,000 unrelated individuals of diverse origins from several population studies (Song et al., 2016; Dayem Ullah et al., 2018). Data from individuals affected with severe paediatric conditions were not included in ExAC, thus the data set can also be used as allele frequency reference for paediatric disease studies (gnomad.broadinstitute.org, n.d.).

To search for the MAF frequency value of each variant, the chromosome location, reference, and alternate allele information were entered into the ExAC database to be interrogated using the annotations from the human genome assembly GRCh37/hg19.

After annotating the missing MAF values, MAF ≤ 1% or that have not been reported on ExAC. A total of 1127 variants were retained for further consideration (Appendix D.4 shows the total variants that were retained), as the adoption of a very low MAF cut-off value is considered to provide a more accurate interpretation of sequenced variants (Kobayashi et al., 2017). At this stage of candidate filtering, recurrent genes from this list were retained with a total 985 variants (Appendix D.5.1) of which 846 were missense variants (Appendix D.5.2). Pathogenic prediction of missense variants was carried by interrogating SIFT (https://sift.bii.a-star.edu.sg/) and Polyphen2 (http://genetics.bwh.harvard.edu/pph2/).
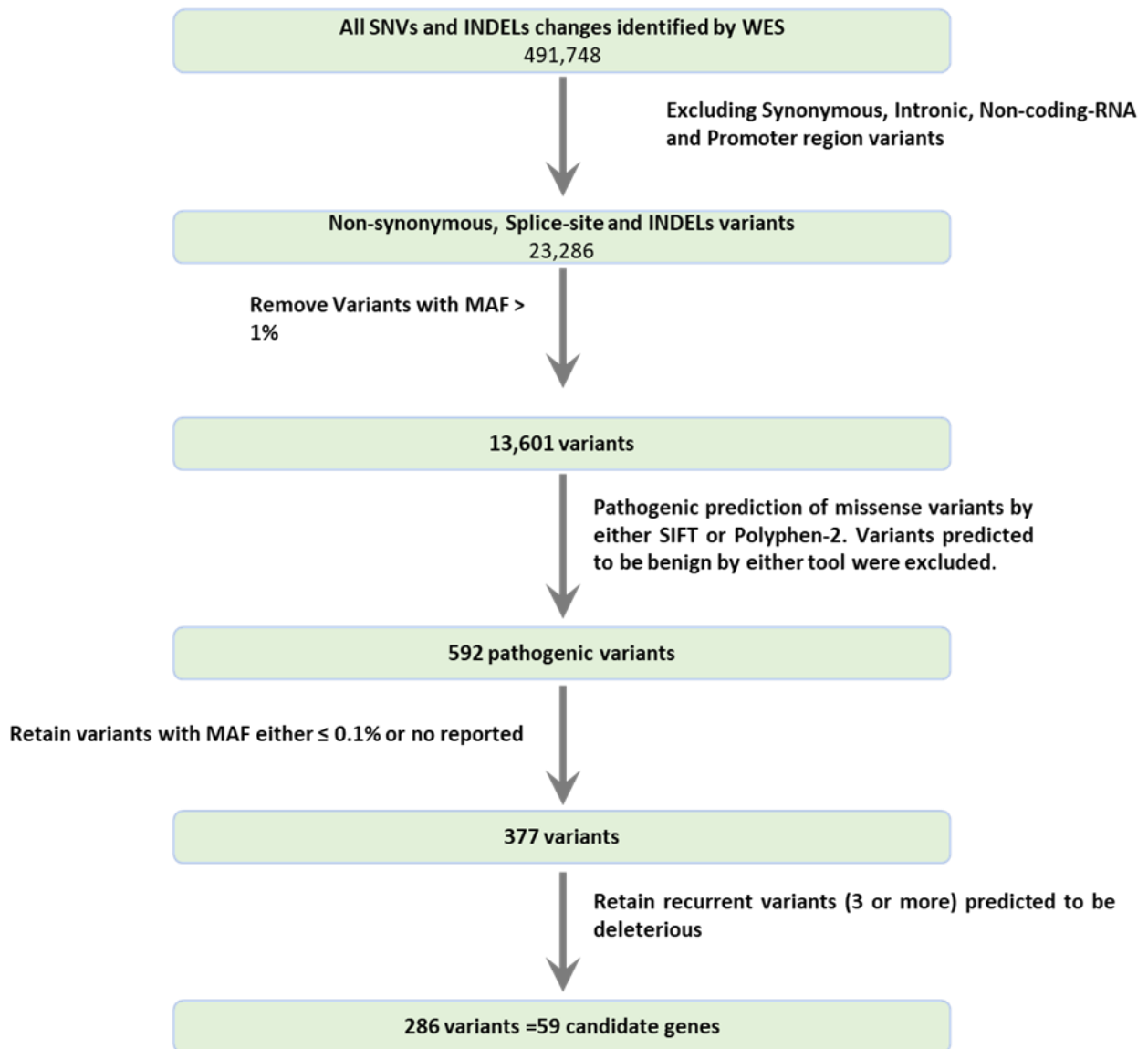
Figure 4.4: WES pipeline summary. The filtering steps of WES data from 26 BBM samples. To reduce the number of variants identified by WES analysis, a stepwise filtering data sorting was carried out. This allowed for the selection of potential metastasis-associated candidate genes. A more detailed schematic representation of the bioinformatic sorting steps is shown in Appendix C.1.

4.5 Predicting the effects on protein function using the SIFT and Polyphen- 2 algorithm

NGS techniques have identified a vast number of variants in cancer studies. However, establishing the importance of those changes can represent a challenge due to the considerable number of data generated by those studies (Ernst et al., 2018). Predicting the potential effect of a missense mutation in a specific protein is crucial to understanding tumour development, progression, and identification of therapeutic targets (Gonzalez-Perez and Lopez-Bigas, 2011; Dong et al., 2014). In silico tools are widely used for determining the possible effect of missense amino acid substitutions (Poon, 2021). Several algorithms have been developed to predict the effect of these changes in the translated protein.

SIFT and Polyphen-2 tools have been built to predict if an amino acid substitution would be deleterious or benign for the protein function (Dong et al., 2014). SIFT performs the predictive assessment by taking into account evolutionary conservation, whereas Polyphen-2 additionally integrates structural information to categorize missense changes as pathogenic or benign (Gnad et al., 2013). SIFT employs multiple sequence homology and the physical and chemical properties between the alternated amino acids to perform a bioinformatic assessment that predicts the effect of an amino acid substitution on the protein function (Vasser et al., 2015). To determine if an amino acid change would be damaging to the protein function the SIFT algorithm assumes that important amino acid localization in a protein structure is highly conserved throughout evolution. Consequently, any substitution at this position is predicted to be deleterious or probably damaging to the translated protein whereas mutations in lower conserve domains are considered to be tolerable or benign (Kumar et al., 2009; Sim

et al 2012). For each amino acid substitution that is analysed with SIFT, a qualitative prediction (deleterious or tolerable) and a score with values within 0 to 1 is generated (Table 4.2) (Ensembl.org, 2014).

The Polyphen-2 algorithm predicts the effect of missense alterations in the protein function using sequence homology, Pfam annotations and protein 3D structure information when available (Adzhubei et al., 2010; Ernst et al., 2018). Like SIFT, Polyphen-2 provides a qualitative prediction and a score (Ensembl.org, 2014).

Name	Category	Qualitaitve prediction	Score	Information Used
SIFT	Functional prediction	Deleterious	< 0.05,	Protein Sequence homology and physico-chemical amino acid properties
		Tolerable	> 0.05-1	
Polyphen-2	Functional prediction	Probaly damging	>0.908	Sequence homology, Pfam annotations and 3D structure information when available
		Possibly damaging	>0.446 and ≤ 0.908	
		Benign	≤ 0.446	

Table 4.2: Summary of bioinformatic prediction tools used in this study to determine the consequence of missense substitutions on the protein function.

From pathogenic prediction of 846 missense variants, 454 were predicted to be "probably damaging" for the resulting protein (Appendix D.5.3). In this study variants predicted to be probably damaging or possibly damaging by Polyphen-2 were referred

to as "PD", SIFT deleterious prediction was reported as "D" (Damaging). Tolerable or benign predictions from either tool were reported as "B" (Benign).

For this study, missense changes predicted to be deleterious for the resulting protein by either tool were retained. Subsequently, genes with 3 or more variants that can be damaging for the translated protein i.e., missense predicted to be damaging by either SIFT or Polyphen-2, exon indels and splice site variants were retained for further consideration. AT this stage, a total of 593 (Appendix D.5.4) variants that were retained

All BBM samples were exome sequenced without matching normal DNA. Consequently, to rule out potential germline polymorphisms only exceedingly rare deleterious variants with either a MAF ≤ 0.1% (0.001) or NR (no reported) on ExAC were kept. Following these filtering steps, genes with 3 or more deleterious variants were retained for further analysis. Genes were arranged into 3 tables: with the chosen candidates being shown in Tables 4.3, 4.5, and 4.7. Additional information presented in these tables was obtained from NCBI (www.ncbi.nlm.nih.gov, n.d.), UniProt (UniProt.org,2019) and Ensembl (www.ensembl.org, 2020). Genes are displayed in ascending order according to protein size.

GENES WITH 5 OR MORE DELETERIOUS VARIANTS								
GENE SYMBOL	PROTEIN NAME	NM_NUMBER	PROTEIN IDENTIFIER	GENE SIZE (BP)	CDS (BP)	PROTEIN SIZE (AA)	TRANSCRIPT ID	BIOLOGICAL FUNCTION
GOLGA8K	Golgin subfamily A member 8K	NM_001282493.2	D6RF30	5174	1893	630	ENST00000512626.3	Golgi organization
HYDIN	Hydrocephalus-inducing protein homolog	NM_001270974.2	F8WD03	21046	1364	725	ENST00000288168.14	Epithelial cell development
KCNN3	potassium calcium-activated channel subfamily N member 3	NM_002249.6	NP_002240.3	13034	2196	731	ENST00000271915.9	protein heterodimerization activity and calcium-activated potassium channel activity
EP400	E1A-binding protein p400	NM_015409.5	A0A0A0MR80	12289	9372	985	ENST00000333577.8	Uncharacterised protein
NEFH	Neurofilament heavy polypeptide	NM_021076.4	NP_066554.2	3795	3063	1020	ENST00000310624.7	peripheral nervous system neuron axonogenesis
TCHH	Trichohyalin	NM_007113.3	NP_009044.2	6985	5829	1943	ENST00000614923.2	Keratinization
SCN10A	Sodium channel protein type 10 subunit alpha	NM_006514.4	NP_006505	6626	5868	1956	ENST00000449082.3	Ion transport, Sodium transport, Transport
RP1L1	Retinitis Pigmentosa 1-Like 1 Protein	NM_178857.6	NP_849188.4	8014	7203	2400	ENST00000382483.4	Required for the differentiation of photoreceptor cells
COL6A3	Collagen alpha-3(VI) chain	NM_004369.4	NP_004360.2	10532	9530	3177	ENST00000295550.9	Cell adhesion
BRCA2	Breast cancer type 2 susceptibility protein	NM_000059.4	NP_000050.3	11954	10254	3418	ENST00000380152.8	Cell cycle, DNA damage, DNA recombination, DNA repair
HSPG2	Basement membrane-specific heparan sulfate proteoglycan core protein	NM_005529.7	NP_005520.4	13176	13173	4391	ENST00000374695.8	Angiogenesis
FAT1	Protocadherin Fat 1	NM_005245.4	NP_005236.2	14776	13764	4588	ENST00000441802.7	Cell adhesion
RYR3	Ryanodine receptor 3	NM_001036.6	NP_001027.3	15568	14610	4870	ENST00000634891.2	Calcium transport, Ion transport
RYR1	Ryanodine receptor 1	NM_000540.3	NP_000531.2	15400	15114	5038	ENST00000359596.8	Calcium transport, Ion transport
TTN	Titin	No available	No available	No available	No available	No available	No available	

Table 4.3: Candidate genes with 5 or more deleterious variants identified by WES of 26 BBM samples. A total of 15 genes were found to have 5 or more deleterious variants within the analysed tumour samples. Presented is information for each gene, including gene symbol, protein name, gene accession number (NM_number), the protein identifier (either NP_number or UniProt code), the coding region (CDS) size, protein size, and transcript ID. Genes are organised in ascending order by protein size.

AMINO ACID CHANGES IN GENES WITH 5 OR MORE DELETERIOUS VARIANTS												
GOLGA8K	HYDIN	NEFH	TCHH	SCN10A	RP1L1	COL6A3	BRCA2	HSPG2	FAT1	RYR3	RYR1
K606Q	R4383P	A304V	L538P	R14L	H542Q	A919V	K169R	S673F	Q843H	S554Y	A882V
A568T	H4270Y	R352S	E1246Q	W358R	D588Y	R1252C	S636T	K1296Q	D1280A	T2508M	Q1568H
P551L	Q4225H	P872A	R1354P	Y724D	R920W	R1632W	A1237P	G2032A	T2377M	E3119K	R1679H
K463E	R3748C		E1357G	W827S	E967K	D2385H	G1529R	V3760M	V3147G	R3194H	P2195L
Q344R	R3481W		R1541H	C1523Y		R2459W	R2034C		E3852D	R4376Q	R3539H
	P2932L		R1843W					S2709C			
	L755F										

Table 4.4 List of Amino acid substitutions (missense mutations) determined to be potentially pathogenic by either SIFT or Polyphen2 in genes with 5 or more deleterious variants.

								GENES WITH 4 DELETERIOUS VARIANTS	

GENE SYMBOL	PROTEIN NAME	NM_NUMBER	PROTEIN IDENTIFIER	GENE SIZE (BP)	CDS (BP)	PROTEIN SIZE (AA)	TRANSCRIPT ID	BIOLOGICAL FUNCTION
TJP3	Tight junction protein 3	NM_001267560.2	NP_001254489.1	3076	2760	919	ENST00000541714.7	Cytoskeletal Signaling and Sertoli-Sertoli Cell Junction Dynamics.
SEMA5A	Semaphorin-5A	NM_003966.3	NP_003957.2	11755	3222	1074	ENST00000382496.10	Differentiation, Neurogenesis
HPS5	Homo sapiens HPS5 biogenesis of lysosomal organelles complex 2 subunit 2	NM_181507.2	NP_852608.1	4840	3390	1129	ENST00000349215.8	May play a role in organelle biogenesis associated with melanosomes, platelet dense granules, and lysosomes.
AATK	Apoptosis associated tyrosine kinase	NM_001080395.3	NP_001073864.2	5461	4125	1374	ENST00000326724.9	Brain development, protein phosphorylation
LMTK3	lemur tyrosine kinase 3	NM_001388485.1	NP_001375414.1	5144	4383	1460	ENST00000598924.1	Protein phosphorylation
ALS2CR11	C2CD6/C2 calcium dependent domain containing 6	NM_001168221.2	NP_001161693.1	5702	5463	1820	ENST00000439140.6	Membrane trafficking and signal transduction
SI	Sucrase-isomaltase	NM_001041.4	NP_001032.2	6012	5481	1827	ENST00000264382.8	polysaccharide digestion, sucrose catabolic process
DYSF	Dysferlin	NM_003494.4	NP_003485.1	6952	6240	2080	ENST00000258104.8 DYSF-201	Plasma membrane organization
CACNA1H	calcium voltage-gated channel subunit alpha1 H	NM_021098.3	NP_066921.2	8219	7062	2353	ENST00000348261.11	Regulation of ion transmembrane transport
LRBA	Lipopolysaccharide-responsive and beige-like anchor protein	NM_006726.4	NP_006717.2	10353	8589	2863	ENST00000357115.9	Protein localization to phagophore assembly site, mitophagy
FAT2	Protocadherin Fat 2	NM_001447.3	NP_001438.1	14710	13047	4349	ENST00000261800.6	Homophilic cell adhesion via plasma membrane adhesion molecules
DNAH2	Dynein axonemal heavy chain 2	NM_020877.5	NP_065928.2	14563	13284	4427	ENST00000572933.6	ATP hydrolysis activity and microtubule motor activity
APOB	Apolipoprotein B-100	NM_000384.3	NP_000375.3	14121	13689	4563	ENST00000233242.5 APOB-201/ ENST00000399256.4 APOB-202	Cholesterol metabolism, Lipid metabolism, Lipid transport, Steroid metabolism, Sterol metabolism, Transport
ABCA13	ATP-binding cassette sub-family A member 13	NM_152701.5	NP_689914	17188	15174	5058	ENST00000435803.6	ceramide transport. Positive regulation of cholesterol transport
FCGBP	Fc gamma binding protein	NM_003890.2	NP_003881.2	16407	16218	5405	ENSG00000275395	May be involved in the maintenance of the mucosal structure as a gel-like component of the mucosa
KMT2D	Histone-lysine N-methyltransferase 2D	NM_003482.4	NP_003473.3	20635	16611	5537	ENST00000301067.12	Transcription, Transcription regulation
HMCN1	Hemicentin-1	NM_031935.3	NP_114141.2	18368	16905	5635	ENST00000271588.9	Cell cycle, Cell division, Sensory transduction, Vision

Table 4.5: Candidate genes with 4 deleterious variants identified by WES of 26 BBM samples. In total, 17 genes were found to have 4 deleterious variants within the analysed tumour samples. Presented information for each include gene symbol, protein name, NCBI gene ascension number (NM_number), the protein identifier (either NP_number or UniProt code), the coding region (CDS) size, protein size, transcript ID and the protein biological function. Genes are organised in ascending order by protein size.

AMINO ACID CHANGES IN GENES WITH 4 DELETERIOUS VARIANTS												
SEMA5A	HPS5	AATK	ALS2CR11	SI	CACNA1H	LRBA	FAT2	DNAH2	APOB	FCGBP	KMT2D	HMCN1
N210K	M166I	P428T	M330T	S431F	P686L	S653F	R133C	R1626W	I160M	P1150L	P998T	D201G
N595K	K755R	E787K	E1065K	V577G	A690V	F850S	A3383V	V2418E	E605K	A1514V	V2052I	V707F
T892M	E1016K	A1128D		R596T	R1105C	Q1396E	Y4193N	A3431T	S2682I	S1633F	P3665A	A1433T
A1003V	M115I			V1334G	P1912R		R4255W	R3524Q	I4341N	R5341C		G3566D

Table 4.6. List of Amino acid substitutions (missense mutations) determined to be potentially pathogenic by either SIFT or Polyphen2 in genes with 4 deleterious variants.

GENES WITH 3 DELETERIOUS VARIANTS								
GENE SYMBOL	PROTEIN NAME	NM_NUMBER	PROTEIN IDENTIFIER	GENE SIZE (BP)	CDS (BP)	PROTEIN SIZE (AA)	TRANSCRIPT ID	BIOLOGICAL FUNCTION
ATXN1	ataxin 1	NM_000332.4	NP_000323.2	10606	2448	815	N/A	Negative regulation of transcription, negative regulation of transcription by RNA polymerase II
KIAA1522	Uncharacterized protein KIAA1522	NM_020888.3	Q9P206	5493	3282	1094	ENST00000401073.7	Cell differentiation
LAMC2	laminin subunit gamma 2	NM_005562.3	NP_005553.2	5398	3582	1193	ENST00000264144.5	Cell adhesion, differentiation, migration, signaling
ZNF208	zinc finger protein 208	NM_007153.3	NP_009084.2	9088	3843	1280	ENST00000397126.9	Transcription regulation by RNA polymerase II
DSPP	Dentin sialophosphoprotein	NM_014208.3	NP_055023.2	4331	3906	1301	ENST00000651931.1	Biomineral tissue development
CEP170B	centrosomal protein 170B	NM_001112726.3	NP_001106197.1	6727	4665	1554	ENST00000414716.8	microtubule organization
SSC5D	scavenger receptor cysteine rich family member with 5 domains	NM_001144950.2	NP_001138422.1	4977	4722	1573	ENST00000389623.11	Innate immune response
CPAMD8	C3 and PZP-like alpha-2-macroglobulin domain-containing protein 8	NM_015692.5	N/A	5983	5659	N/A	ENST00000651564.2	Eye development
KIAA1549	KIAA1549	NM_020910.3	NP_065961.2	12450	5805	1934	ENST00000440172.5	May play a role in photoreceptor function
ATG2A	Autophagy-related protein 2 homolog A	NM_015104.3	NP_055919.2	6318	5814	1938	ENST00000377264.8	Autophagosome assembly,
KIAA1244	Brefeldin A-inhibited guanine nucleotide-exchange protein 3	NM_020340.5	NP_065073.3	14859	6531	2177	ENST00000251691.5	Actin cytoskeleton organization, regulation of ARF protein signal transduction
ABCA4	ATP binding cassette subfamily A member 4	NM_000350.3	NP_000341.2	7328	6822	2273	ENST00000370225.4	Lipid transport, photoreceptor cell maintenance
CACNA1H	Voltage-dependent T-type calcium channel subunit alpha-1H	NM_021098.3	Q9P0X4	8219	7059	2353	ENST00000348261.11	Regulation of ion transmembrane transport
SPTA1	Spectrin alpha chain	NM_003126.4	NP_003117.2	8018	7257	2419	ENST00000643759.2	Regulation of cell shape
NCOR2	Nuclear receptor corepressor 2	NM_001206654.2	NP_001193583.1	8475	7512	2504	ENST00000429285.6	Negative regulation of transcription, DNA-templated, negative regulation of androgen receptor signalling pathway
COL6A5	Collagen alpha-5(VI) chain	NM_001278298.2	NP_694996.5	8969	7845	2526	ENST00000312481.11	Cell adhesion

Gene	Protein name	NM_number	Protein ID	CDS	Protein size	Transcript ID	Function
IGSF10	Immunoglobulin superfamily member 10	NM_178822.5	NP_849144.2	8562	7872	2623 ENST00000282466.4	Cell differentiation, regulation of neuron migration
FLNC	Filamin-C	NM_001458.5	NP_001449.3	9159	8175	2725 ENST00000325888.13	Ssarcomere organization
SRRM2	serine/arginine repetitive matrix 2	NM_016333.4	NP_057417.3	9044	8259	2752 ENST00000301740.13	mRNA splicing, via spliceosome
TNRC18	Trinucleotide repeat-containing gene 18 protein	NM_001080495.3	NP_001073964.2	10839	8904	2968 ENST00000430969.6	Chromatin binding and transcription cis-regulatory region binding
PRUNE2	Protein prune homolog 2	NM_015225.3	NP_056040.2	12612	9267	3088 ENST00000443509.6	Apoptotic process
DNAH12	Dynein axonemal heavy chain 12	NM_001366028.2	Q6ZR08	12145	11883	3092 ENST00000351747.6	Microtubule-based movement
LAMA2	Laminin subunit alpha-2	NM_000426.4	NP_000417	9696	9366	3122 ENST00000421865.3	Cell adhesion
EYS	Eyes shut homolog	NM_001142800.2	NP_001136272.1	10590	9435	3144 ENST00000503581.6	Detection of light stimulus involved in visual perception, skeletal muscle tissue regeneration
MYO15A	Unconventional myosin-XV	NM_016239.4	NP_057323.3	11811	10590	3530 ENST00000647165.2	Actin filament organization, inner ear morphogenesis
CUBN	Cubilin	NM_001081.4	NP_001072.2	11927	10872	3623 ENST00000377833.10	Lipoprotein, vitamin and iron metabolism
FLG	Filaggrin	NM_002016.2	NP_002007.1	12793	12183	4061 ENST00000368799.2	Establishment of skin barrier
PKD1	polycystin 1, transient receptor potential channel interacting	NM_000296.4	P98161	14137	12909	4303 ENST00000262304.9	Cell adhesion, epithelial development
DYNC2H1	dynein cytoplasmic 2 heavy chain 1	NM_001377.3	NP_001368.2	13683	12924	4307 ENST00000375735.7	Cilium assembly, golgi organization
STARD9	StAR related lipid transfer domain containing 9	NM_020759.3	NP_065810.2	15637	14103	4700 ENST00000290607.12	Microtubule-based movement, spindle assembly
SYNE1	spectrin repeat containing nuclear envelope protein 1	NM_182961.4	NP_892006.3	27708	26394	8797 ENST00000367255.10	Spermatogenesis, golgi organization

Table 4.7: Candidate genes with 3 or more deleterious variants identified by WES analysis of 26 BBM samples. A total of 31 genes were found to have 3 deleterious variants within the analysed tumour samples. Table information for each includes gene symbol, protein name, NCBI gene ascension number (NM_number), the protein identifier (either NP_number or UniProt code), the coding region (CDS) size, protein size, transcript ID and the protein biological function. Genes are organised in ascending order by protein size.
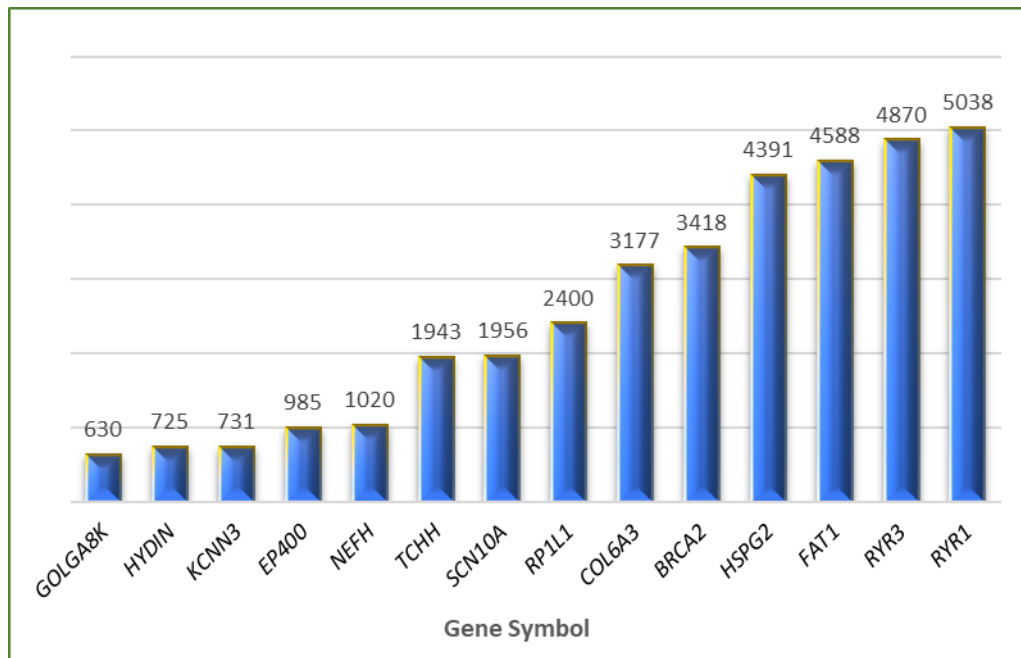
Figure 4.5: Protein size of candidate genes with 5 or more probably damaging variants. 14 candidate genes with 5 or more deleterious variants displayed in ascending order according to the protein size indicated by numbers (number of amino acids) at the top of the data bars.
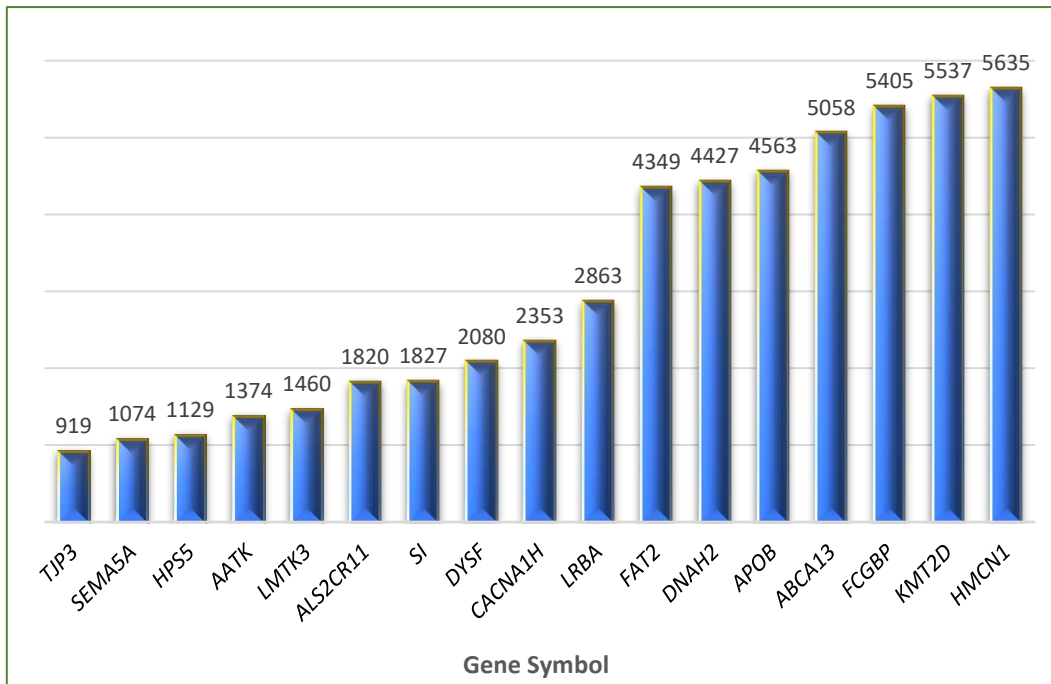
Figure 4.6: Protein size of candidate genes with 4 probably damaging variants.

17 candidate genes with 4 deleterious variants displayed in ascending order according to the protein size indicated by numbers (number of amino acids) at the top of the data bars.
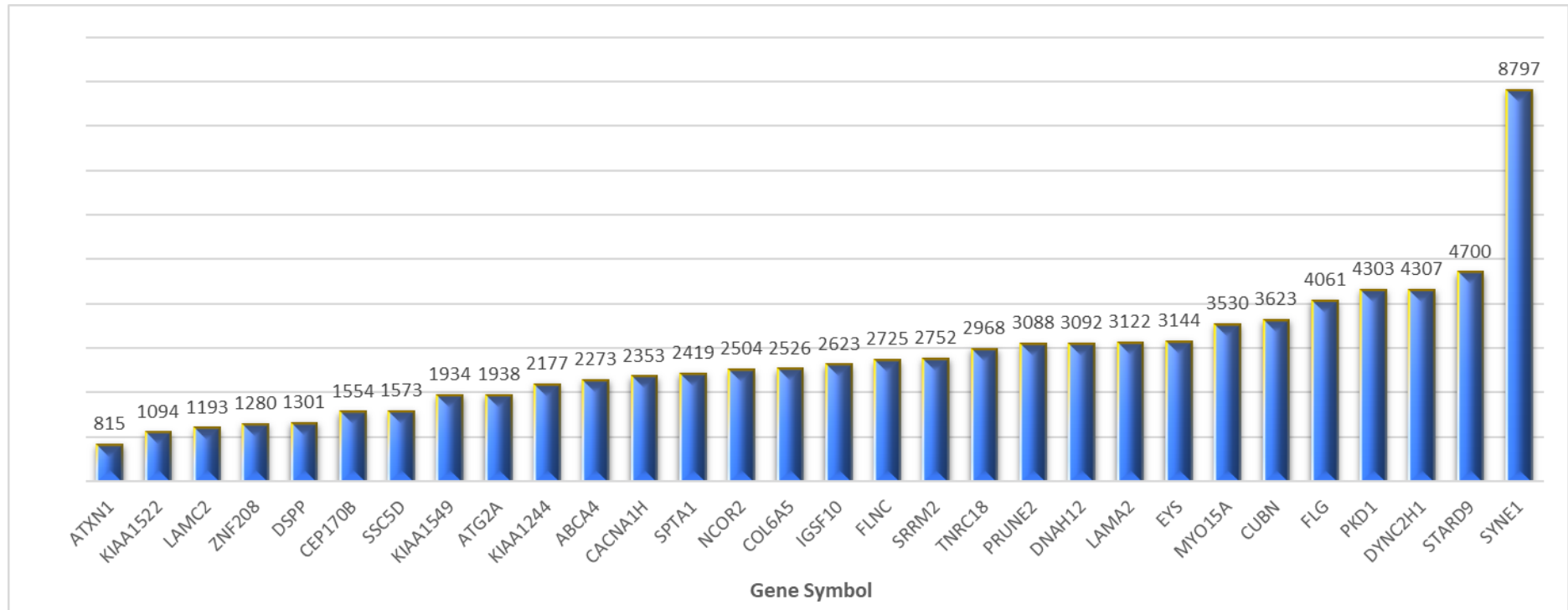
Figure 4.7: Protein size of candidate genes with 3 probably damaging variants. 30 candidate genes with 3 deleterious variants

displayed in ascending order according to the protein size indicated by the amino acid number at the top of the data bars.

4.6 Confirmation of variants by Sanger sequencing

The use of NGS in cancer research has been essential in obtaining a deeper understanding of the complexities underlying tumour development and progression. However, NGS is not error-free. Correctly filtering the large amount of data generated by these studies and the subsequent validation is essential to be able to correctly evaluate and characterize these results (De Cario et al., 2020).

Inaccurate results can be due to several factors such as the quality of the sequenced samples, and processing and storage conditions. Identifying clinically relevant variants from background "noise" can represent a challenge, as a consequence of poor targeting such as low read depth, especially in GC- enriched areas that could result in false-positive read, inadequate exon coverage that can produce false-negative results and inaccurate alignment to the reference genome (Luthra et al., 2015, Di Resta et al., 2018; Hu et al., 2021; Horgan et al., 2022). In the current study Whole Exome Sequencing was carried out at high read depth in order to minimise the likelihood of getting false positives.

Optimised filtering of WES data has been designed to reduce the number of false-positive variants. However, there are no internationally established validation protocols for WES analysis, hence Sanger sequencing validation is still widely used for variant confirmation (Chang et al., 2017; De Cario et al., 2020; Fernandez - Rozadilla 2021; Lascar et al., 2022).

To validate WES data, validation of mutated variants was carried out using Sanger sequencing (Sanger sequencing electropherograms are shown in Appendix E and in Chapter 8 BIG3 gene sequencing). Custom primers to flank the desired region were designed following the guidelines previously described (Materials and Methods

section Chapter 2) and DNA was amplified by PCR. Afterwards, the products were sequenced using chain termination cycle sequencing kit v3.1 (Applied Biosystems) following the manufacturer's protocol ass described in the Materials and Methods Chapter 3.

In this study some variants were validated because several studies have shown that Sanger sequencing in not completely necessary when validating WES data as consistently, results from next generation technique studies have been largely accurate (McCourt et al., 2013; Strom et al., 2014; Beck, Mullikin and Biesecker, 2016; De Cario et al., 2020). Additionally, Sanger sequencing is expensive and time consuming so it would have been a colossal effort to validate all of the candidate variants identified in this study using WES.

Furthermore, this study, due to the considerable number of variants identified by WES, independent variants present in more than 3 BBM samples were excluded from the list of potential candidate variants as they were more likely to be false positive or common polymorphisms.

4.7 Discussion

To assess if the BBM tumours from this study harbour some of the same genomic characteristics as primary breast tumours, the 20 most mutated genes in primary breast cancer were searched for within the WES of the 26 BBM of this cohort. Known primary breast cancer drivers were found mutated in across the 26 BBM samples from this study, including PIK3CA, Tp53, CDH1 and KMT2C. Most of the changes were concentrated outside the coding exon. However, variants, clustered into 15 genes were within the exons or at splice site regions (KMT2C, ZFHX3, LRPB1, PIK3CA, ALK,

GATA3, PTPRT, MAP3K1, GRIN2A, NF1, MED12, ESR1, PTEN, ARID1A, ERBB2, KMT2C, ZFHX3, LRPB1, PIK3CA and ALK).

Several studies have been carried out to determine the mutational landscape of brain metastases evolving from primary breast tumours (Cosgrove et al., 2022; Morgan, Giannoudis and Palmieri, 2021; Ali et al., 2021). In summary, a recent meta-analysis of the genomic landscape of breast to brain metastases, the data from 13 primary breast sequencing studies comprised of 164 brain metastases were compared. 268 genes were found mutated in BBM samples of which 22 (8%) were reported mutated in five or more BBM including TP53 (52%), PI3KCA (22%), KMT2C (6%) and ZFHX3(5%) (Morgan, Giannoudis and Palmieri, 2021). These are well-documented tumour initiator genes that are likely to drive primary tumour growth and then persist in metastasising cells (Bailey et al., 2018; Porta-Pardo, Valencia and Godzik, 2020).

PIK3CA is part of the family of the Phosphatidylinositol 3-kinase. This protein family is critical for the regulation of cellular growth, transformation, adhesion, apoptosis, survival, and motility (Karakas, Bachman, and Park 2006). Mutations of this gene have been reported to be present in over one-third of primary breast cancer subtypes (Zardavas, Phillips and Loi, 2014; Dirican, Akkiprik and Özer, 2016).

An immunohistochemistry study by Adamo and colleagues on 52 BBM and 12 matched primary tumours found that PIK3CA was active in most BBM regardless of the breast cancer subtype (Adamo et. al., 2011; Batalini et al. 2020). In a retrospective study of 307 HR+/HER2- breast cancer patients they found that PIK3CA mutations in primary breast cancer may be associated with the progression to brain metastasis as they reported that patients with PIK3CA mutations had higher incidence of brain metastasis (Fitzgerald et al., 2019).

TP53 mutations have been characterised in a large number of cancer studies. TP53 is a tumour suppressor gene vital for the control of numerous cellular pathways associated with proliferation and survival (Reinhardt and Schumacher, 2012). A study determined that important metastatic processes such as cell motility and adhesion were influenced by TP53 binding to the promoter regions of the gene (Powell, Piwnica-Worms, and Piwnica-Worms, 2014). TP53 mutations might play a role beyond tumour initiation and progression, they might also help tumours to develop genomic characteristics that will promote the growth of the metastatic malignancy. A mutational profiling study of 18 primary breast tumours and 42 brain metastases from breast tumours (15 with matched pairs) found PIK3CA and TP53 genes to be frequently mutated in both the primary and the metastatic malignancy (Lee et. al., 2015).

KMT2C is a histone methyltransferase protein with an essential role in the gene expression regulation. This protein is frequently found to be dysregulated in cancer and somatic mutations in these and other histone modifier proteins have been linked to oncogenesis by tumours genome and exome sequencing studies (Fagan and Dingwall, 2019; Prado, Bennett, and Licht, 2022). In the present analysis, of the 20 genes found commonly in primary breast cancer, 168 KMT2C variants were identified across the 26 BBM samples. Most variants were found in the non-coding region of the exons. However, nineteen mutations were found to be located within the exon or the splice-site region which suggest that these mutations may have been implicated in the development of the brain tumour. A Whole Exome Sequencing paired analysis of lung cancer and brain metastases identified KMT2C mutations in 25% of lung cancer and 50% of brain metastases (Liu et.al., 2021). These findings suggest that KMT2C mutations may be positively selected during metastatic tumour evolution. The presence of these mutations in the 26 BBM supports the premise that brain

metastases carry some of the mutations commonly found in the primary malignancy. These mutations may have been necessary for tumour evolution from breast cancer to brain metastases promoting survival and proliferation of prometastatic cells and the establishment of the brain tumour. However, additional new genomic changes may have been necessary to support tumour colonisation of the brain as the tumours cell have to adapt brain microenvironment (Yates et al., 2017; Fecci et al., 2019)

Even though the top 20 genes mutated in primary breast cancer were frequently mutated in the BBM samples of this study they were not considered for further analysis as these genes have been widely characterised in primary breast tumours and metastatic tumours and this study aimed to identify rare novel genomic alterations that may be metastasis-specific changes. Therefore, to characterise genomic changes that may be metastasis-specific, variant annotations were used to prioritise variants. There are many tools accessible to aid with the prioritisation of variants and selection of cancer-specific candidates such as the 1000 Genomes project (Auton et al.,2015; Aganezov et al., 2022) and the ExAC database (Lek et al., 2016; Guo et al.,2018). This study is comprised of a cohort of 26 BBM samples and due to the small number of matched normal tissue to confirm the somatic or germline nature of the variants, filtering variants by frequency is not completely informative to rule out common variants in the general population. Therefore, the MAF reported for each variant on the ExAC database was used as an external control to exclude variants frequently found in the general population. A MAF of <1% is considered an appropriate cut-off value to identify uncommon variants that might be disease-specific (Rabbani, Tekin and Mahdieh, 2013; Niroula and Vihinen, 2019).

It has been demonstrated that metastatic tumours continue to evolve and develop genomic aberrations that were not present in the primary tumours and that these

changes allow them to adapt to sustain growth and proliferation in the tissue microenvironment ((Boire et al., 2019; Craig et al., 2020). Cancer evolution results from the accumulation of oncogenic changes (Hanahan et al, 2022). We hypothesised that metastasis-specific gene mutations will be infrequently found in the primary tumour as they would have developed after the primary tumour cells have left the originating site. In order to identify rare genomic changes that may be involved in brain tumour development and colonisation, a MAF ≤0.1% was applied to the recurrent pathogenic variants. Variants with a MAF value less or equal to 1% considered uncommon in the general population (Rabbani, Tekin and Mahdieh, 2013; Lee et al., 2015). However, these analyses usually have a larger cohort consisting of paired primary tumours and brain metastases, and sometimes matched normal tissue samples which allow more accurate characterisation of the genomic landscape of both tumours. A comparative whole-exome sequencing study of primary lung cancer and matched brain metastasis found mutational discrepancies between the malignancies. More mutations were found in the metastasis than in the primary tumour including genes previously described as either lung cancer-associated (KRAS, ROS1 and STK11) or metastasis-specific genes (CCDC178, RUNX1T1, MUC2) (Tomasini et.al., 2020). Following the application of the inclusion and exclusion criteria, 286 variants were categorised into 63 candidate genes. All the genes were considered to be suitable metastasis-specific candidates as they have 3 or more deleterious variants with a MAF≤0.1% which makes them infrequent in the general population. Initially, from this extensive list of candidates, genes that had more than 5 mutated variants were considered fitting candidates because these mutations were frequent in the BBM but not in the general population. Initially, 16 genes had 5 more damaging variants. However, EP400, and KCNN3 were removed from the list as there was inconsistency

between the transcripts from the WES data and the annotations on Ensembl, NCBI and UniProt. Subsequently, twelve genes, BRCA2, HYNDIN, NEFH, RP1L1, TCHH, COL6A3, RYR1, RYR3, HSPG, SCN10A, FAT1, and GOLGA8K were the most frequently mutated across the 26 BBM samples with either MAF ≤0.1% or found not reported on the EXAC database and predicted to be deleterious for the resulting protein. Interestingly, in a systemic review of the genomic landscape of breast to brain metastatic tumours it was found that BRCA2, FAT1 and COL6A3 are associated with 22% of breast to brain metastases. They compiled immunohistochemistry, copy number alteration and Prediction Analysis of Microarray 50 data of breast to brain metastasis studies (Morgan, Giannoudis and Palmieri, 2021). Which suggests that these genes can be suitable metastasis-associated genes (Corti et al.,2022)

From the list of genes with 4 deleterious variants, KMT2D was selected as a candidate of interest as it belongs the family of histone–lysine N-methyltransferase (KMT2). Mutations in this protein family are among the most common genomic mutations in cancer (Xie et al., 2022; Anjanappa et al., 2017). In the present study, rare, mutated variants in KMT2D were found in 4 out of 26 (15%) BBM samples. KMT2D mutations were reported in 4% of breast-to-brain metastases in a review of the current treatment landscape of BBM (Corti et al.,2022). A whole-exome sequencing study of metaplastic or sarcomatoid carcinoma from different organs (breast, oesophagus, lung, and kidney) revealed frequent mutations of KMT2D in tumours but not in the normal tissue. These changes correlated with low KMT2D gene expression, large tumour size and adverse prognosis (Zheng et al., 2021). Knockout experiments in mice have shown that KMT2D loss results in formation of lymphomas (Zhang et al., 2015; Ortega-Molina et al., 2015). It can be inferred that the alterations identified in the BBM samples may be implicated in tumour development.

Additionally, from the list of genes with 3 damaging variants, KIAA124/ARFGEF3/BIG3 were prioritised for further analysis because it has been described that BIG3 is involved in the regulation of neurotransmitter receptor activity (Liu et al., 2016). Liu and colleagues reported that BIG3 knockout mice showed enhanced inhibitory synaptic activity that resulted from upregulation in postsynaptic GABA receptor activity. Additionally, three separate studies have described that the way tumour cells colonise the brain is by modulating the expression of neurotransmitter receptors (Venkatesh et al., 2019; Venkataramani et al., 2019; Zheng et al. 2019). Thus, these findings suggest that BIG3 may be a metastasis-associated gene.

4.8 Conclusion

This study analysed the sequencing data of 26 BBM samples. The data analysis was limited in part by not having matching normal tissue for most of the samples used in this research. The addition of normal tissue sequencing analysis would have made it possible to accurately filter out germlines from somatic mutations. However, gathering the appropriate set sample for cancer studies can be difficult due to the amount of time between the primary malignancy diagnosis and the metastatic tumour development. Also, to add statistical significance to the identified variants, a bigger number of sample sets that include tumours and matched normal tissue would need to be sequenced to generate a larger cohort study.

In summary, known mutated drivers of primary breast cancer along with several potential metastasis-specific genes were identified by WES of 26 BBM samples. These changes are hypothesised to be the result of late genomic evolution that might have been necessary for metastatic tumour progression and adaptation to the brain microenvironment.

Chapter 5 - Amino Acid mutation signature of metastasis-associated genes

5.1 Introduction

In the last decade, cancer genome studies have made possible the identification of cancer "driver" mutations that have helped to characterize the molecular mechanisms involved in cancer development and progression (Sondka et al., 2018; Dietlein et al., 2020; Martínez-Jiménez et al., 2020). Cancers are classified according to the tissue of origin or their nucleotide substitution signature, however, these classifications lack understanding of how cancer cell adaptation takes place (Van Hoeck et al., 2019). It has been described that cell adaptation is determined at the protein level, and that to understand how cancer cells are able to survive and thrive in different tissues it is necessary to identify the changes at the proteomic level that could be driving these processes (Szpiech et al., 2017). Therefore, describing the amino acid mutation signature of metastatic cancers might lead to a better understanding of the mutational patterns involved in cancer progression and proliferation as well as to provide new diagnostic markers and therapeutic targets.

Here is presented an analysis of the amino acid substitutions identified in 14 metastases-associated genes and how these alterations might be relevant for the metastatic process.

5.2 Results

BIG3	BRCA2	KMT2D	GOLGA8K	NEFH	COL6A3	TCHH	HSPG2	SCN10A	FAT1	RP1L1	RYR3	HYDIN	RYR1
A1329V	A1237P	M3398V	A568T	A304V	A919V	E1357G	A4071V	C1523Y	D4218G	D588Y	E3119K	F4120L	A882V
C1058F	G1529R	P3665A	E447Q	P872A	D2385H	E1246Q	A3396V	P1045T	D1280A	E967K	R3194H	H4270Y	D3310E
R1582T	K169R	P2557L	E434D	R352S	R3088K	E494Q	G2950R	R14L	E3852D	H542Q	R4376Q	K2144E	E3583Q
V1203G	K3326*	P2382S	K606Q		R2459W	L538P	G2032A	W827S	K835Q	Q2010*	S554Y	L755F	P2195L
V1510A	R2034C	P998T	K463E		R1632W	Q1535E	K1296Q	W358R	Q843H	R1622Q	T2508M	M524I	Q1568H
V1598M	S636T	P630S	M455V		R1252C	Q277H	P1019L	Y724D	T2377M	R920W		P2932L	R3539H
V2162L	S2697N	V2052I	P551L			R1843W	Q2618L		V3147G			Q4225H	R1679H
	S2709C		Q344R			R1541H	R2977W		Y1250C			R4383P	
	P1088S		V602M			R1354P	R2004H					R3748C	
						R820W	R1919C					R3481W	
						R552S	R1090Q					S2047T	
							S673F					S1344T	
							T361N						
							V3760M						
							W444*						

Table 5.1: Amino acid mutation pattern of BBMs. Distribution of 111 amino acid substitution among the 14 breasts to brain metastasis-associated genes identified in this study. The most common substitution is arginine (R) for tryptophan (W), followed by exchange of alanine (A) for valine (V). substitutions present in more than one gene are shown in bold.
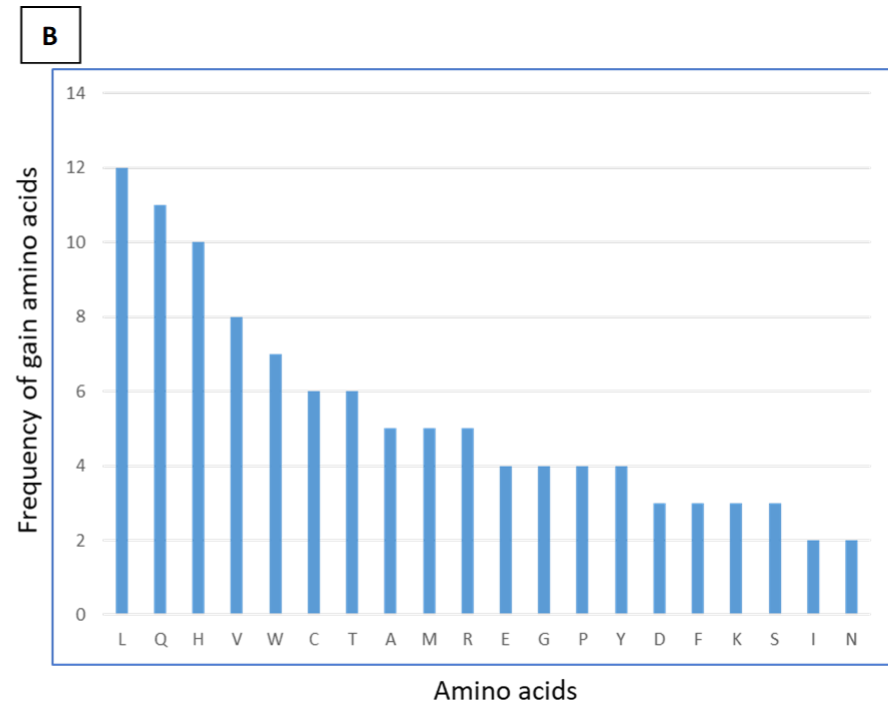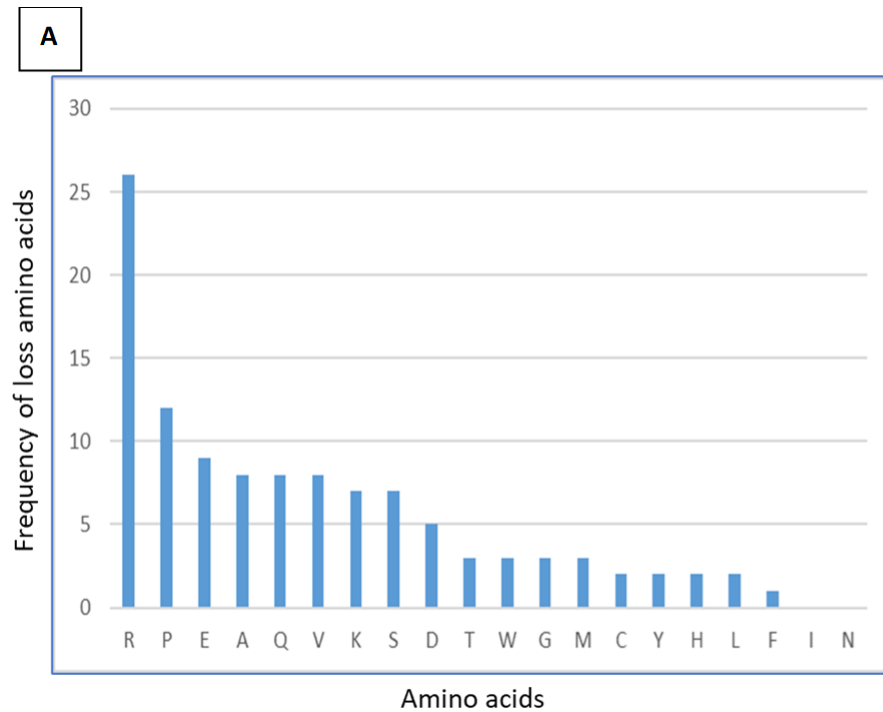
Figure 5.1: Amino acid changes identified in BBM samples following whole exome sequencing data. Overview of the affected amino acids by nonsynonymous mutations in the 14 analysed proteins. For this analysis, all amino acid substitutions regardless of their pathogenicity were counted. A) Frequency of amino acid loss in the analysed 14 proteins. Most mutations resulted in the loss of arginine (R), proline (P) and glutamic acid (E). B) Frequencies of gained amino acids in the analysed proteins. The most frequently

gained amino acids were leucine (L), glutamine (Q) and histidine (H). For this analysis, all amino acid substitutions regardless of their pathogenicity were considered. Data is presented in descending order from left to right.
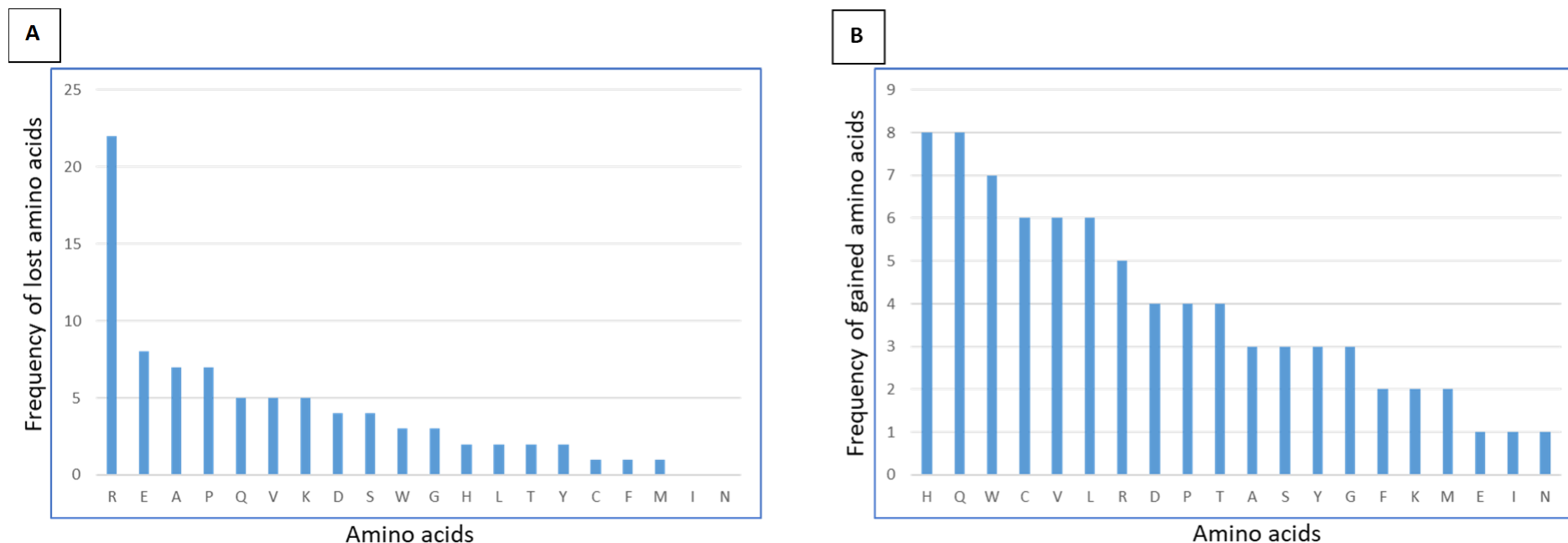
Figure 5.2: Frequency of amino acid changes predicted pathogenic. Overview of the affected amino acids by nonsynonymous mutations predicted to be pathogenic for the protein consequence by either SIFT or Polyphen-2. A) Arginine (R), glutamic acid (E), alanine (A) and proline (P) were the most frequently affected amino acids. B) Frequencies of amino acids gain in the analysed 14

proteins. Histidine (H), glutamine (Q) and tryptophan (W) were the most frequently gained amino acids. Data is presented in descending order from left to right.
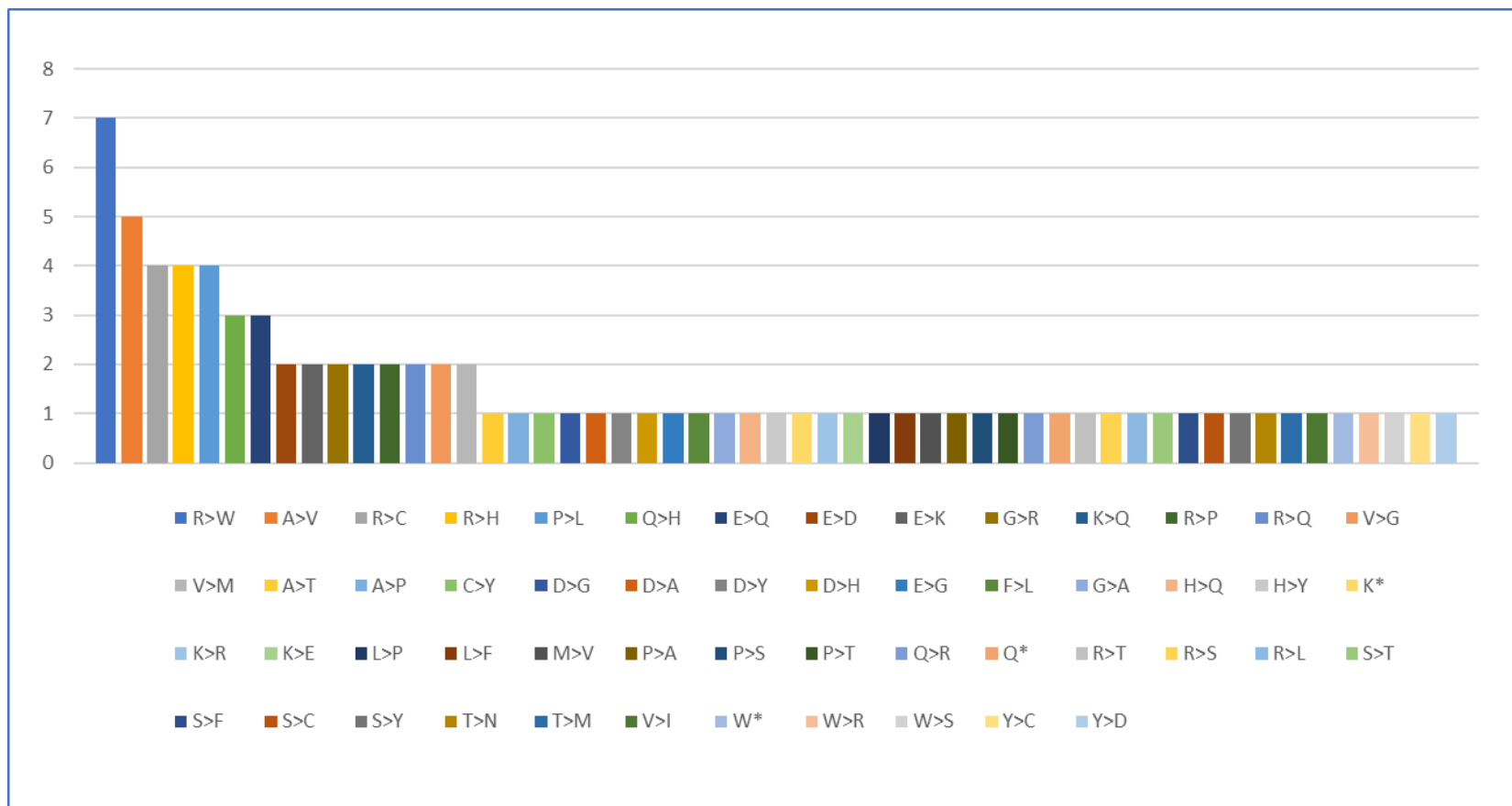
Figure 5.3: Frequency distribution of damaging amino acid changes. Analysis of amino acid substitutions present in the 14 analysed proteins that were predicted to be pathogenic by either SIFT or Polyphen-2. Changes of arginine>tryptophan, alanine>valine, arginine to either cysteine (C), histidine (H), and proline>leucine were the most frequent substitutions in the analysed proteins. Data is presented in descending order from left to right. *Nonsense amino acid mutations.

5.3 Discussion

Understanding the biological processes involved in cancer development and progression is of utmost importance. In the past years, genome databases such as TCGA and CoSMIC, have compiled the data from different cohorts' studies, facilitating the analysis of cancer cells' mutational patterns (Hutter and Zenklusen, 2018; Tate et al., 2019). These have provided insight into the nucleotide substitutions and corresponding amino acid changes of somatic mutations in cancer and also how these changes regulate cancer cell behaviour (Blokzijl et al., 2018; Tan, Bao, and Zhou, 2015; Alexandrov et al., 2013). These studies have been carried out using data generated from the primary tumour or paired normal-primary tumour sequencing. In the case of metastases, analysing amino acid substitutions can provide valuable information on the role of these alterations in metastatic cancer behavioural patterns, as it has been determined that cellular adaptations occur at the proteomic level, and that this is not different for cancer cells (Tsuber et al., 2017). These changes allow the cells to adapt variations in intracellular processes, such as changes in metabolism, intracellular pH, matrix composition, availability of oxygen and nutrients as well as intracellular pH level variations (White, Grillo-Hill and Barber, 2017).

The present analysis highlighted a set of possible amino acid mutation patterns leading to the mutational landscape of 14 breast to brain metastases-associated genes, which were identified in this study by Whole Exome Sequencing. 111 Amino acid substitutions were computed for these 14 candidate genes (Table 5.1). Overall, loss of arginine (R) followed by loss of proline (P) and glutamic acid (E) and significant gain of leucine (L), glutamine (Q) and histidine (H) dominate the landscape of amino acids introduced by mutations (Figure 5.1). Following prediction by either SIFT or Polyphen-2 the list of substitutions reduced to 84 pathogenic amino acid changes with

arginine still being the most commonly lost amino acid, followed by loss of glutamic acid (E) and alanine (A) with gains of histidine (H), glutamine (Q) and tryptophan (W) (Figure 5.2). Substitution distribution showed that histidine gain was mainly due to loss of either arginine or glutamine, whereas glutamine gain resulted primarily from loss of glutamic acid, lysine, or arginine (Figure 5.3). In addition, tryptophan gain was exclusively due to loss of arginine. Other amino acids were lost and gained, although at a lower rate (Figure 5.2).

Arginine is a positively charged semi-essential amino acid with a significant role in gene expression, protein structure and function, and genome evolution (Nelakurti et al., 2021). Arginine loss is considered to be essential for cancer development and progression as it has been found to be depleted in genes with lower tumour suppressor activity (Tsuber et al., 2017). This amino acid is coded by six different codons and base substitutions at any of the six codons can result in mutations that can modify the protein structure and function (Chen et al., 2021). Substitutions of arginine for cysteine, histidine, glutamine, and tryptophan represent around 75% of all amino acid substitutions in cancer, altering the genome and mitochondria protective functions (Anoosha, Sakthivel and Michael Gromiha, 2016).

Szpiech and colleagues published data in 2017 in which they utilised a non-negative matrix factorisation (NMF) approach to filter and analyse the amino acid mutation characteristics of a cohort of tumour-normal paired samples across 29 primary cancers. The study found that amino acid substitution introduced by mutations were dominated by arginine loss, and that this loss resulted in gain of glutamic acid, histidine, and tryptophan.

A further study carried out by Tsuber and colleagues in 2017 aimed to determine the somatic evolution of cancer. They analysed CCLE (Cancer Cell Line Encyclopaedia) data of amino acid changes that resulted from single nucleotide substitutions in genes that coded for 2164 proteins, which correspond to around a tenth of the human proteome. This analysis found that cancer mutations resulted in a net loss of arginine with gains of cysteine, histidine, and tryptophan. Importantly, to determine if the losses and gains of specific amino acids apply to other cancer types, they analysed CoSMIC data of all nonsynonymous amino acid changes caused by mutations of single nucleotide substitutions. They concluded that the observed amino acid substitutions are universal, meaning that all cancer tissues undergo marked loss of arginine with significant gains of cysteine, histidine, and tryptophan. Loss of arginine was markedly high in tumour suppressor proteins coded by TP53, OBSCN, LRP1B, and TTN, among others.

The amino acid changes identified by the previously mentioned studies and by the present analysis in breast to brain metastases are classified as charge changing mutations that can lead to conformational and electrostatic changes that can alter the protein function resulting in subsequent disruption in protein-substrate, protein-protein and protein-membrane interactions (Ardito et al., 2017; Zheng and Cui, 2016). These patterns of amino acid substitution might have been selected during tumour evolution as they could be advantageous for driving cancer cells' adaption and proliferation in a new microenvironment.

It has been established that these amino acid substitutions in cancer cells are not the result of random events; some cancer types have a selection preference for a particular amino acid substitution, with arginine to histidine having the highest frequency of substitution (Alexandrov et al.,2013; Anoosha, Sakthivel and Michael

Gromiha, 2016; Tsuber et al., 2017; Szpiec et al., 2017). Histidine residue is essential in proteins with pH sensitivity, it has been described that proteins harbouring arginine to histidine substitution are particularly sensitive to intracellular pH (pHi) level, thus altering the protein function (Nelakurti et al., 2021). It has also been described that most cancer cells are characterized by high pHi levels which allows them to increase cell proliferation and survival (Liu, White and Barber, 2020). When looking at proteins with high rates of arginine to histidine substitutions, White et al. (2017), found that an increase in pHi gave the mutant proteins oncogenic capacity. They observed that by increasing the pHi in breast cancer cells with arginine to histidine substitutions in transcription factor p53 led to a decrease in p53 transcriptional activity and diminished DNA damage response (White et al., 2017). This suggests that arginine to histidine mutations in cancer cells might be adaptative and they confer fitness advantage to the rise in pHi levels.

Even though arginine to histidine substitutions have the highest frequency of occurrence among cancer cells they are exclusive to only few cancer types including colorectal, stomach and uterine malignancies (Tsuber et al., 2017). Similarly other cancer types have a predilection for specific substitutions such glutamic acid for lysine in skin and urinary tract malignancies (Anoosha, Sakthivel and Michael Gromiha, 2016).

In the present analysis, substitutions of arginine to tryptophan, followed by alanine to valine were the most frequently identified changes (Figure 5.3). The previously mentioned studies were all carried out on primary tumour data which can be the reason for the contrast in the preferred amino acid substitution seen in the metastases data from this study. This difference in the frequency of substitutions for a particular amino acid might reflect specific functional and nutritional requirements specific to the tumour

microenvironment. As with arginine to histidine, arginine to tryptophan substitution alters the charge of the residue, thus it is expected to significantly affect the protein structure and function. The effect that arginine to tryptophan or virtually any other substitutions might have in the resulting protein and the role they may play in the metastatic process has not been studied yet. It can be argued that because these amino acid substitutions were predicted to be damaging (See Figure 5.3) by the prediction tools (SIFT and Ployphen-2) used in this study these residues are likely to be highly conserved throughout the evolution so any mutation affecting them will result in detrimental consequences. Tryptophan residues in proteins are particularly important as they modulate protein-protein and protein-peptide interactions as well as protein binding and recognition (Palego et al., 2016). As with histidine, tryptophan's charge is determined by variations in pHi levels; elevated pH cause tryptophan to be negatively charged (Cardenas et al., 2019) this could suggest that these types of amino acid substitutions might be selected during tumour evolution to drive and sustain tumour progression as cancer cells are characterised by having high pHi levels (Ward et al., 2020). It has been described that cancer cells may modulate cellular pH levels to promote tumour progression and invasion via regulation of ion channels, ion transporters and membrane trafficking (Andersen, Moreira, and Pedersen, 2014). Damaging amino acid changes in this study were located in protein domains or conserved amino acid residues involved in the regulation of the cellular metabolic pathways, which could be the key to understanding how cancer cells bypass growth and nutrient restriction and how they develop resistance to cancer therapies. Studying these changes can result in better understanding of the many cellular processes and mechanisms involved in metastatic development.

5.4 Conclusion

Here, the preferred amino acid substitutions of a cohort of 26 breast to brain metastases were determined. Several studies have explored the amino acid mutation signature of primary cancers, however, to date, studies on metastatic tumours have not been carried out. Understanding the implications of specific amino preferences by different cancer types could provide a better understanding of the mutational patterns of these malignancies. Particularly, functional work is needed to determine the consequences on the proteins' structure and function due to substitutions in the amino acids. Additionally, determining if these substitutions are characteristic of metastatic tumours and if they happen in combination with other mutational patterns might provide an insight on the selective pressures that take place during tumour evolution.

Chapter 6 - Mutation distribution within protein domains of metastasis-associated genes

6.1 Introduction

Normal cells accumulate somatic mutations that are essential for transformation into cancerous cells, and these alterations might also be decisive for disease progression and proliferation in a new environment (Tsuber et al., 2017). Most mutations have been described to be "passenger" alterations that do not contribute to cancer progression (Rentzsch and Orengo, 2013). However, some mutations that can change the protein function "drivers" will be selected during tumour evolution as they promote cancer growth and proliferation (Miller et al, 2015). Cancer studies have been focused on understanding the mutational landscape of cancer by analysing gene expression patterns across tumours (Lawrence et al., 2013; Alexandrov et al., 2013).

Nevertheless, genes can carry out multiple molecular roles in cancer progression, it can be the case that it is not the gene itself that is driving cancer development and differentiation but a specific gene function that might be conferring an advantage to cancer cells (Peterson et al., 2010). The categorization of cancer-associated variants concerning the relationship within the protein domains might be useful as it allows systemic assessment of genes' common biological functions (Cheng et al., 2014).

Protein domains are conserved fundamental homologous units of a protein that have been formed through evolution by recombination, duplication, or both. They determine protein structure, function, interactions, and evolutionary design (Rentzsch and Orengo, 2013). Due to the integrated recombination of domains during protein evolution, usually, large proteins consist of several domains (most small proteins have

one domain structure) that allow them to direct multiple cellular functions (Wang et. al., 2021). Domains can work in collaboration or autonomously, they can also have similar biological characteristics, but they can be found in proteins that have contrasting functions (Engin et al., 2013). Accurate representation of protein domains is vital in understanding the protein's biological activity and predicting and creating protein structures (Peng et.al., 2014). It has been indicated that mutations within a particular domain more than within an individual gene are likely to trigger similar structural and functional impacts (Yang et al., 2015). Linking mutations among several genes to a domain family or a set of domains can help identify other functional alterations and understand pathways that might be under selective pressure in cancer cells (Gauthier et al., 2015).

To assess, if the identified variants are likely to be driving metastasis, the conserved domains of 14 proteins of interest were analysed. It is hypothesised that for these changes to be metastasis drivers they should be located within or close to the protein regions that determine protein function (conserved domains), suggesting that alterations in these parts of the protein could lead to modification in the protein consequence that might be supporting the metastatic process (Miller et al., 2015).

For this analysis, the information in the NCBI CDD (Conserved Domain Database) (https://www.ncbi.nlm.nih.gov, n.d.) was used for finding highly conserved domains of the proteins of interest. The NCBI CDD contains manually curated data of highly conserved domain families of human proteins. To search for the conserved domains a query containing the protein in FASTA format was submitted to the CDD database. With the obtained results, a diagram for each protein was constructed, conserved domains were plotted, and mutated variants identified in the BBM samples from this study were mapped to establish their relative position within the protein structure.

Interpro (Blum et al., 2020), a database containing information of protein families and domains was used to determine the function of conserved protein domains. The mutation aligner data based (www.mutationaligner.org, n.d.) was used to determine domain mutation hot spots in cancer. Interpro (Interpro EMBL-EBI, 2019) database for the classification of protein families was used to obtain information about domain function. Additionally, the mutation distribution of the BBMs was compared with the mutational distribution within the same protein position in primary tumours that recurrently metastasise to the brain (breast, lung, and skin). Data on the primary tumours' mutation distribution was obtained from the CoSMIC database.

6.2 Results

6.2.1 Candidate metastasis-associated proteins and their conserved protein domains

Protein	Domains	Cancer types
GOLGA8K	*Smc Super family*	Glioblastoma, Liver Hepatocellular Carcinoma, Uterine Carcinoma
	*GM130_C	
	*GOLGAL5 Super family	
NEFH	*Filament Super family*	Melanoma, Head and Neck Carcinoma, Uterine Carcinoma, Lung Adenocarcinoma, Breast Invasive Carcinoma
	*PTZ00121 Super family	

TCHH	*S_100*	Melanoma, Bladder Cancer, Head and Neck Carcinoma, Breast Invasive Carcinoma, Colorectal Adenocarcinoma
	PTZ00121 Superfamily	
SCN10A	*Na_Transport_Cytoplasmic_Super family	
	Na_Transport_Associated	Melanoma, Lung Carcinoma, Breast Invasive Carcinoma, Brain Lower Grade Glioma, Bladder Cancer
	*Na_Channel_gate	
	Ion Transport	Melanoma, Lung Carcinoma, Uterine Carcinoma, Breast Invasive Carcinoma, Head and Neck Carcinoma
KIAA1244/ARFGEF3/BIG3	DCB	
	SEC7	Melanoma, Uterine Carcinoma, Liver Hepatocellular Carcinoma, Breast Invasive Carcinoma, Kidney Renal Clear Cell Carcinoma and Lung Carcinoma
	*PLN03076	
	*DUF_1981	
	*CMB_15	
RP1L1	DCX1_RP1L1	
	DCX2_RP1L1	

	PHA03307 Super family	
	2A1904 Super family	
COL6A3	*vWA_Collagen_alpha3 VI-like	
	*VWFA Super family	
	VWA	Melanoma, Lung Carcinoma, Liver Hepatocellular Carcinoma, Head and Neck Carcinoma, Breast Invasive Carcinoma
	Collagen	Melanoma, Lung Carcinoma, Head and Neck Carcinoma, Kidney Cancer, Breast Invasive Carcinoma
	FN3	Melanoma, Lung Carcinoma, Head and Neck Carcinoma, Colorectal Carcinoma, Uterine Cancer, Breast Invasive Carcinoma
	Kunitz_BPTI	Melanoma, Lung Carcinoma, Breast Invasive Carcinoma, Colorectal Carcinoma, Liver Hepatocellular Carcinoma
BRCA2	*BRCA2* repeat	
	*BRCA2 helical Superfamily	
	*OB1 folds	
	*OB2 folds	
	*OB3 folds	
HSPG2	*SEA*	Bladder Carcinoma, Cervical

		Carcinoma, Lung Carcinoma
	*Lam B	
	*IgL_Perlecan like	
	*LamG	
	*I-set	
	*IG like	
	*Ig Super family	
	*IG	Lung Carcinoma, Head and Neck Carcinoma, Melanoma, Liver Hepatocellular Carcinoma, Uterine Carcinoma
	*EGF	
	*EGF_Lam	
	*EGF_CA	Melanoma, Liver Hepatocellular Carcinoma, Breast Invasive Carcinoma, Lung Carcinoma, Head and Neck Carcinoma
	*LDLa	
	*Laminin EGF	Melanoma, Liver Hepatocellular Carcinoma, Lung Carcinoma, Uterine Carcinoma, Breast Invasive Carcinoma, Colorectal Carcinoma
FAT1	*Cadherin_repeat	
	Cadherin Super family	Melanoma, Lung Carcinoma, Breast Invasive Carcinoma, Uterine Carcinoma, Head and Neck Carcinoma
	*LamG	

		EGF_CA	Melanoma, Liver Hepatocellular Carcinoma, Breast Invasive Carcinoma, Lung Carcinoma, Head and Neck Carcinoma
		*EGF	
RYR3		Ins 145_P3-rec Super family	
		MIR	Melanoma, Lung Carcinoma, Uterine Carcinoma, Breast Invasive Carcinoma, Head and Neck Carcinoma
		*RYDR_ITPR	Lung Carcinoma, Breast Invasive Carcinoma, Melanoma, Liver Hepatocellular Carcinoma, Uterine Carcinoma
		*SPRY1_RyR	
		*SPRY2_RyR	
		*SPRY3_RyR	
		*RyR	Melanoma, Lung Carcinoma, Breast Invasive Carcinoma, Uterine Carcinoma, Head and Neck Carcinoma
		*RIH_associated	
		*RR_TM4-6	
		*Ion_transport	Melanoma, Lung Carcinoma, Uterine Carcinoma, Breast Invasive Carcinoma, Head and Neck Carcinoma

	*EF_hand 7	
	*RyR Super family	
RYR1	Ins 145_P3-rec Super family	
	MIR	Melanoma, Lung Carcinoma, Uterine Carcinoma, Breast Invasive Carcinoma, Head and Neck Carcinoma
	RYDR_ITPR	Lung Carcinoma, Breast Invasive Carcinoma, Melanoma, Liver Hepatocellular Carcinoma, Uterine Carcinoma
	*SPRY1_RyR	
	*SPRY2_RyR	
	*SPRY3_RyR	
	RyR	Melanoma, Lung Carcinoma, Breast Invasive Carcinoma, Uterine Carcinoma, Head and Neck Carcinoma
	*RIH_assoc	
	*RR_TM4-6	
	*Ion_transport	Melanoma, Lung Carcinoma, Uterine Carcinoma, Breast Invasive Carcinoma, Head and Neck Carcinoma
HYDIN	*ASH	
	*Hydin_ADK	
	*TPH Super family	
	*Ptz00121 Super family	
KMT2D	PHD1_KMT2D	
	*PHD2_KMT2D	
	PHD3_KMT2D	
	PHD5_KMT2D	

	PDH1_KMT2C_Like	
	PDH5_KMT2C_Like	
	PHD_Super family	
	*PHA03247 Super family	
	HMG_box	Brain Lower Grade Glioma, Melanoma, Liver Hepatocellular Carcinoma, Breast Invasive Carcinoma, Colorectal Carcinoma
	*FYRN	
	*FYRC	
	SET_KMT2D	Head and Neck Carcinoma, Melanoma, Breast Invasive Carcinoma, Lung Carcinoma, Liver Hepatocellular Carcinoma

Table 6.1: Proteins with their associated domains. Domains present in more than one of the 14 are shown in red. Domains reported in the mutation aligner database (www.mutationaligner.org) as a mutation hot spot in cancer are underlined. Domains that were directly affected by mutations or were located near the mutation location in the BBMs from the present study are marked with *.

From the table above (Table 6.1), 5 proteins were selected for further analysis, considering the protein size and where the identified mutations sit within the protein. The smallest proteins GOLGA8K, NEFH, TCHH, SCN10A and BIG3 were considered suitable candidates as metastasis promoter genes as the mutations were located within the protein functional units or very close to it. It is hypothesised that mutations that could be involved in metastasis tumour development should be located within the

functional region of the protein, resulting in damage or alteration of the protein function leading to increased cell survival, migration, and proliferation, and that ultimately the new tumour microenvironment would favour selection of these cells with a more aggressive phenotype. Mutations in the larger proteins were scattered across the proteins' length. It is expected that mutations in larger proteins is due to genomic instability in tumours due to damage or loss of DNA repair mechanisms and that these mutations will be found dispersed throughout the protein. It is also thought that these modifications are less likely to change the protein function and therefore would not be selected as metastatic drivers. Nevertheless, these larger proteins are worth being studied in future research as they could provide a more complete understanding of the complexities of metastatic tumour development.

Schematic representation of proteins with their conserved domains and identified mutations are shown in figures 6.1 to 6.5 for GOLGA8K, NEFH, TCHH, SCN10A and BIG3. For the rest of the proteins please see Appendix F.

For the present analysis, mutations predicted to be damaging that are upstream or are close to the conserved domain are considered equally relevant than those sitting within the domain structure as damaging mutations thin the domain are also in highly conserved amino acid residues outside the protein domains and they cause damage to the domain and the protein (Miller et al., 2015).

6.2.2 Schematic representation of mutations positions within metastases-associated proteins.
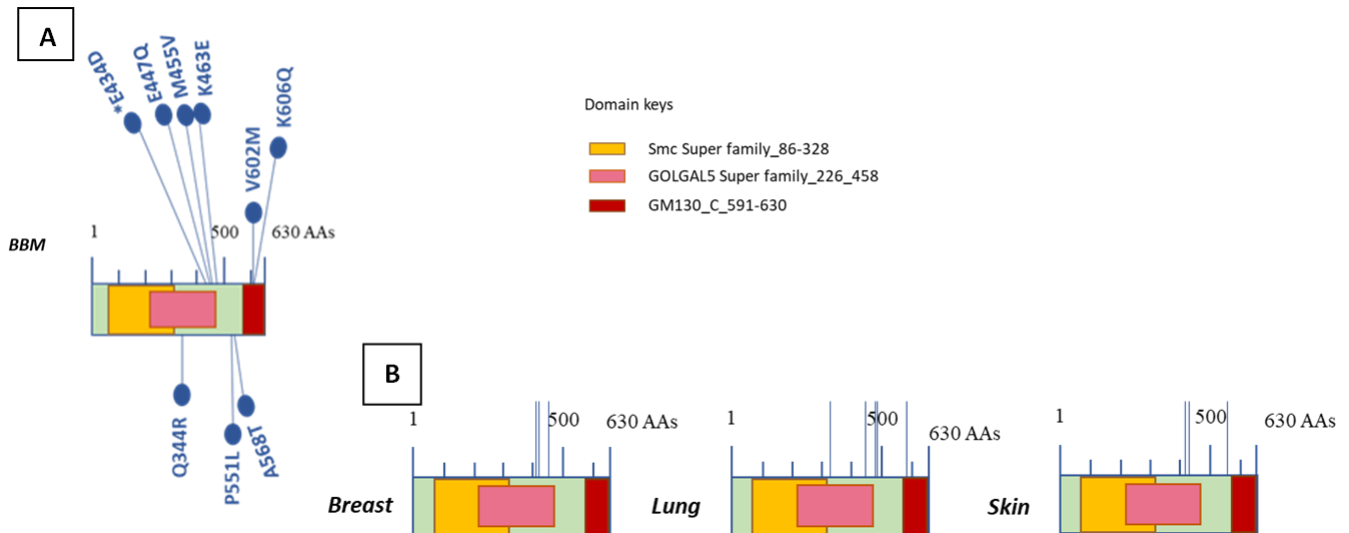
Figure 6.1: Schematic representation of GOLGA8K mutations within the protein domains. The GOLGA8K protein is formed by 630 amino acids and 3 conserved domains. A) Amino acid changes found in BBMs mapped in their relative position within the protein. All amino acid substitutions fall either within a specific conserved domain or in the area in between domains. B) GOLGA8K mutation distribution that falls within the same area of the ones found in BBMs in this study. Data on primary tumours was obtained from CoSMIC database. Amino acid substitutions M455V and E4567Q found in the BBMs were present in breast and lung primary tumours, respectively. All Missense mutations are presented in blue in both A and B figures. (*) Amino acid change predicted to be benign by either SIFT or Polyphen-2.
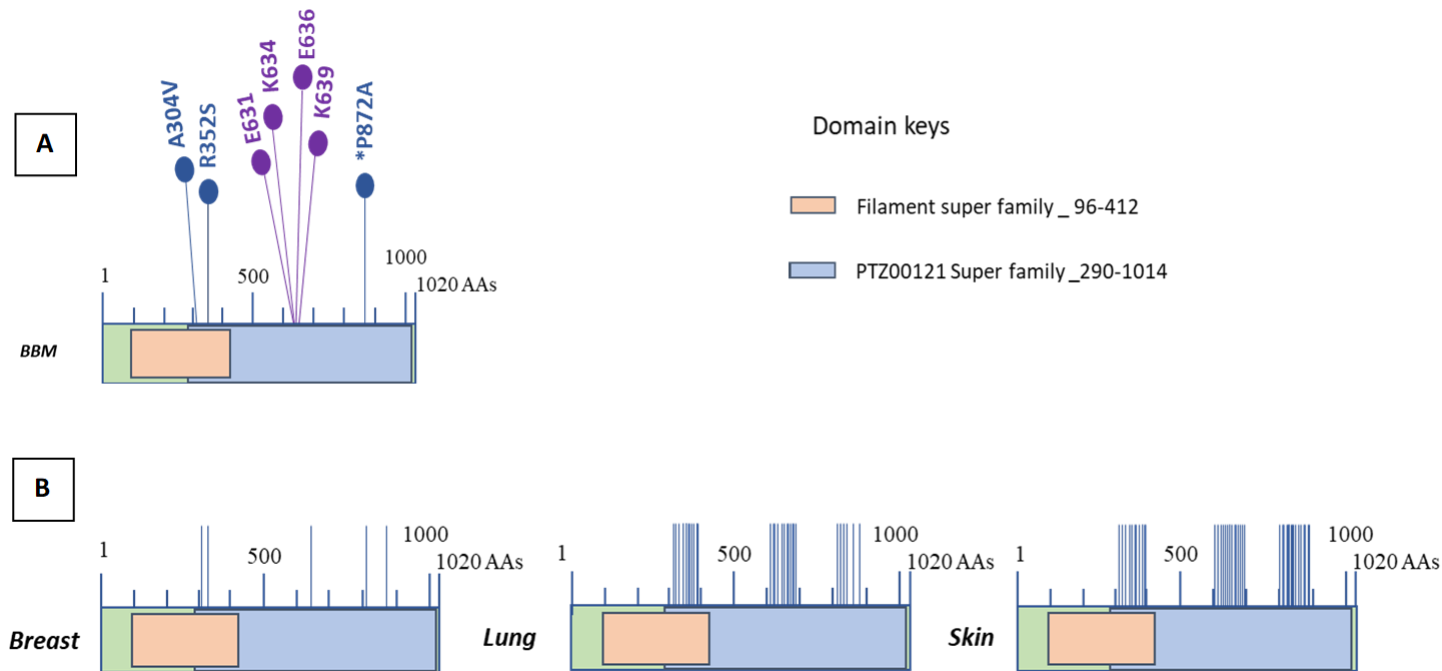
Figure 6.2: Schematic representation of NEFH mutations distribution in within the protein domains. NEFH protein is formed by 1020 amino acid residues and two conserved domains. A) Amino acid changes are shown in blue and purple corresponding to missense and frameshift mutations, respectively. * Amino acid changes predicted to be benign by either SIFT or Polyphen-2. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.
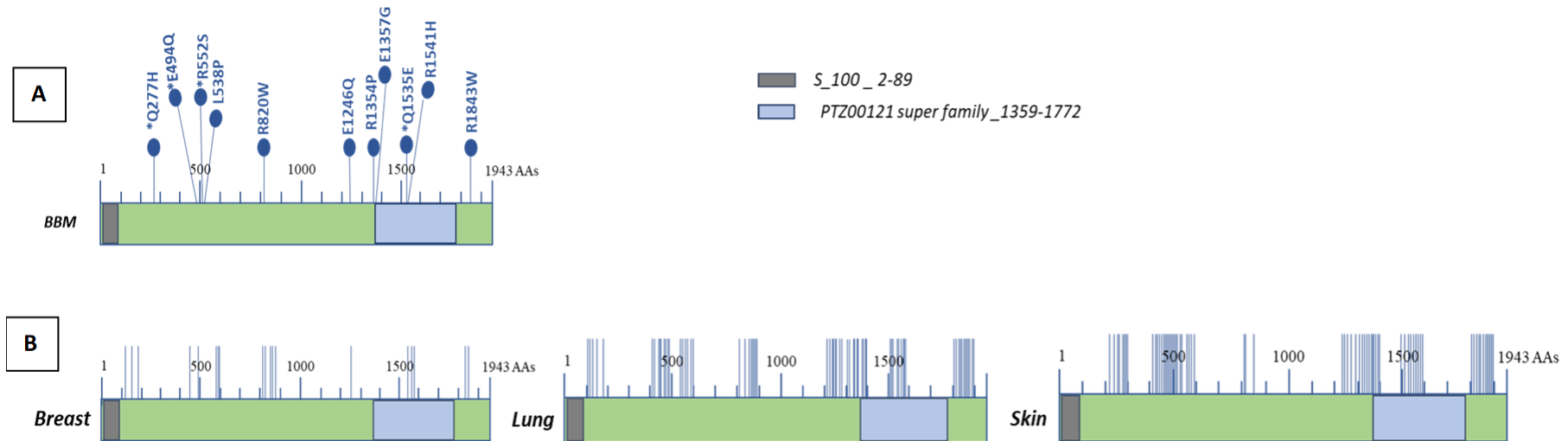
Figure 6.3: Schematic representation of TCHH mutations distribution in within the protein domains. TCHH protein is formed by 1943 amino acid residues and two conserved domains. A) Amino acid changes are shown in blue corresponding to missense mutations. * Amino acid changes predicted to be benign by either SIFT or Polyphen-2. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.
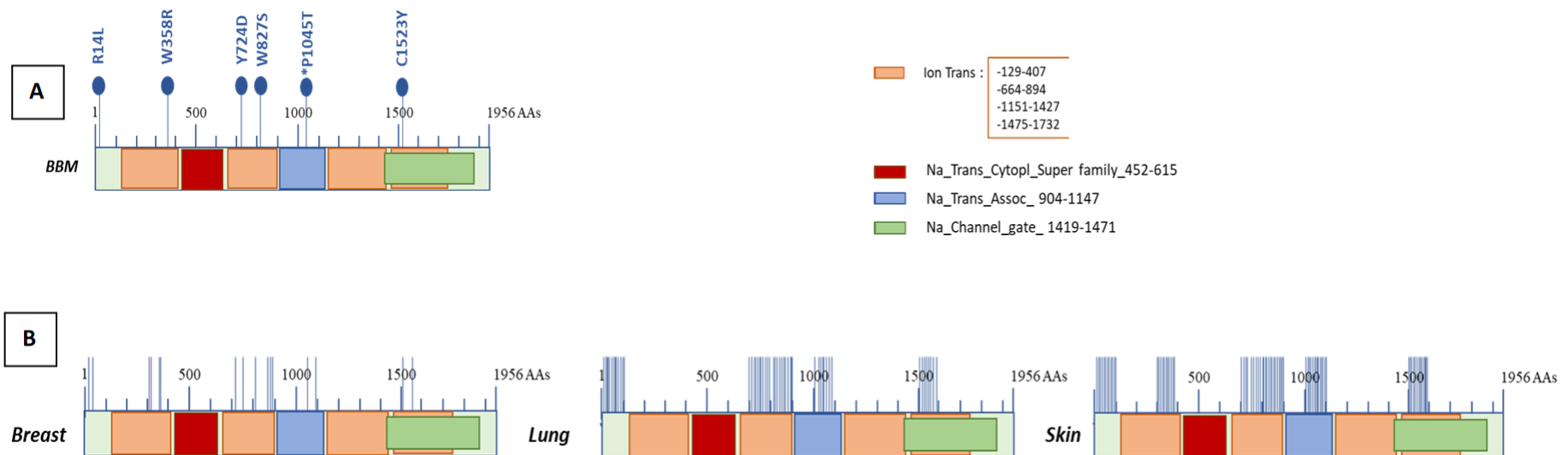
Figure 6.4: Schematic representation of SCN10A mutations distribution in within the protein domains. SCN10A protein is formed by 1956 amino acid residues and four conserved domains. A) Six amino acid changes were novel or have a MAF ≤0.01%, five of which were predicted to damage the resulting protein product. Three damaging mutations and one benign are within the Ion Trans domain. Mutation C1523Y is within the overlapping region of the Ion Trans domain and the Na_Channel_gate domain. Mutations are shown in blue corresponding to missense alteration. * Amino acid change predicted to be benign by either SIFT or Polyphen-2. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.
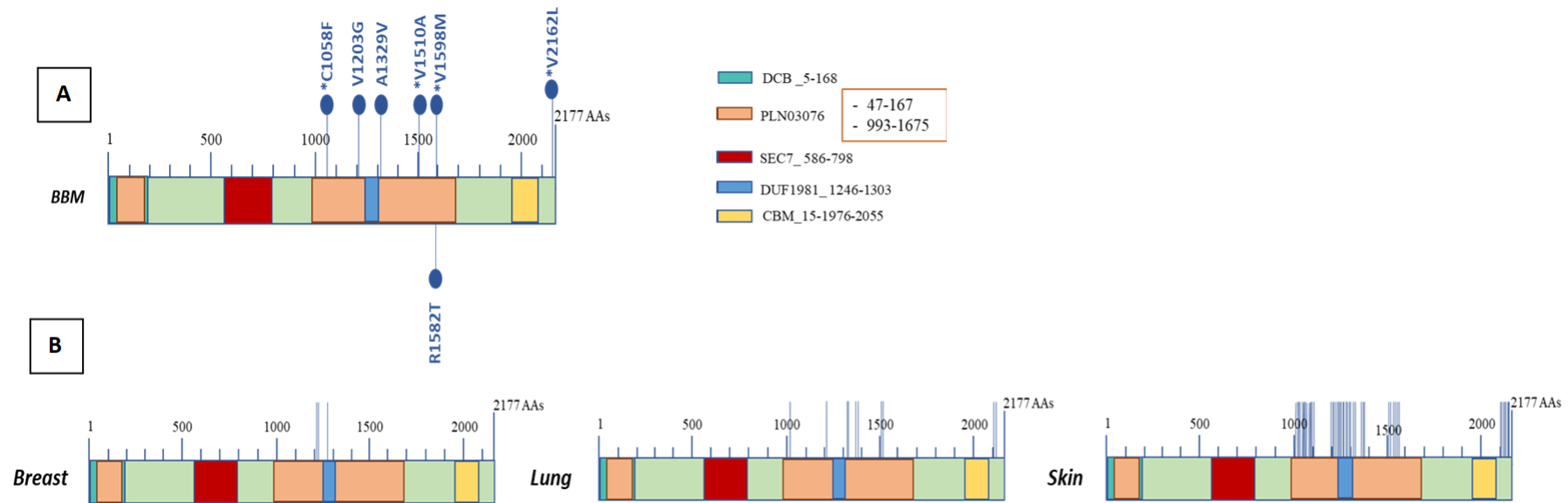
Figure 6.5: Schematic representation of BIG3 mutations distribution in within the protein domains. Schematic representation of BIG3 protein formed by 2177 amino acid residues. A) Seven amino acid changes are represented in the protein structure. Three missense changes, V1203G, A1329V and R1582T were predicted to be damaging for the protein. * Amino acid change predicted to be benign by either SIFT or Polyphen-2. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.

6.3 Discussion

The mutations presented in this study were considered rare or novel variants due to their very low MAF or because they were not found reported on the ExAC database, respectively. Most of the damaging mutations were localised either within a conserved domain structure or in its surrounding area. Hence, it can be suggested that they are most likely to be somatic variants that may be involved in metastatic progression. Knowledge of the implications that these mutated variants might have in cancer progression is of utmost importance. Accordingly, because domains are the functional unit of the proteins, and mutations targeting these regions are likely to be damaging to the protein function; analysing the metastases-associated variants identified in this study in the context of where they sit within the protein can help to interpret the consequences of these mutations in the protein function. Additionally, correlation of this information with previous domain studies helps in the identification of novel genes that might have been overlooked because they are not frequently mutated in cancer, but they harbour mutations in protein domains that have been identified as frequently mutated in well-known cancer genes which suggest that some mutations are necessary for tumour adaptation and survival, therefore they are selected during tumour evolution.

Studies in primary tumours have identified frequently mutated protein domains. In a study carried out by Hashemi and colleagues the preferentially mutated domains and protein coding regions in primary tumours were identified. They analysed the data from the TCGA of 29 cancer types and found that primary tumours have preference for protein domains that regulate DNA repair and cell metabolism. They concluded that proteins involved in DNA repair pathways are the most mutated proteins in cancer and that the p53 domain encoded by the p53 gene was the number one domain affected,

regardless of cancer type. This was followed by the tm_4 whose role in cancer development needs further investigation (Hashemi et al., 2017). Another study, looking at preferred domain mutations within proteins in various cancer types, found that p53 and PI3K domains were significantly mutated (Nehrt et al., 2012). Moreover, protein domain hot spots for cancer mutations were identified in a study of 21 cancer types by Yang and colleagues. They identified 59 tumour driver genes that were not previously identified as such by comparing their domain mutation data with previous studies of domain mutational landscape in melanoma, breast, lung, and colon cancer (Yang et al., 2015).

A total of nine missense mutations in the GOLGA8K protein (eight of which were predicted to be damaging for the resulted protein) were represented in their relative position within the protein (Figure 6.1). These amino acid changes were found to be within or in close proximity to the three protein domains (Smc, GOLGAL5 and GM130_C). Specifically, four out of the nine mutations were localised within the GOLGAL5 Super family domain, and two mutations were within the GM130_C domain.

GOLGA8K is a small protein coding gene formed by 630 amino acid residues and 3 conserved domains, it is involved in the organization of the Golgi apparatus (National Center for Biotechnology Information, 2017). The Golgi apparatus controls intracellular trafficking, and damage to proteins and domains involved in the regulation of Golgi organisation and function result in fragmentation of the Golgi apparatus which led to the dysregulation of several cellular processes including cell glycosylation, kinase activity, Ras GTPases, and aberrant myosin motor proteins activity that result in cancer development and subsequent progression to metastatic disease (Petrosyan, 2015). To evade the host immune system, cancer cells take control of Golgi-mediated glycosylation. Aberrant Golgi glycosylation was reported to be involved in tumour

angiogenesis and increased metastatic potential of gastric, breast, prostate tumours, and liver cancer cell lines (Liu et al., 2021; Dresang et al., 2022).

The GM130_C domain is the C-terminal that is bound by the GRASP65 PDZ domain. GRASP65 protein mediates stacking of the Golgi apparatus during mitosis, which is an essential step for the correct formation of the Golgi apparatus (Manjunath, Ramanujam and Galande, 2017). Incorrect stacking will result in a defective Golgi machinery that in turn will result in disruption of protein trafficking, processing, and secretion (Tan, Yang and Wang, 2010). Evidence suggests that GM130 is implicated in the regulation of cell division and migration through the regulation of microtubule and spindle organization, cell glycosylation and by controlling microtubule and mitotic spindle organization (Zhang et al., 2011). In a breast cancer study, it was observed that loss of GM130 in breast cancer cells resulted in increased cell motility and higher invasion rates leading to tumour progression (Baschieri et al., 2015).

Mutations of proteins containing Smc domain have been linked to cancer as they result in failed DNA damage response and aberrant DNA-protein interactions that result in genomic instability (Wu and Yu, 2012). Smc is part of the chromosome segregation ATPase family that controls cell cycle, cell, and chromosome division. Abnormal chromosome segregation during mitosis results in chromosomal instability which is essential for cancer development (Gisselsson, 2008; Matityahu and Onn 2017).

The role of the GOLGAL5 Super family domain have not been yet determined, however it can be hypothesised as with any other Golgi related proteins that damaging mutations could result in alterations of protein trafficking and possible aberrant DNA damage response that can lead to cancer development and progression.

NEFH is a neurofilament heavy chain protein that assembles with other neurofilament proteins (NEFM and NEFL) to form the cytoskeleton of mature neurons, where it has a major regulatory role in the intracellular processes of axons and dendrites, and thus can be employed as a biomarker to assess neuronal damage (Koudonas et al., 2022). Mutations of neurofilament proteins have been implicated in diseases affecting motor neurons and the peripheral nervous system as well as with the development and progression of several cancer types (Alholle et al., 2013; Huang et al., 2013, Koudonas et al., 2022).

NEFH protein is formed by two conserved domains spanning from amino acid position 96 to 1014. The filament Super family domain is an intermediate filament protein that regulates protein-protein interactions, and assembly of the supramolecule (Kornreich et al., 2015). Cancer progression and metastasis development have been linked to changes in the expression of intermediate filament proteins (Joosse et al., 2012; Chen et al., 2014). In an experimental setting, breast cancer cells with epigenetically silenced NEFH showed a metastasis supporting phenotype (Calmon et al., 2015). In another study carried out in metastatic prostate cancer arising from different primary sites, it was concluded that genes coding for intermediate filament or associated proteins were largely down regulated in all metastatic prostate cancer tumours regardless of the tissue of origin of the primary tumours (Samaržija, 2021). These findings suggest that the damaging amino acid changes at position A304 and R352S in the NEFH protein can be considered as promoters of tumorigenesis and metastasis as they are positioned in the area where the two protein domains, Filament Super family and PTZ00121, overlap each other. These domains are essential for the proteins' regulatory role in cell orientation, cell-cell interaction, and cell structural

stability, protein-protein interactions among other mechanisms (Strouhalova et al., 2020; Sharma et al.,2019).

TCHH is a structural protein that crosslinks with keratin to regulate and maintain the structural integrity of hair follicles (Mukamel et al., 2021). Some studies have associated TCHH with cancer development and progression (Dong et al., 2021; Yang et al., 2019) and resistance to cancer therapy (Yang et al., 2018). In a study of colon adenocarcinoma tumours, TCHH mutations were associated with tumour lymphatic invasion and metastasis as it was found that downregulation of this gene promoted proliferation of infiltrating lymphocytes, however the exact mechanisms involved in this have not been described yet (Shi et al., 2021).

Eleven missense mutations affecting the TCHH protein were down stream of the s-100 domain which is part of the family of EF-hand calcium binding proteins that regulated several cellular processes including cell homeostasis, differentiation, and proliferation (Allgöwer et al., 2020). Of these eleven amino acid changes, 7 (L538P, R820W, E1246Q, R1354P, E1357G, R1541H and R1843W) were predicted damaging by SIFT or Polyphen-2 which confirm these mutations sit within highly conserved domain or amino acid residue. Mutation R1541H directly affects the PTZ00121 domain; L1354P and E1357G are upstream of this domain. Thus, it is believed that these mutations will likely lead to structural and functional alterations of the TCHH protein resulting in aberrant crucial cellular processes that will favour cancer progression.

NEFH and TCHH share the Super family domain PTZ00121 (refer to Figures 6.2 and 6.3) PTZ family domains mediate protein-protein interactions and several forms of cell signalling function (Basu, Poliakov and Rogozin, 2009). Physiological cell signalling

pathways dictate cell survival, motility, death, proliferation, and differentiation, these pathways are disrupted in cancer which allows for cancer cells to adapt and survive in the tissue microenvironment (Sever and Brugge, 2015). In this study three frameshift mutations on the NEFH protein, and two and three damaging missense mutations for NEFH and TCHH respectively were found within this domain. Thus, it can be said that these mutations might be implicated in the development and progression of breast to brain metastasis, as damaging mutations in conserved regions are likely to be selected during tumour evolution. Additionally, in the case of NEFH, the frameshift mutations are predicted to change protein conformation by shifting the reading frame, thus producing an aberrant protein. Furthermore, these three mutations were affecting a highly conserved protein region which makes them good candidates for metastasis associated mutations.

The SCN10A protein is involved in cell sodium transport by directing sodium channel formation (National Centre for Biotechnology Information, 2017). This protein has been reported to be involved in cardiac conduction and mutations have been associated with aberrant cardiac rhythm (Abou Ziki et al., 2017). The exact role this protein plays in cancer remains unknown, but mutations in the SNC10A gene have been reported to be implicated in the progression of peripheral neuropathy in patients with colorectal cancer following chemotherapy (Kong et al., 2019).

The SCN10A protein is formed by three sodium transport and one ion channel domain. The damaging amino acid changes W358R, Y724D and W827S occur in the Ion transport domain and C1523Y is in the overlapping region of Ion transport and sodium (Na) channel gated domain. R14L is upstream from all the protein domains. The Ion-transport, Na-transport-associated domains have been reported as a hot spot for mutations across several cancer types including Melanoma, Lung Carcinoma, Uterine

Carcinoma, Breast Invasive Carcinoma, Head and Neck Carcinoma, Bladder Cancer, and Brain Lower Grade Gliomas (data obtained from mutational aligner database) which suggest these mutations may confer a selective functional advantage to tumour cells.

All the damaging mutations in the BIG3 protein were found within the region PLN0376 and DUF1981 domains and upstream from the CBM domain (Figure 6.5). As of this date, the activity of these domains is still undetermined (Interpro EMBL-EBI, 2019). However, several members of the DUF domains from different protein types have been reported in the mutational aligner database as mutation hot spots for various types of cancer including Liver, Lung, Breast, Bladder, and Colorectal Carcinomas among others (www.mutationaligner.org,n.d.). It can only be hypothesised because these domains are structural components of the BIG3 protein and they are conserved through evolution, any damaging mutation targeting them might also be selected during tumour evolution and these mutations might have some degree of association with cancer progression. The function of BIG3 will be further discuss in chapter 8.

One major requirement for cancer progression is the control of the metabolic machinery by tumour cells (Pascual, Domínguez and Benitah, 2018). Cancer cell metabolism differs from that of normal cells; they benefit from a dysregulated oxidative metabolism, which is characterised by aerobic glycolysis and decreased oxidative phosphorylation (the Warburg effect). This has been extensively studied across a variety of tumour types as a major factor influencing tumour cell survival (Liberti and Locasale, 2016; Vander Heiden, Cantley and Thompson, 2009; Pascale et al., 2020). By regulating cell metabolism, cancer cells are capable of sustaining growth and proliferation by adapting to any metabolic restriction that may arise in the tumour microenvironment, which results in of the more aggressive phenotypes being selected

for survival during tumour evolution (Damaghi et al., 2020; Damaghi and Gillies 2017). Mutations in protein domains such as ion channel transport, sodium transport, or Golgi apparatus that regulate protein-protein interactions and membrane trafficking pathways hence controlling cell metabolism, are crucial for maintaining cancer development and proliferation (Staub and Rotin, 1997; Caterino et al., 2017).

Establishing the mutational alterations that control the development of brain metastases can be challenging for many reasons, one of which is the lack of access to match them to their primary tumour samples. The CoSMIC database has data from thousands of somatic mutation studies in primary cancer, consequently, this data could be used to compare mutational patterns between metastatic and primary tumours.

When it comes to the origin of the metastatic disease regardless of the primary tumour of origin there is still much that remains unknown. Two models (the linear progression and parallel models) have been used to described metastatic development (Naxerova and Jain, 2015). The linear model states that metastatic precursor cells leave the primary tumour late and that metastatic cells are the result of clonal selection resulting in a genetically related metastatic and primary tumour (Hutchinson, 2015). On the other hand, the parallel model considers metastases as an early event resulting in the independent evolution of the primary and the metastatic tumours (Krøigård et al., 2017).

Following analysis of CoSMIC data of primary tumours that frequently metastasise to the brain (lung, breast, and skin) it was found that BBMs harbour mutations that correlate with the mutational pattern of these primary tumours in the same protein area (Figures 6.1 to 6.5 part B for schematic representation of the mutations within the

proteins). Some of the damaging amino acid changes found in GOLGA8K, NEFH, TCHH and BIG3 proteins in the BBM samples were reported in the skin, lung and breast tumour data obtained from CoSMIC. GOLG8K mutations E447Q and M455V were reported in lung and breast tumours, respectively. NEFH R352S was found in melanoma. Mutations E1246Q and R1354P on TCHH protein, and P1045T on SCN10A were reported for lung cancer. Other changes occur at the same amino acid position, yet the amino acids gained differed from the ones in the BBM samples. Nonetheless, the occurrence of these amino acid changes in these primary tumours might first help to paint the mutational trajectory that could be determining brain metastatic tumour development, and second it suggests that these mutations might be necessary for cancer development and progression. However, it is not possible to establish if the mutations identified in the BBMs of this study are the result of an early or late event; sequencing of the matching primary tumour would be required to more accurately characterise the evolutionary pattern of metastatic development.

6.4 Conclusion

Domain analysis can be a powerful tool when analysing cancer mutation data as it can help identify novel genes that have not been previously linked to cancer, Additionally, domain mutation data can be used to prioritize mutations for functional studies, that could be used for the development of target therapy or the use of already existing therapy targeting alterations with the same biological effect.

Most of the domains identified in this study have not been implicated in cancer susceptibility, and the mechanism by which this group of proteins might be involved in the metastatic process are not completely understood. However, cancer studies have

indicated that mapping mutations to specific protein domains can help to identify rare potential cancer drivers that could have been overlooked by gene focus analysis, as domain studies can provide a wider understanding of the functional consequence of mutations (Nehrt et al., 2015; Yang et al, 2015; Hashemi et al., 2017). The results of the present study may be a step forward towards an improved understanding of the complexities of the metastatic process by identifying domain mutational patterns that could be specific to breast to brain metastatic tumours. These results could further be used to create a database of the mutational landscape of brain metastatic tumours that could aid in the development of target therapy to treat these malignancies by supplying a more specific target region.

It would be ideal to further explore the distribution of mutations in protein domains with a larger group of samples of matched primary and metastatic tumours from different tissue sites to be able to create a comprehensive profile of the mutational pattern of metastatic tumours.

Future work will be required to look in-depth into the domain architecture by multiple domain sequence alignment and by characterising the functional implications of these mutations by performing mutational pathway alignment followed by laboratory confirmation of these amino acid alterations.

Chapter 7 - Whole Exome Sequencing of BBMs identified BIG3 as a novel metastasis-associated gene.

7.1 Introduction

In the present analysis, seven mutations matching the filtering criteria previously described in chapter 4 were identified in a member of the ERFGEF (guanine-nucleotide exchange factors for ADP-ribosylation factor GTPases) protein family member ERFGEF3 also known as BIG3 or KIAA124. The seven mutations, three of which were predicted to be damaging for the protein product were successfully validated by Sanger sequencing in the laboratory (see Figures 1 and 2) in the BBM samples. DNA extracted from matched patient's blood when available were also Sanger sequenced to determine if the mutations were somatic or germline.

Data of primary breast tumours from the CoSMIC database showed that BIG3 mutations were present in less than five per cent of the total breast cancer samples analysed in their database, which indicates that this gene is not frequently mutated in primary breast tumours. Studies have reported that BIG3 is highly upregulated in breast tumours and that this is associated with adverse prognosis (Kim et al., 2009; Yoshimura et al., 2017; Yoshimaru, Nakamura and Katagiri, 2021. On the other hand, investigations into the role of BIG3 in metastatic brain tumours from any primary source have not been carried out so far. The present analysis was carried out with the goal of identifying the role played by BIG3 in the development of breast to brain metastases. Furthermore, beyond to its assumed role in tumour formation, a role in the regulation of neuronal neurotransmitter activity has been suggested (Venkatesh et al., 2019; Venkataramani et al., 2019; Zheng et al. 2019). This is going to be investigated in this chapter in relation to tumour metastasis.

7.2 Results

Following Whole Exome Sequencing and Insilico analysis of 26 BBM samples data, BIG3 was selected as one of 14 metastasis associated candidate genes. Seven mutations 4 with a MAF ≤ 0.1% and 4 were not reported (NR) on the ExAC database. These 7 mutations were selected for pathogenic prediction using SIFT and Polyphen-2 Insilico. 3 mutations with no data on the ExAC (no reported mutations) were predicted pathogenic (Table 7.1) the 4 remaining mutations were predicted to be benign for the protein consequence.

BBMs	Mutation type	Gene section	Nucleotide change	Amino acid change	MAF	Pathogenic prediction
12438	Missense	19	c.3173G>T	C1058F	NR	Benign
2613	Missense	21	c.3608T>G	V1203G	NR	Probably damaging
712	Missense	24	c.3986C>T	A1329V	NR	Probably damaging
12138	Missense	28	c.4529T>C	V1510A	0.00002471	Probably damaging
12146	Missense	29	c.4745G>C	R1582T	NR	Benign
119	Missense	30	c.4792G>A	V1598M	0.0003994	Benign
756	Missense	34	c.6484G>T	V2162L	0.000008258	Benign

Table 7.1: Mutations in BIG3 found in 7 BBM samples. Three pathogenic mutations predicted to be damaging were found not reported on ExAC which suggest these are novel and are likely to be cancer associated. One not reported variant, and three variants with a MAF of less than 0.1% were predicted to be benign. Sanger sequencing validation is shown in figures 7.1 and 7.2.

7.2.1 Sanger sequencing electropherograms of missense mutations identified in BIG3 gene
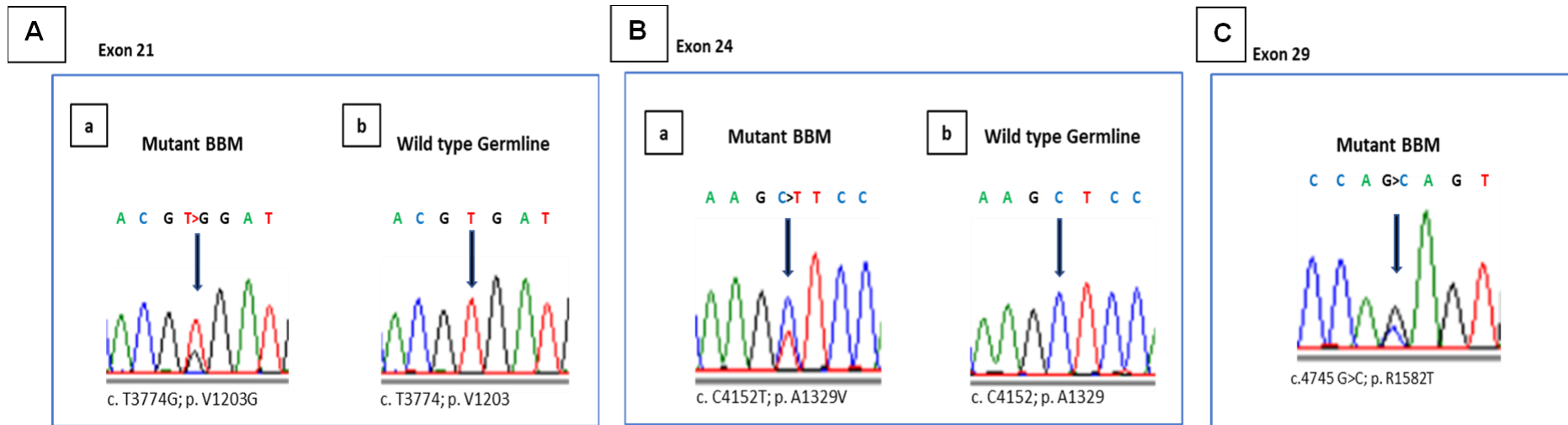
Figure 7.1: Sanger sequencing electropherograms of 3 BIG3 mutations predicted to be damaging mutations. Mutations found in BIG3 three exons in three different BMM samples. A) Image for exon 21 shows: a) nucleotide change from T to a G on the BBM sample, b) Wild type germline are the results from patient's blood. B) Mutation identified in exon 24: a) Image for the BBM nucleotide change of C to T. b) chromatograph from Patient's blood DNA C) Nucleotide change from G to C found in exon 29.
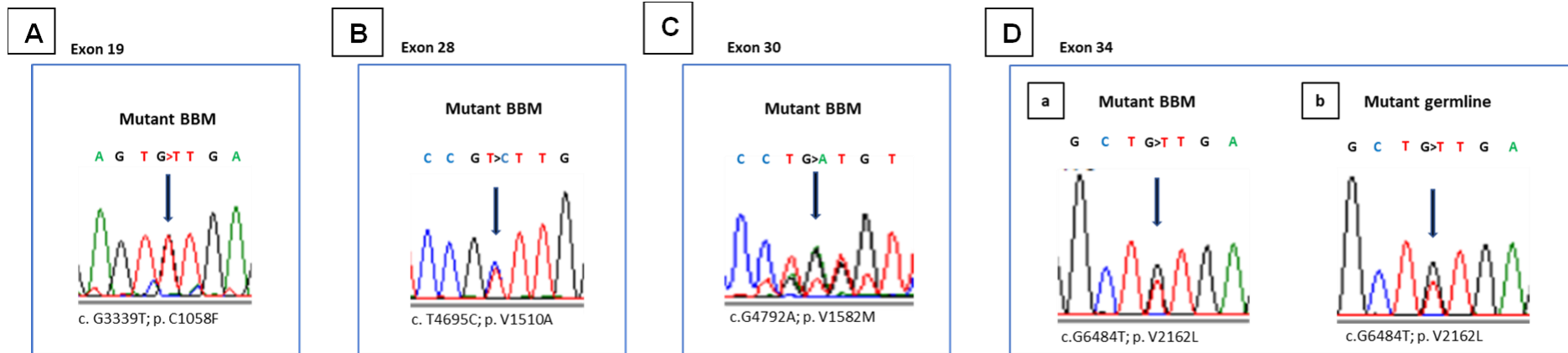
Figure 7.2 Sanger sequencing electropherograms of BIG3 mutations predicted to be benign. Four different exons in four different BMM samples. A, B and C show mutations found at exons 19, 28 and 30, respectively. D) Image for exon 34. a) mutant BBM. b) the mutation was identified in matched blood DNA.

7.2.2 BIG3 silencing using CRISPR-cas9 lentiviral vector gene editing

To investigate the functions of BIG3 the CRISPR-cas9 gene editing system was used to removed BIG3 from the MCF7 breast cancer cell line. Before CRISPR-cas9 gene editing was carried out, the expression of BIG3 was determined in five breast cancer cell lines by RT-PCR.
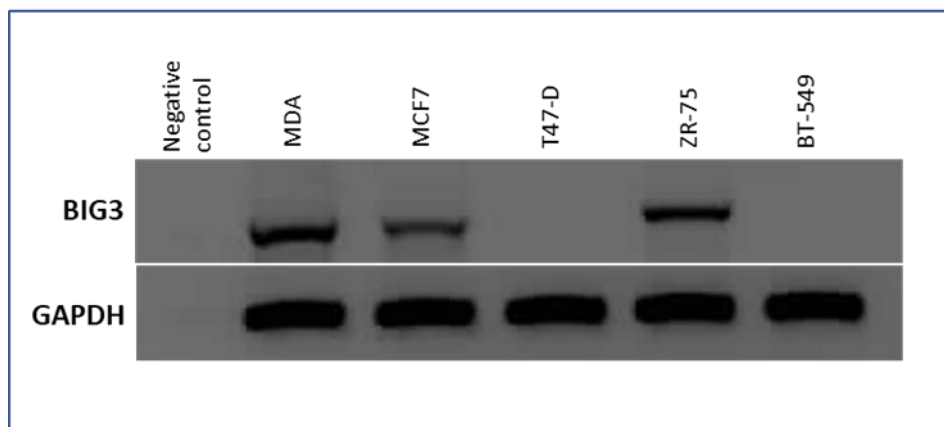
Figure 7.3: BIG3 was found expressed in MDA, MCF7 and ZR-75 cell lines.
Expression data was obtained by RT-PCR. Equal loading was determined by GAPDH.

Plasmid for CRISPR-cas9 gene editing was bought from abm goods Canada. Bacterial transformation followed by miniprep and were carried out to purify the plasmid DNA. Plasmid was then digested to ensure plasmid identity.
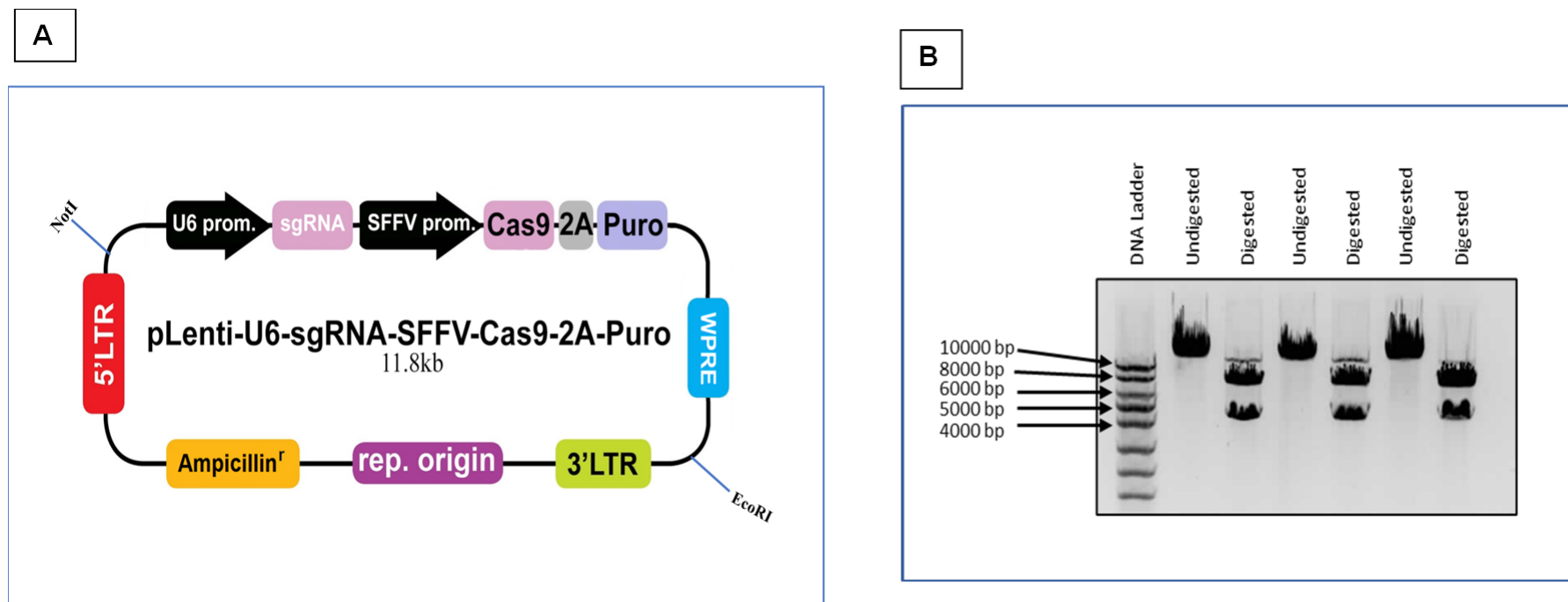
Figure 7.4: Schematic representation of BIG3 lentiviral vector. A) Schematic representation of BIG3 lentiviral vector system's composition with the approximate restriction digestion sites, marked with the cutting enzymes used to digest the plasmid DNA. B) Validation of plasmid purification before carrying out cell line transfection was performed by double digesting plasmid with endonucleases EcoRI and NotI. The undigested fragment produced a band above the ten thousand base mark in the PCR ladder

which matches the expected size of more than eleven thousand. Two fragments, one of around seven thousand base pairs and a smaller of around four thousand can be observed following digestion, confirmed the plasmid's identity. The vector target sequences, and the protocol used for endonuclease restriction digestion was described in the Materials and Methods Chapter 2.

Confirmation of CRISPR-cas9 gene editing was carried out using RT-PCR and analysis. Following transfection, MCF7 cells transfected with BIG3 and MCF7 transfected with a vector containing a scrambled control gRNA sequence were grown under specific tissue culture conditions (see chapter 2 Material and Methods) and clones were isolated. Each BIG3 transfected clone was given a number from 1 to 47 to identify them. Following puromycin antibiotic selection, seven clones survived (9,15,19,28,29 and 47). RNA was extracted from these seven populations of cells, and the control clone transfected with the scrambled vector, to carry out cDNA synthesis for RT-PCR confirmation. RT-PCR results confirmed successful gene editing shown in Figure 7.5.
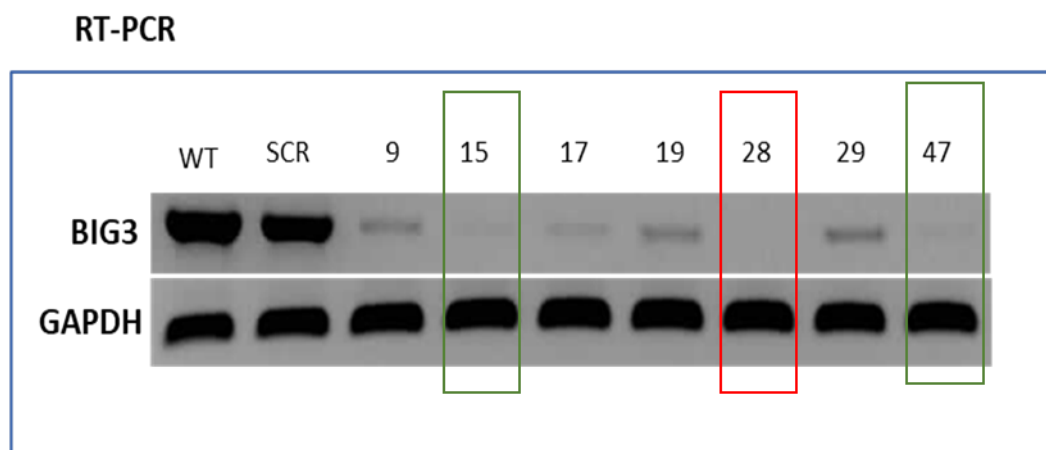
Figure 7.5: Confirmation of BIG3 expression in following CRISPR- knockout at the RNA and protein level. RT-PCR shows no expression of BIG3 in cell population 28 (red box) and a marked decrease in BIG3 expression can be seen in clone 15 (green box) and 47. clones 9, 17, 19 and 29 also show decreased BIG3 mRNA expression when compared with non-transfected cells containing

wild type BIG3 (WT) and scrambled (SCR). GAPDH expression was used to determine equal loading. RT-PCR shows no expression of BIG3 in cell population 28 (red box) and a marked decrease in BIG3 expression can be seen in clone 15 (green box) and 47. clones 9, 17, 19 and 29 also show decreased BIG3 mRNA expression when compared with non-transfected cells containing wild type BIG3 (WT) and scrambled (SCR). GAPDH expression was used to determine equal loading.
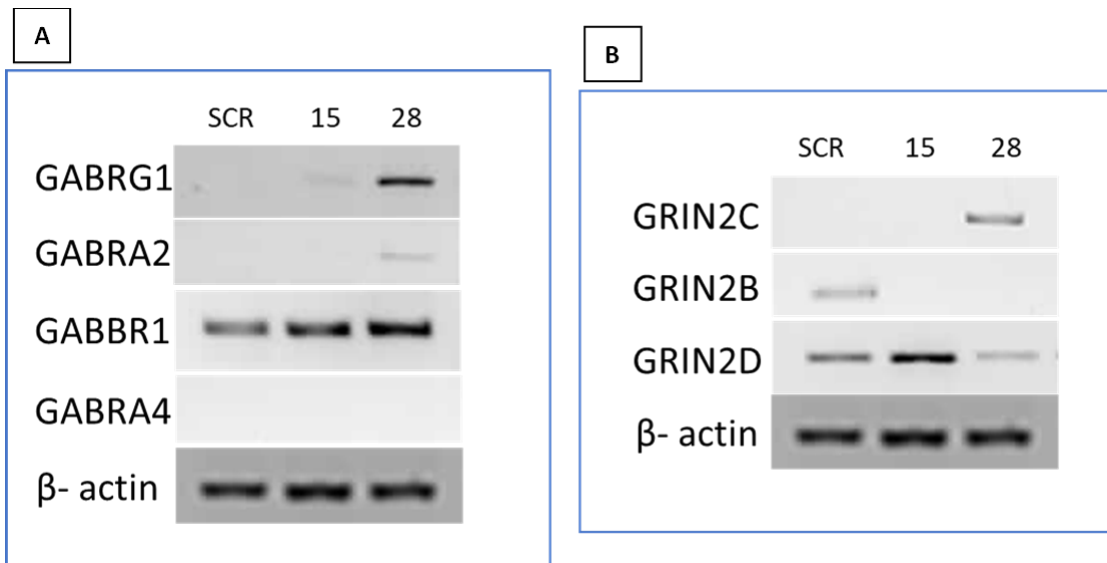
Figure 7.6: RT-PCR expression data of Neurotransmitter receptor subunits in MCF7 breast cancer cell line. A) Expression data of gamma aminobutyric acid receptors (GABARs) subunits in MCF7 expressing BIG3 (SCR) and BIG3 knockout (28) and knockdown (15) MCF7. Upregulation of GABRG1 is observed in 28. GABRA2 expression is shown only in 28. GABBR1 subunit expression is present in the SCR, 15 and 28. None of the MCF7 cells, wild type or knock-out show expression of GABRA4. B) Upregulation of Glutamate receptors (NMDARs) subunit GRIN2C is shown in the BIG3 knockout clone, 28. Downregulation of GRIN2B is observed in MCF7 knockdown and knockout clones. GRIN2D downregulation is observed in the BIG3 Knockout clone, 28. Expression was determined by RT-PCR. Equal loading was determined by β- actin.

7.3 Discussion

BIG3 is a signal transduction protein coding gene encoded by 34 exons mapped to chromosome six forward strand (NCBI) that belongs to the ADP-ribosylation factor (ARF) family of guanine-nucleotide-binding (G) proteins in charge of the regulation of cellular organelles' development, and membrane dynamics (Pipaliya et al., 2019). BIG3 is expressed in healthy individuals in pancreatic islets where it regulates systemic glucose homeostasis, and in the brain where it regulates neurotransmitter secretion from hippocampal neurons (Liu et al., 2013; Li et al., 2014; Chigira et al., 2019).

BIG3 is found to be overexpressed in breast cancer cells where it regulates oestrogen transcriptional activity (Yoshimaru et al., 2015; Kim et al., 2009). BIG3 overexpression in breast tumours is associated with poor prognosis because it blocks PHB2 by binding to it in the cell cytoplasm, thus preventing its translocation into the cell nucleus where PHB2 binds and represses the transcriptional activity of oestrogen receptor alpha (ERα) (Yoshimaru et al., 2015; Chigira et al., 2019). The BIG3-PHB2 interaction results in dephosphorylation of PHB2 which results in the loss of PHB2 tumour suppressor activity leading to the activation of oestradiol (E2)/ERα pathway which promotes tumour growth, proliferations, and resistance to breast cancer treatment with antioestrogen agents such as tamoxifen and xanthohumol (Yoshimaru, Nakamura and Katagiri, 2021; Yoshimaru et al., 2014) thus, targeting BIG3-PHB2 interaction has been suggested as a plausible therapeutic approach to treat breast cancer patients (Yoshimaru et al., 2014). Figure 7.7 shows a representation of BIG3/PHB2 interactions.
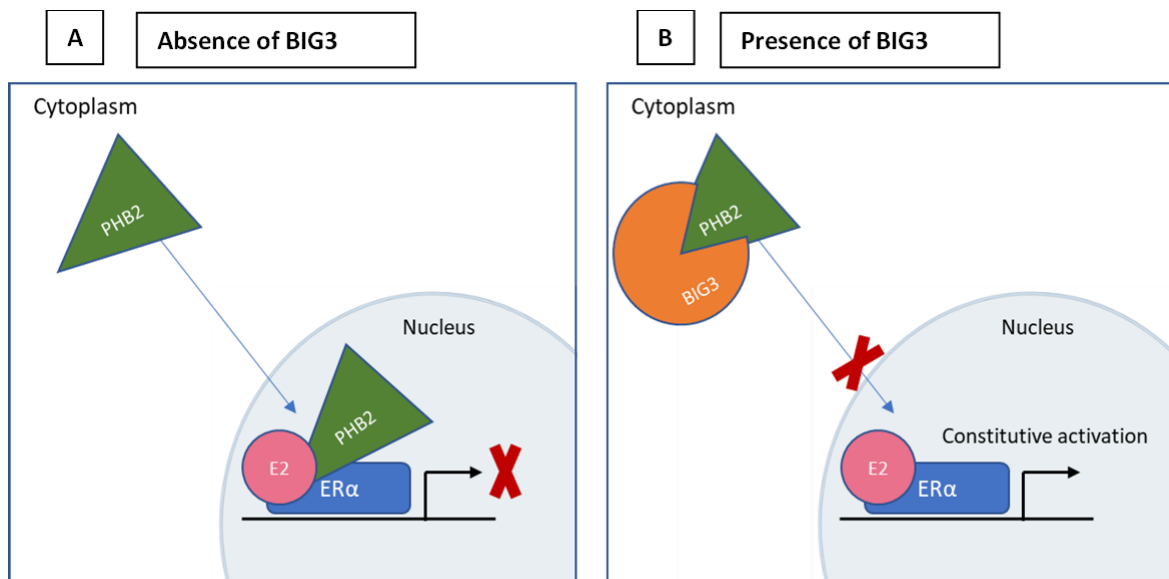
Figure 7.7: Schematic representation of BIG3 regulation of ERα signalling pathway. When BIG3 is not present PHB2 translocates into the cell nucleus and binds to ERα repressing its transcriptional activity. B) BIG3 binds to PHB2 in the cytoplasm inhibiting PHB2 translocation into the cellular nucleus which results in activation of ERα signalling pathway. Image adapted from Kim et al., 2009 and Yoshimaru et al., 2017.

Based on these formerly discussed findings it can be inferred that BIG3 has oncogenic effect in breast cancer and, that BIG3 mutations could positively impact patients' recovery and survival. For the present study, the question to answer is how are BIG3 mutations implicated in the development and proliferation of breast to brain metastases?

The study of the mechanism driving the metastatic process in an ever-evolving subject. It is known that for primary tumour cells' colonization of the brain to take place, the cells must cross the blood brain barrier (BBB). Research have suggested thatspecific genes and membrane transport molecules give tumour cells the capability of infiltrating the BBB (Arvanitis, Ferraro and Jain, 2019; Bailleux, Eberst and Bachelot,

2020). An in vitro study carried out by Bos and colleagues using HUVECs (human umbilical vein endothelial cells) identified several genes that appear to be implicated in creating a more permeable BBB such COX2 (Cyclooxygenase-2) and EGFR (Epidermal growth factor receptor). Additionally, they found that ST6GALNAC5 (2-6-sialyltransferase) expression allows breast cancer cells to attach to brain epithelial cells, thus facilitating migration of breast cancer cells into the brain (Bos et. al., 2009). Additionally, well-known cancer associated genes TP53, PIK3CA, KMT2C, and RB1 were found to be highly mutated in patients with breast to brain metastasis in a panel of 167 matched primary breast tumours and brain metastases (Morgan, Giannoudis and Palmieri, 2021). However, the association of a group of genes with the metastatic process does not ensure that they provide tumour cells with the necessary evolutionary advantage to be selected by the brain microenvironment. For brain colonization to take place, cancer cells must evolve and acquire new genomic alterations that will support their adaptation and growth in the new microenvironment (Neman et al., 2014; Li et al., 2022). On this note, it has been described that metastatic brain tumours evolve independently from the primary tumour, acquiring genomic changes that allow for the colonization of the brain (Bastianos et al., 2015; Iwamoto et al., 2019). In the case of BIG3 it can then be hypothesised that breast cancer cells may develop BIG3 mutations after leaving the primary tumour site where BIG3 has been found to be highly expressed, and that these alterations give an evolutionary selective advantage to breast to brain metastatic cells allowing them to create a supportive tumour microenvironment.

Characterizing the genomic alterations that are required for tumour colonisation of the brain has been the subject of intense research. Proposed mechanisms include the upregulation of neurotransmitter receptors by tumour cells, which allows them to

control the neuronal signalling network, resulting in the activation of cellular pathways that favour the development of tumours inside the brain; This has been described in both primary brain tumours, and breast to brain metastatic tumours (Neman et al., 2014; Brosnan and Sanders, 2018; Venkatesh et al., 2019; Venkataramani et al., 2019; Zheng et al., 2019; Li et al., 2022). It was stated that metastatic breast cancer cells have the capability of modifying the brain microenvironment to their benefit by displaying a neuron-like signalling phenotype (Neman et al., 2014). These metastatic cells overexpressed gamma aminobutyric acid receptors (GABAR) that allow the tumour cells to use GABA neurotransmitter as their source of energy (Neman et al., 2014; Brosnan and Sanders, 2018). The expression of this neural-like phenotype by breast cancer cells may be giving them the proliferative advantage needed for brain colonization.

Zheng and colleagues showed that breast to brain metastatic cells' brain colonisation is determined by the activation of NMDARs signalling. Cancer cells achieve this by replacing the astrocytes at the neuronal synaptic site resulting in the formation of pseudo synapses between the tumour cells and the pre- and post-synaptic neurons (Figure 7.8). This sabotage to the synaptic signalling system results in increased release of glutamate at the synaptic cleft activating the NMDAR/glutamate axis which promotes the growth of brain metastatic tumours (Zheng et al., 2019). Upregulation of NMDARs has been associated with increased tumour growth and proliferation in several cancer types (Du et al., 2020). In agreement with these findings, blocking NMDARs have shown therapeutic benefit reducing invitro tumour progression and proliferation (Du et al., 2020; Zheng et al., 2019; Li and Hanahan, 2013). NMDAR activation in physiological conditions have been demonstrated to be neuroprotective

(Hahn, Wang, Margeta, 2015) thus it can be suggested that by expressing NMDA receptors tumour cells disguise themselves as part of the brain's resident cells.
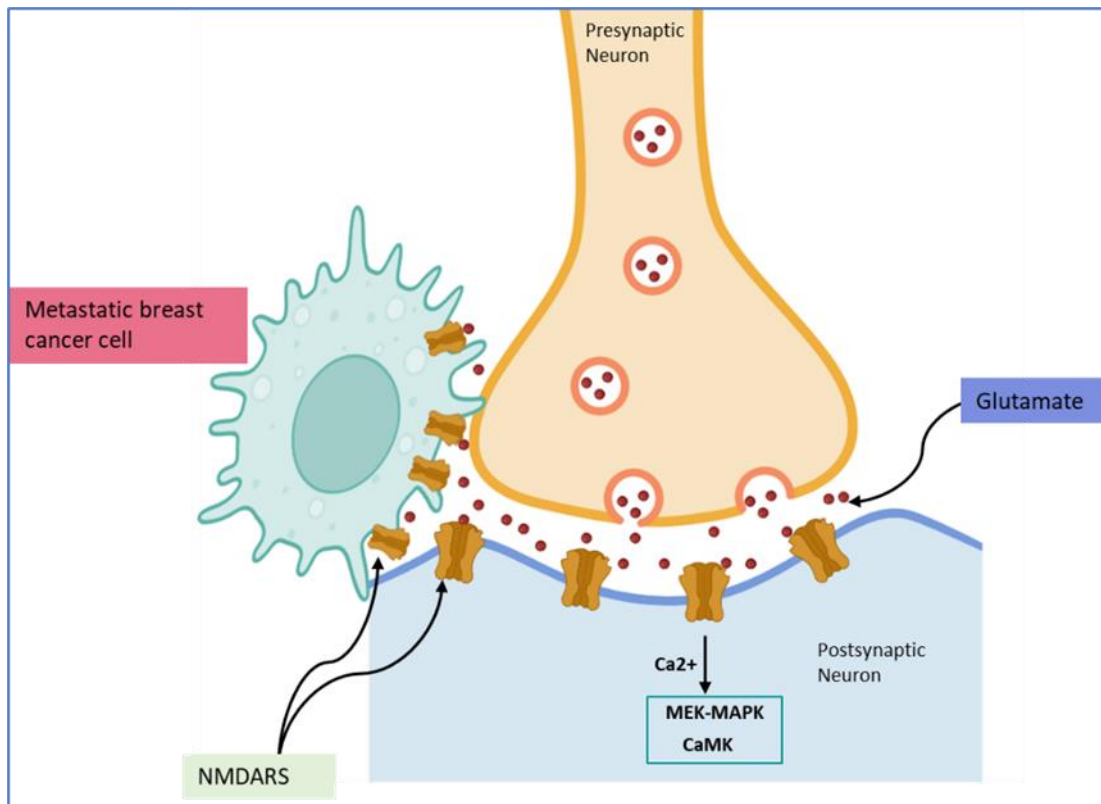
Figure 7.8: Schematic representation of NMDAR dependent brain colonisation. Metastatic breast cancer cells mimic astrocyte like expression of NMDA receptors and supplant the astrocytes at the synaptic cleft. Activation of NMDARs results in influx of calcium (Ca^{2+}) followed by the downstream activation of the two major signalling pathways, MEK-MAPK (mitogen activated protein kinase) and CaMK (Ca^{2+}/calmodulin-dependent protein kinase) resulting in promotion of tumour invasion and progression. Image created with Birender.com, adapted from Zheng et al., 2019, and Li and Hanahan 2015.

To elucidate the role of BIG3 in breast to brain metastases CRISPR-cas9 editing of MCF7 breast cancer cell lines was carried out following the protocol described in the materials and methods section. One successful knockout and two knockdowns were confirmed by RT-PCR and western blot analysis (see figure 7.5). Expression of NMDA (N-methyl-D-aspartate receptors) (GRIN2B, GRIN2C, GRIN2D) and GABA (GABRG1, GABRA2, GABABR1, GABRA4) neurotransmitter receptor subunits in a BIG3 knockout/knockdown MCF-7 breast cancer cell line was determined using RT-PCR. The results confirmed that neurotransmitter subunits were differentially expressed in BIG3 knockout and knockdown MCF7 cells compared to those expressing endogenous BIG3 (showed in Figure 7.6).

Expression of Glutamate receptor subunit GRIN2C, and GABA receptor subunit GABRA2 was observed only in BIG3 knockout MCF7. GABRG1 expression was seen in both the knockout and the knockdown. GABBR1 and GRIN2D subunits were found expressed in the wild type as well as the knockout and knockdown. GRIN2B upregulation was observed in MCF7 expressing wild type BIG3.

Brain tumours regulate synaptic activity in the brain via universal mechanisms (neurotransmitter receptor upregulation). There are several paths through which tumour achieve control of the neuronal signalling network. These which include tumour-neuron- interactions that result in enhance electrical and synaptic neuronal activity or via tumour upregulation of neurotransmitter/neurotransmitter subunits (Venkatesh et al., 2019; Venkataramani et al. 2019; Zheng et al., 2019). BIG3 upregulation of neurotransmitter subunits suggest that BIG3 mutations may be one the ways through which breast to brain metastases achieve colonisation of the brain.

Zheng and colleagues when analysing the expression of six glutamate receptors in breast cancer patients, they found glutamate neurotransmitter subunit GRIN2B to be upregulated in more than forty 40 percent of the sample analysed. Additionally, following knockout of GRIN2B in a MDA231 breast cancer cell line, they concluded that to create a supportive brain microenvironment, interaction between neurons and metastatic brain tumour cells must take place resulting in GRIN2B/ NMDAR activation (Zheng et al., 2019). It was found that NMDAR activation generates a Ca^{2+} influx into the post synaptic neuron, which results in the activation of two major signalling pathways MEK-MAPK and CaMK, both of which are linked with cancer development (Monroe, Basheer and Gibert, 2021). Activation of MEK-MAPK and CaMK pathways regulates cell survival, proliferation, migration, and differentiation, (Han et al., 2021; Martinelli et al., 2017; Brzozowski and Skelding, 2019). Activation of these pathways has been linked to poor prognosis in tumours as it is related to the development of therapeutic resistance (Martinelli et al., 2017; Monroe, Basheer and Gibert, 2021). Drugs targeting the MEK-MAPK pathway and its regulatory molecules in combination with other therapeutic agents have shown promising pre-clinical outcomes in the treatment of several tumours (Han et al., 2021; Monroe, Basheer and Gibert, 2021; Simpkins et al., 2018; Yamasaki et al., 2022).

Characterising the role played by BIG3 in brain colonisation by breast tumour cells can offer an opportunity to understand the complexities behind brain metastatic tumour development as well as the opportunity for the exploration of possible therapeutic targets. The role of BIG3 in the regulation of neuronal signalling was described by Liu and colleagues. They determined that BIG3 is found in the lysosomes in hippocampal neurons and that the absence of BIG3 results in an increased in the number of lysosomes in those neurons which results in enhance GABA-dependent synapses due

to the increased expression of GABA receptors (Liu et al., 2016). Lysosomes are regulators of several crucial cellular mechanisms, including initiation of the immune response, regulation of cell metabolism, control of gene expression, cellular migration, and apoptosis (Bonam, Wang and Muller, 2019). Lysosomes are implicated in cancer development, progression, migration, and proliferation (Tang et al., 2020). They provide tumour cells the necessary mechanisms to control cellular signalling, catabolic and metabolic pathways that allow cancer cells to redirect energy sources, evade immune attack, and to consolidate the creation of a tumour supporting microenvironment by regulating extra and intracellular pH levels and the remodelling of the extracellular matrix (Michiels et al., 2020; Hongu and Oskarsson 2020; Saftig and Puertollano, 2021).

This research, including the present study, revealed that there is bidirectional communication between the tumour cells and the brain cellular network., showing that cancer cells that invade the brain use the neurotransmitter's signalling pathways to modify the brain's microenvironment to support tumour adaptation, resulting in unrestrained cell growth and proliferation within the brain (Jiang et al., 2019; Kuol et al., 2018). One of such mechanism may be BIG3 loss by mutation.

7.5 Conclusion

Metastatic tumours continue to be the primary cause of deaths due to cancer. Although advances in genomic characterisation have made possible the identification of genes implicated in metastasis tumour development, is still necessary to understand the factors and mechanisms behind metastatic tumour colonisation and survival. The present analysis has suggested a mechanism (BIG3 mutation) that facilitates crosstalk between the metastatic tumour cells and the brain neuronal network. The study of

brain tumour metastases has revealed that tumours co-opt the synaptic signalling to create a brain microenvironment that will support their metabolic demands to sustain tumour growth and proliferation.

The present analysis showed that neurotransmitter subunits are differentially expressed in BIG3 knockout and knockdown MCF7 breast cancer cell line which suggest that BIG3 mutations may play a significant role in metastatic brain tumour survival and proliferation in the brain microenvironment. Further experiments are required to establish a pattern of neurotransmitter subunits regulated by BIG3. It was previously described that BIG3 in found in the lysosomes of hippocampal neurons, therefore determining the expression of lysosomal marker (Lamp1) in BIG3 expressing cells and knockout/knockdown cell lines, would be the next step to further assess the role of BIG3 in the regulation of neurotransmitter receptor in breast to brain metastases. Additionally animal xenograft studies with a BIG3 wild type and knockdown genotype could be beneficial in determining if BIG3 promotes brain tumour development.

CHAPTER 8- General conclusion

Cancer is a heterogenous disease in which an interplay of several factors influences the development and the eventual progression to metastatic tumours. Tumour angiogenesis is essential for the progression of the primary tumour. Worldwide cancer statistics show that breast cancer is the most diagnosed cancer in women. However due to the advances in early diagnosis and the efficacy of therapy, primary cancer of the breast does not tend to be lethal. Death in these patients is usually the result of metastatic brain tumours that develop years or decades after the primary tumour diagnosis. Understanding the genomic changes and the mechanisms behind this progression may offer an opportunity to predict the development of, or to stop their growth.

8.1 Role of the transcription factor ATF2 in the regulation of angiogenic gene expression in VEGF-Stimulated endothelial cells

8.1.1 ATF2 regulatory function of Notch signalling ligands

ATF2 is thought to be a novel molecular target for antiangiogenic therapy. This molecule is a phosphorylation-dependent regulator of cellular processes such as the DNA damage response elicited by ionizing radiation or ultraviolet light (Bohoumik et al.,2005, Lopez-Bergami, Lau and Ronai, 2010). The link between ATF2 and VEGF-induced angiogenesis has been supported by in vitro and in vivo studies (Fearnley et al., 2014; Bus et al., 2018; Fearnley et al., 2019; Wang et al., 2021). VEGF regulates the angiogenic and inflammatory process via activation of specific extracellular signals leading to phosphorylation of ATF2 that modulates the expression of angiogenesis-associated genes (Fearnley et al., 2014; Gozdecka et al., 2014). Also, as part of the pro-angiogenic transcription program in nasopharyngeal carcinomas, ATF2

transcriptional activity is stimulated by the Epstein–Barr nuclear antigen 1 (EBNA1) (Lau and Ronai, 2012).

Here it was shown that ATF2 transcriptional activity may play a role in angiogenesis via negative regulation of DLL4 Notch signalling pathway ligand in endothelial cells. The Notch pathway is a central signalling mechanism that controls cell fate and differentiation thus playing a pivotal role during vascular development, and the angiogenic process (Hasan et.al., 2017). The Notch ligand-receptor mechanism is activated via cell-cell interactions when the receptor's extracellular domain interacts with the corresponding ligand found on the adjacent cells (Dufrain et.al., 2008). Complete knockout of Notch ligands Notch1 and Notch4 was found to cause lethal vascular defects (Krebs et.al., 2000).

In this study, it was observed that ATF2 negatively regulate the expression of DLL4. The molecular mechanisms by which ATF2 modulates DLL4 expression are still uncharacterised. It can be hypothesised that the negative regulation of DLL4 observed in this study could be the result of direct ATF2 binding to DLL4 gene regulatory regions or indirect activation by ATF2 of a DLL4 repressor that would result in decreased DLL4 expression. The negative regulatory effect of ATF2 in gene expression has been demonstrated by several studies. A Chromatin immune precipitation assay (ChIP) determined that ATF2 has transcriptional repression activity on interferon-$\beta1$ (IFN$\beta1$) by direct binding to the gene promoter region (Lau et al., 2015) and another study carried out in Macrobrachium Nipponese found that antimicrobial peptide genes are negatively regulated by ATF2 via regulating the expression of TNF (Zhang et al., 2020).

Another possible explanation for ATF2 negative regulation of DLL4 is that ATF2 may be regulating the expression of specific miRNAs that would target the DLL4's three prime untranslated region resulting in expression of the gene being reduced. In a study conducted to assess the role of microRNA-30b (miR-30B) in the regulation of cellular morphogenesis via transforming growth factor beta 2 (TGFβ2) regulation, it was found that cells depleted of ATF2 presented inhibited miR-30b- dependent TGFβ2 expression. This suggested that ATF2 activation (phosphorylation) was required for miR-30b up-regulation of TGFβ2. Conversely, the effects of miR-30b on ATF2 were carried out by JDP2, an ATF2 repressor molecule, this resulted in decreased JDP2 mRNA expression (Howe, Kazda and Addison, 2017).

When assessing for DLL1, SNAI2 and PPARG expression data showed no statistical significance. This could potentially be due to the activity of another ATF family member such as ATF7 since it also targets the ATF2 DNA binding site (Liu et al., 2016). Preliminary analysis showed no significant alteration in ATF7 RNA expression in HUVECs with functionally suppressed ATF2 (Prof. A Armesilla, personal communication). Hence it can be argued that ATF7 might be compensating for suppressed ATF2 activity in the regulation of DLL1 in the experiments carried out using si-RNA mediated suppression where no significant changes in gene expression were observed.

8.1.2 ATF2 role in endothelial cell tube formation

Furthermore, it was investigated the potential role of ATF2 during tubule formation using an organotypic co-culture assay. The mechanism underlying tubule formation and development involves the signalling of several molecules such as growth factors and their receptors (Iruela-Arispe and Beite, 2013). RTKs are major regulators of cell

proliferation, differentiation, and apoptosis (Fearnley et.al. 2020). RTKS regulatory functions are ubiquitous or cell-specific, and their signal transduction regulation is required to modulate cell behaviour and responses to extracellular signals (Lemmon and Schlessinger, 2010).

For the functional analysis, adenovirus-infected GFP-HUVECs lacking functional AFT2 were seeded on top of a monolayer of human derived fibroblast (HDFα), stimulated with bFGF, and quantification of tubule formation determined by anti-CD31 antibody staining. Pairwise comparison between the groups showed no statistical significance in tubule formation between the Ad-GFP and Ad-AFT2 controls. However, there was a significant increase in tubule formation when comparing HUVECs Ad_GFP control and Ad_GFP stimulated with bFGF. Tubule formation in Ad_ATF2 HUVECs stimulated with bFGF showed a slight increase in tubule formation when compared with Ad_GFP control.

8.2 Whole-Exome Sequencing analysis of 26 Breast to Brain Metastases

Analysis of Whole Exome Sequencing data from 26 breast to brain metastases showed that known primary breast cancer drivers were found mutated across the 26 BBM samples from this study, including PIK3CA, Tp53, CDH1 and KMT2C. Most of the changes were concentrated outside the coding exon. However, variants, clustered into 15 genes, were within the exons or at splice site regions (KMT2C, ZFHX3, LRPB1, PIK3CA, ALK, GATA3, PTPRT, MAP3K1, GRIN2A, NF1, MED12, ESR1, PTEN, ARID1A, ERBB2, KMT2C, ZFHX3, LRPB1, PIK3CA and ALK).

Several studies have been carried out to determine the mutational landscape of brain metastases evolving from primary breast tumours (Cosgrove et al., 2022; Morgan, Giannoudis and Palmieri, 2021; Ali et al., 2021). In summary, a recent meta-analysis

of the genomic landscape of breast to brain metastases, the data from 13 primary breast sequencing studies (comprised of 164 brain metastases) were compared. 268 genes were found mutated in BBM samples of which 22 (8%) were reported mutated in five or more BBM including TP53 (52%), PI3KCA (22%), KMT2C (6%) and ZFHX3(5%) (Morgan, Giannoudis and Palmieri, 2021). These are well-documented tumour initiator genes that are likely to drive primary tumour growth and then persist in metastasising cells (Bailey et al., 2018; Porta-Pardo, Valencia and Godzik, 2020).

Even though the top 20 genes mutated in primary breast cancer were frequently mutated in the BBM samples of this study they were not considered for further analysis as these genes have been widely characterised in primary breast tumours and metastatic tumours and this study aimed to identify rare novel genomic alterations that may be metastasis-specific changes. Therefore, to characterise genomic changes that may be metastasis-specific, variant annotations were used to prioritise variants. MAF <1% is considered an appropriate cut-off value to identify uncommon variants that might be disease-specific (Rabbani, Tekin and Mahdieh, 2013; Niroula and Vihinen, 2019). We hypothesised that metastasis-specific gene mutations will be infrequently found in the primary tumour as they would have developed after the primary tumour cells have left the originating site. In order to identify rare genomic changes that may be involved in brain tumour development and colonisation, a MAF ≤0.1% was applied to the recurrent pathogenic variants. Variants with a MAF value less than or equal to 1% are considered uncommon in the general population (Rabbani, Tekin and Mahdieh, 2013; Lee et al., 2015).

Following the application of the inclusion and exclusion criteria, 286 variants were categorised into 63 candidate genes. All the genes were considered to be suitable metastasis-specific candidates as they have 3 or more deleterious variants with a

MAF≤0.1% which makes them infrequent in the general population. These variants are believed to be the result of late genomic evolution that might have been necessary for metastatic tumour progression and adaptation to the brain microenvironment which would favour brain colonisation by tumour cells.

Twelve genes, BRCA2, HYNDIN, NEFH, RP1L1, TCHH, COL6A3, RYR1, RYR3, HSPG, SCN10A, FAT1, and GOLGA8K were the most frequently mutated across the 26 BBM samples with either MAF ≤0.1% or found not reported on the EXAC database and predicted to be deleterious for the resulting protein. Interestingly, in a systemic review of the genomic landscape of breast to brain metastatic tumours it was found that BRCA2, FAT1 and COL6A3 are associated with 22% of breast to brain metastases. The researchers compiled immunohistochemistry, copy number alteration and prediction analysis of 50 microarray data of breast to brain metastasis studies (Morgan, Giannoudis and Palmieri, 2021). This suggests that these genes can be suitable metastasis-associated genes (Corti et al., 2022).

Additionally, from this list of metastasis-associated candidate genes; KIAA124/ARFGEF3/BIG3 was prioritised for further analysis because it has been described that BIG3 is involved in the regulation of neurotransmitter receptor activity (Liu et al., 2016). Liu and colleagues reported that BIG3 knockout mice showed enhanced inhibitory synaptic activity that resulted from upregulation in postsynaptic GABA receptor activity. Additionally, three separate studies have described that the way tumour cells colonise the brain is by modulating the expression of neurotransmitter receptors (Venkatesh et al., 2019; Venkataramani et al., 2019; Zheng et al. 2019). Thus, these findings suggest that BIG3 may be a metastasis-associated gene.

8.2.1 Amino acid mutation signature of metastasis-associated genes

Presented here was an analysis of the amino acid substitutions identified in 14 metastases-associated genes and how these alterations might be relevant for the metastatic process.

111 amino acid substitutions were computed for these 14 candidate genes. Overall, loss of arginine (R) followed by loss of proline (P) and glutamic acid (E) and significant gain of leucine (L), glutamine (Q) and histidine (H) dominate the landscape of amino acids introduced by mutations (Figure 6.1). Following prediction by either SIFT or Polyphen-2 the list of substitutions reduced to 84 pathogenic amino acid changes with arginine still being the most commonly lost amino acid, followed by loss of glutamic acid (E) and alanine (A) with gains of histidine (H), glutamine (Q) and tryptophan (W) (Figure 6.2). Substitution distribution showed that histidine gain was mainly due to loss of either arginine or glutamine, whereas glutamine gain resulted primarily from loss of glutamic acid, lysine, or arginine (Figure 6.3). In addition, tryptophan gain was exclusively due to loss of arginine. Other amino acids were lost and gained, although at a lower rate (Figure 6.2).

Szpiech and colleagues published data in 2017 in which they utilised a non-negative matrix factorisation (NMF) approach to filter and analyse the amino acid mutation characteristics of a cohort of tumour-normal paired samples across 29 primary cancers. The study found that amino acid substitution introduced by mutations were dominated by arginine loss, and that this loss resulted in gain of glutamic acid, histidine, and tryptophan.

Another study carried out by Tsuber and colleagues in 2017 aimed to determine the somatic evolution of cancer. They found that cancer mutations resulted in a net loss

of arginine with gains of cysteine, histidine, and tryptophan. Importantly, to determine if the losses and gains of specific amino acids apply to other cancer types, they analysed CoSMIC data of all nonsynonymous amino acid changes caused by mutations of single nucleotide substitutions. They concluded that the observed amino acid substitutions are universal, meaning that all cancer tissues undergo marked loss of arginine with significant gains of cysteine, histidine, and tryptophan. Loss of arginine was markedly high in tumour suppressor proteins coded by TP53, OBSCN, LRP1B, and TTN, among others.

It has been established that these amino acid substitutions in cancer cells are not the result of random events; some cancer types have a selection preference for a particular amino acid substitution, with arginine to histidine having the highest frequency of substitution (Alexandrov et al.,2013; Anoosha, Sakthivel and Michael Gromiha, 2016; Tsuber et al., 2017; Szpiec et al., 2017).

In the present analysis, substitutions of arginine to tryptophan, followed by alanine to valine were the most frequently identified changes (Figure 6.3). The previously mentioned studies were all carried out on primary tumour data, which could be the reason for the contrast in the preferred amino acid substitution seen in the metastases data from this study. This difference in the frequency of substitutions for a particular amino acid might reflect specific functional and nutritional requirements specific to the tumour microenvironment. As with arginine to histidine, arginine to tryptophan substitution alters the charge of the residue, thus it is expected to significantly affect the protein structure and function.

8.2.2 Mutation distribution within protein domains of metastasis-associated genes

Cancer studies have been focused on understanding the mutational landscape of cancer by analysing gene expression patterns across tumours (Lawrence et al., 2013; Alexandrov et al., 2013). Nevertheless, genes can carry out multiple molecular roles in cancer progression, it can be the case that it is not the gene itself that is driving cancer development and differentiation but a specific gene function that might be conferring an advantage to cancer cells (Peterson et al., 2010). The categorisation of cancer-associated variants concerning the relationship within the protein domains might be useful as it allows systemic assessment of genes' common biological functions (Cheng et al., 2014).

To assess if the identified variants are likely to be driving metastasis, the conserved domains of 14 proteins of interest were analysed. It is hypothesised that for these changes to be metastasis drivers they should be located within or close to the protein regions that determine protein function (conserved domains), suggesting that alterations in these parts of the protein could lead to modification in the protein consequence that might be supporting the metastatic process (Miller et al., 2015).

For this analysis, the information in the NCBI CDD (Conserved Domain Database) (https://www.ncbi.nlm.nih.gov, n.d.) was used for finding highly conserved domains of the proteins of interest. To search for the conserved domains a query containing the protein in FASTA format was submitted to the CDD database. With the obtained results, a diagram for each protein was constructed, conserved domains were plotted, and mutated variants identified in the BBM samples from this study were mapped to establish their relative position within the protein structure. Interpro (Blum et al., 2020),

a database containing information on protein families and domains was used to determine the function of conserved protein domains. The mutation aligner data base (www.mutationaligner.org, n.d.) was used to determine domain mutation hot spots in cancer. Interpro (Interpro EMBL-EBI, 2019) database for the classification of protein families was used to obtain information about domain function. Additionally, the mutation distribution of the BBMs was compared with the mutational distribution within the same protein position in primary tumours that recurrently metastasise to the brain (i.e.. breast, lung, and skin). Data on the primary tumours' mutation distribution was obtained from the CoSMIC database.

Most of the domains identified in this study have not been implicated in cancer susceptibility, and the mechanism by which this group of proteins might be involved in the metastatic process are not completely understood. However, cancer studies have indicated that mapping mutations to specific protein domains can help to identify rare potential cancer drivers that could have been overlooked by gene focus analysis, as domain studies can provide a wider understanding of the functional consequence of mutations (Nehrt et al., 2015; Yang et al, 2015; Hashemi et al., 2017). The results of the present study may be a step forward towards an improved understanding of the complexities of the metastatic process by identifying domain mutational patterns that could be specific to breast to brain metastatic tumours. These results could be further used to create a database of the mutational landscape of brain metastatic tumours that could aid in the development of target therapy to treat these malignancies by supplying a more specific target region.

8.2.3 Potential role of BIG3 in the regulation of neurotransmitter subunits

BIG3 gene mutations were identified in breast to brain metastatic tumours after Whole Exome Sequencing. Afterwards, CRISPR cas9 editing of MCF7 breast cancer cell lines was carried out to elucidate the role of BIG3 in breast to brain metastases. Expression of NMDA (N-methyl-D-aspartate receptors) (GRIN2B, GRIN2C, GRIN2D) and GABA (GABRG1, GABRA2, GABABR1, GABRA4) neurotransmitter receptor subunits in a BIG3 knockout/knockdown MCF-7 breast cancer cell line was determined using RT-PCR. The results confirmed that neurotransmitter subunits were differentially expressed in BIG3 knockout and knockdown MCF7 cells compared to those expressing endogenous BIG3. Expression of glutamate receptor subunit GRIN2C, and GABA receptor subunit GABRA2 was observed only in BIG3 knockout MCF7. GABRG1 expression was seen in both the knockout and the knockdown. GABBR1 and GRIN2D subunits were found expressed in the wild type as well as the knockout and knockdown. GRIN2B upregulation was observed in MCF7 expressing wild type BIG3.

These mutations are thought to be involved in tumour-neuronal network integration via upregulation of neurotransmitter receptors subunits, which has been reported by Venkatesh et al., 2019; Venkataramani et al. 2019 and Zheng et al., 2019 as a mechanism used by primary and metastatic brain tumours to adapt to the tumour microenvironment to sustain tumour growth and colonisation of the brain. BIG3 upregulation of neurotransmitter subunits suggest that BIG3 mutations may be one the ways through which breast to brain metastases achieve colonisation of the brain.

8.3 Further directions

ATF2 can represent the future for antiangiogenic therapy. Here it was shown that this transcription factor may play a role in angiogenesis by negatively regulating DLL4 Notch signalling pathway ligand in endothelial cells. Future work would include establishing ATF2 binding sites in the regulatory region of DLL4 to determine the molecular mechanisms behind this interaction. Additionally, the effect of ATF2 suppression in DLL4 target genes could be determined to further understand the implications of ATF2 in the regulation of angiogenesis.

Organotypic coculture assay showed that ATF2 participates in tubule formation in endothelial cells. ATF2 suppression resulted in significantly decreased tubule development in GFP-HUVECs lacking functional ATF2 and stimulated with the pro-angiogenic molecule bFGF when compared to the control. Therefore, in future work, coculture assays could be used to characterise the breast to brain metastasis angiogenic process, as coculture studies have been shown to be useful when determining the angiogenic potential of tumours and also assessing their sensitivity to antiangiogenic therapy. Additionally, coculture assays can be used to assess tumour and extracellular matrix interactions which could provide a better understanding of tumour interactions with their microenvironment.

Particularly, functional work is needed to determine the consequences on the proteins' structure and function due to substitutions in the amino acids. Additionally, determining if these substitutions are characteristic of metastatic tumours and if they happen in combination with other mutational patterns might provide an insight on the selective pressures that take place during tumour evolution.

It would be ideal to further explore the distribution of mutations in protein domains with a larger group of samples of matched primary and metastatic tumours from different tissue sites to be able to create a comprehensive profile of the mutational pattern of metastatic tumours.

Future work will be required to look in-depth into the domain architecture by multiple domain sequence alignment and by characterising the functional implications of these mutations by performing mutational pathway alignment followed by laboratory confirmation of these amino acid alterations.

Additional experiments are required to establish a pattern of neurotransmitter subunits regulated by BIG3. It was previously described that BIG3 is found in the lysosomes of hippocampal neurons, therefore, determining the expression of lysosomal marker (Lamp1) in BIG3 expressing cells and knockout/knockdown cell lines would be the next step, in order to further assess the role of BIG3 in the regulation of neurotransmitter receptors in breast to brain metastases. Additionally animal xenograft studies with a BIG3 wild type and knockdown genotype could be beneficial in determining if BIG3 promotes brain tumour development.

8.4 Limitations

The data analysis was limited in part by not having matching normal tissue for most of the samples used in this research. The addition of normal tissue sequencing analysis would have made it possible to accurately filter out germlines from somatic mutations. However, gathering the appropriate set sample for cancer studies can be difficult due to the amount of time between the primary malignancy diagnosis and the metastatic tumour development. Also, to add statistical significance to the identified variants, a

larger number of sample sets that include tumours and matched normal tissue would

need to be sequenced to generate a larger cohort study.

Appendix

Appendix A.1– List of Primers Used for Sanger Validation

Gene	Exon	Primer	Primer sequence
BIG3/ARFGEF3	19	F	CATGAGAACACAGGAGGGAAGGACTG
		R	CGTCCCTGGCATGACTGAAACCA
	21	F	GCTAAGTGAACCACCACATCCTTCTTG
		R	GCCATGAGGTATAACAAAACAGTG
	24	F	GCACCATGCTGGAGCCAAGCC
		R	GATGCGTGCTTGTTCCAGCACGTCA
	28	F	GCCACTGAAGACCATATATTCCCCCAT
		R	GAAGGCTGGCACTGACCATGTTGT
	29	F	GGCAGGAGGAGGATATGGCCATT
		R	GCTGACTGAACACTGAAATGAAGATT
	30	F	GGAATTCAAGACTCCTTTGTGG
		R	GAGACAAAAGTACTGACCCAG
	34	F	CTCTCAAGTGTTCAAATAAAACCAGGG
		R	CAAACCTTAGTATGGCCCCAATGC
HERC1	16	F	GATCTATCAGGGACCTTGAAACTCAG
		R	GCTAACACTACCCCCAACCACGA
	33	F	GGTATAATCAAACTTCGCATGCGAGTT
		R	GCTGTTAAGAAATACCTTTTAGGAAAAACC
	43	F	CTGTTCAAGGGATAAAGATGTGTTTCAAAT
		R	CTGTTAATGTGGAGGGACTGAGCA
	70	F	CTTATTTATGTCTGCCACCATCTAAGAAAAT
		R	CATGGGGTACTGCTAAAAAGAAAGG
KMT2D	29	F	GGGATGAGGGCAAGGGAGACA

		R	GAGGGATATGGGACACAGCCTTA
	31	F	GGCACCTTCTCCTCCAAGTCAC
		R	GAACCGACGGAGGGCGTAGT
BRCA2		F	CACAGTTTTGGTAGCTTC
	11 (V1)	R	CAGCATCTCTGCATTCCTCA
	11 (V2)	F	GATTGGTCAGGTAGACAGCAG
		R	GTAATTTCTTCAACAAAAGTGCAAG
BRCA2	18	F	GAAACAATATATTCCTAGCTACA
		R	GTACATCTAAGAAATTGAGCATC
	22	F	GTGAGAAACTGATTACATTAACCA
		R	GTGGATTTTGCTTCTCTGATATAAA
MATN2	4	F	GAGGGTAGGGACTGCAGCTAA
		R	GTGGGTGAGGCATCACAGATG
	11	F	GTGTACTGTAGGATAAGGGCATG
		R	GCTGTGTGGCATTGACTTCTG
	13	F	GCTTTTTCCCAACCCTGAGTATGA
		R	GTGTTTTCACAGGTGCTGCAG
COL6A3	7	F	GCTCAGTACAGCGATGATGTCAAG
		R	TCTGCTTCAGGTTACTTGCTGGC
	9	F	GGTGTTAGGATCCCCTGCCTG
		R	TGCACTTCATCCTTGCTGGAATGG
	36	F	CAAGACACTTTCGGCCGGATGC
		R	TCTCTGAGCTGTGGGGATGCTC
NEFH	2	F	CATAGGCAGTCTGGCTGTCTGC
		R	CATTAGAACCCAGTCCAGGTGTGT

Appendix A.2- List of primers used for RT-PCR

Gene	Primer	Primer Sequence
BIG3/ARFGEF3	F	GAGCTGCACCTGGGCCCT
	R	GAACTCTGATGTGGGTCAC
GABRG1	F	TGTGACTGCGATGGATCTCT
	R	AGTTGAACAGGGCAAAAGCG
GABRA2	F	TGTGCCATGCCCTGTAATTG
	R	AACATGGCTCAAGGGGATCA
GABBR1	F	AACCAGACCATTACCGACCA
	R	TTGGGCTGTGAGTTCTGGAT
GABRA4	F	AGTGTGGACCCCTGATACTT
	R	TCCGTCTGAGGTGGAAGTAA
GRIN2B	F	GCCAACTTAGCTGCCTTCAT
	R	CGCTTCCACCCAGAATCTTT
GRIN2C	F	CGCTGGTCTTCAACAACTCA
	R	AGTTGAGGACAGCAGCATCA
GRIN2D	F	GTCGCCGTCACTGTTTTCAT
	R	GTTGCTGCGGATGTTCTTCT

Appendix B

Appendix B.1- DLL4 expression data in HUVECs infected with ad-GFP and AD_ATF2_AA

DLL4 Expression	0 hours		0.5 hours		1 hour		3 hours	
	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)
Exp1A	0.944408761	1.192205179	4.757963738	7.76018949	3.513588813	5.185866803	1.255132656	2.375628319
Exp1B	1.058863536	1.402868205	5.787915507	6.488360782	3.018003651	5.760151809	1.222433361	2.88198418
Exp2A	0.992864569	1.681422666	2.251026079	5.557377856	3.115719739	4.529125754	1.085844361	2.383291531
Exp2B	1.007186712	1.893963665	2.733174811	5.187917181	3.072632798	9.736777703	0.990951273	2.465276097
Exp3A	0.955587367	1.754203124	3.497162341	6.198707997	3.283526732	5.264051039	2.765654329	5.132911729
Exp3B	1.04647679	2.027171726	3.337737787	6.480229564	3.585064447	9.908667262	2.412321145	4.722902048
Mean	1.000897956	1.658639094	3.72749671	6.278797145	3.26475603	6.730773395	1.622056188	3.326998984
SD	0.046444808	0.311070691	1.317633896	0.89396805	0.238695398	2.427444779	0.763214679	1.260647501
SEM	0.018961014	0.126994078	0.537921785	0.364960928	0.097446988	0.991000181	0.311581088	0.514657187

Appendix B.2-DLL1 expression data in HUVECs infected with ad-GFP and AD_ATF2_AA

DLL1 Expression	0 hours		0.5 hours		1 hour		3 hours	
	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)
Exp1A	1.160911218	NA	3.483816694	7.525149667	3.290553739	4.481128906	1.050753662	1.922284008
Exp1B	0.861392314	3.020428563	2.692028664	6.882388791	3.010760491	8.164624411	0.981361461	1.41765049
Exp2A	1.043598194	2.916083634	1.66172347	5.997142467	2.683083626	6.898024999	0.63326881	1.799038602
Exp2B	0.958223199	2.78965794	1.511867384	5.051283081	2.449350787	11.06998375	0.641178532	1.418487584
Exp3A	1.202803767	2.916083634	2.901307298	3.371038027	2.788007989	7.487033189	1.005605378	2.683401122
Exp3B	0.831390812	2.78965794	2.486656234	7.550062815	NA	6.820873888	1.461047206	3.215373016
Mean	1.009719917	2.886382342	2.456233291	6.062844142	2.844351327	7.486944858	0.962202508	2.076039137
SD	0.153454997	0.098035634	0.752810249	1.630656488	0.320934523	2.150480132	0.306656807	0.725698265
SEM	0.06264774	0.043842869	0.307333497	0.665712724	0.143526282	0.877929837	0.125192117	0.296265076

Appendix B.3- SNAI2 and PPARG expression data in HUVECs infected with ad-GFP and AD_ATF2_AA

SNAI2 Expression	0 hours		0.5 hours		1 hour		3 hours	
	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)
Exp1A	0.713139699	NA	1.128407215	2.29435047	0.54390967	1.498779259	0.581175881	0.899518706
Exp1B	1.4022498	0.274312823	0.730547749	2.415863829	0.54390967	1.047321077	0.458431105	0.527193531
Exp2A	0.505894231	0.049504155	0.629371002	0.039512692	0.961691739	0.0702611	0.865474962	0.427917037
Exp2B	1.976697774	0.366820306	0.866106802	0.122324014	0.180727985	0.06614136	0.414145465	0.484365282
Exp3A	0.62352471	0.266469788	2.594298124	1.896835376	0.983388193	0.221434532	1.659762778	0.606218287
Exp3B	1.60378568	0.357419637	2.459582209	0.214735268	2.731827323	0.465477966	4.750314504	0.535054699
Mean	1.137548649	0.262905342	1.401385517	1.163936942	0.990909097	0.561569215	1.454884116	0.58004459
SD	0.605796301	0.127880974	0.888807168	1.151743634	0.904181055	0.587550014	1.678610228	0.167250364
SEM	0.247315304	0.05719011	0.362854007	0.47019737	0.36913037	0.239866289	0.685289756	0.068279675

PPARG Expression	0 hours		0.5 hours		1 hour		3 hours	
	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)	Ad-GFP	Ad-ATF2(AA)
Exp1A	0.881756882	NA	1.27534476	3.203860823	0.405671889	1.119547844	1.114311607	1.061270588
Exp1B	1.134099456	0.572526945	1.591950459	4.102862002	0.314321886	1.961374057	1.20756307	1.006190521
Exp2A	0.500422912	0.210262045	3.572046804	0.304840828	4.431988037	0.528731584	1.308328548	1.04040629
Exp2B	1.998309783	0.659188374	2.037092078	0.430949335	1.964894726	0.614149729	1.368219856	1.78757542
Exp3A	0.747978154	0.622615708	4.307023993	1.578990295	5.751974932	0.419844418	1.875332556	1.022017397
Exp3B	1.336937443	0.431989521	4.508034958	0.413161901	6.085932306	0.582584447	1.49123494	0.50734801
Mean	1.099917438	0.499316518	2.881915509	1.672444197	3.159130629	0.87103868	1.394165096	1.070801371
SD	0.528262184	0.183171286	1.421990569	1.628394399	2.608058543	0.586606551	0.269118576	0.409588541
SEM	0.215662133	0.074779364	0.580525219	0.664789229	1.064735442	0.239481121	0.109867199	0.167213821

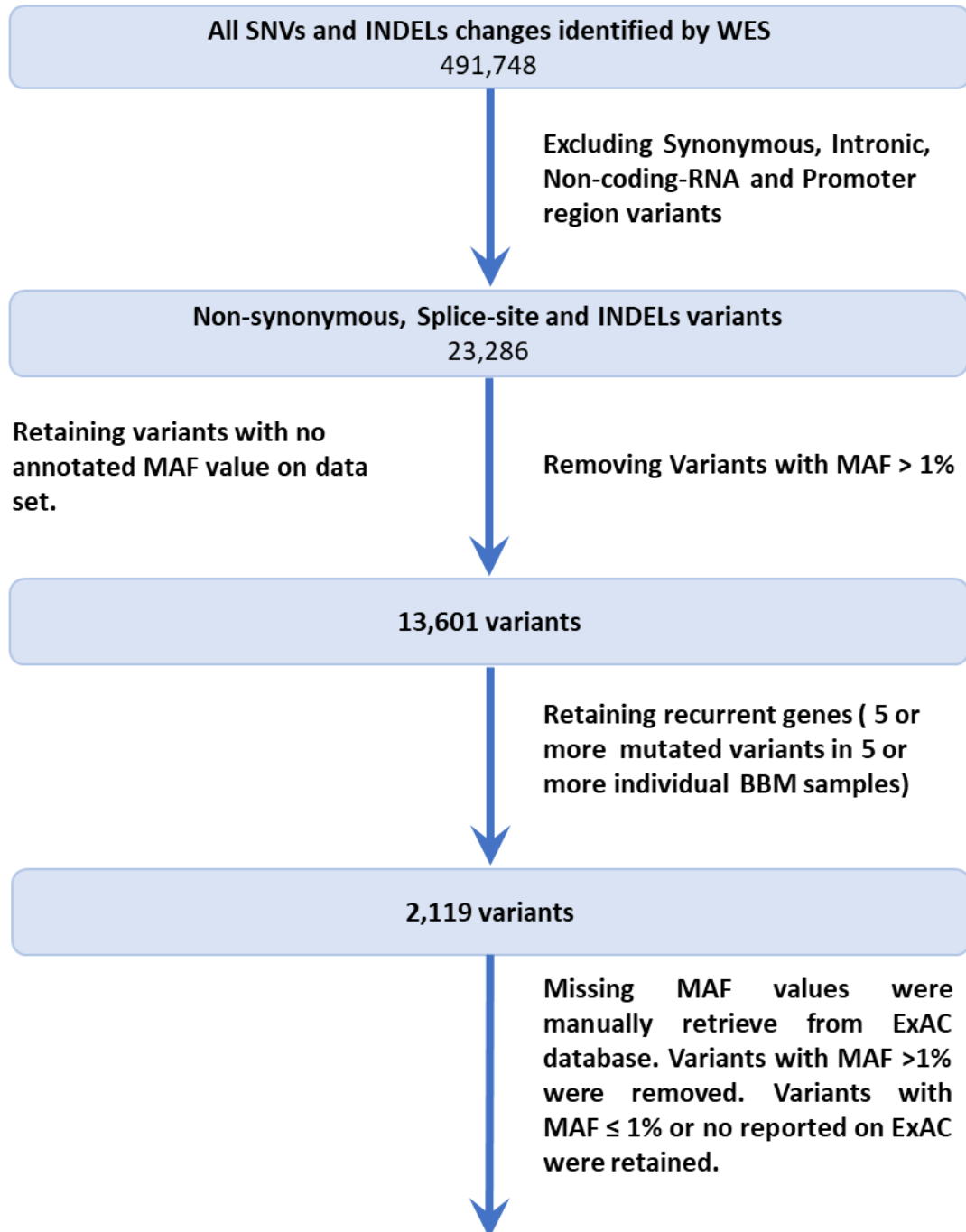**Appendix B.4- DLL4 expression data in HUVECs transfected with si_ATF2 or si_NT**

DLL4 Expression	0 hours		0.5 hours		1 hour		3 hours	
	si-NT	si-ATF2	si-NT	si-ATF2	si-NT	si-ATF2	si-NT	si-ATF2
Exp1A	0.783551442	1.479904494	2.601213332	5.12409062	3.632084789	5.4740698	2.62858284	3.555149622
Exp1B	1.276240393	1.385887294	2.642810362	4.831472148	5.043972445	7.213047006	1.013561742	4.387664968
Exp2A	0.786721514	1.874769833	3.551687285	11.51215047	3.81623971	4.843450874	1.553566371	1.361783646
Exp2B	1.271097818	2.753195876	3.004553422	6.20145398	1.856871221	6.655295814	0.847224711	2.958729174
Exp3A	0.68490449	4.687540941	2.960554524	10.52127706	5.983971586	11.14092803	1.167161614	1.472785185
Exp3B	1.460057591	3.316586457	3.86112521	5.27414845	5.097890006	12.60340539	1.614053202	1.382979828
Mean	1.043762208	2.582980816	3.103657356	7.244098788	4.23850496	7.988366153	1.470691747	2.519848737
SD	0.329116349	1.27653412	0.50390839	2.974531125	1.459603159	3.156821443	0.64155156	1.302545906
SEM	0.134361187	0.521142872	0.205719739	1.214347247	0.595880494	1.288766958	0.261912328	0.531762139

Appendix B.5- DLL1 expression data in HUVECs transfected with si_ATF2 or si_NT

DLL1 Expression	0 hours		0.5 hours		1 hour		3 hours	
	si-NT	si-ATF2	si-NT	si-ATF2	si-NT	si-ATF2	si-NT	si-ATF2
Exp1A	0.643853031	1.814659979	1.335563012	4.017314219	2.16659987	3.101637188	1.50044464	1.594197733
Exp1B	1.55314948	1.971403717	1.353289062	2.883941379	1.428275806	4.503585071	1.249140887	1.608843185
Exp2A	1.361720634	0.940709622	1.978542345	2.091459912	1.526903677	1.716832053	0.677760985	1.584119345
Exp2B	0.734365019	0.837571162	0.97117151	0.715996502	2.427951124	2.274595199	0.549674739	1.07649552
Exp3A	1.212610505	0.651867374	2.481628098	2.265292014	2.596962772	1.233879643	1.215382326	0.931628803
Exp3B	0.824667109	1.042080924	1.989309304	1.521334232	1.869584386	3.242535959	1.192997474	0.966820955
Mean	1.055060963	1.209715463	1.684917222	2.249223043	2.002712939	2.678844186	1.064233509	1.293684257
SD	0.372014344	0.547042634	0.557687549	1.13443968	0.476427327	1.183468525	0.368246861	0.334389221
SEM	0.15187422	0.22332922	0.227674989	0.46313306	0.194500642	0.483149002	0.150336151	0.136513828

Appendix C

Appendix C.1- WES Pipeline – Detailed Version

All SNVs and INDELs changes identified by WES
491,748

Excluding Synonymous, Intronic, Non-coding-RNA and Promoter region variants

Non-synonymous, Splice-site and INDELs variants
23,286

Retaining variants with no annotated MAF value on data set.

Removing Variants with MAF > 1%

13,601 variants

Retaining recurrent genes (5 or more mutated variants in 5 or more individual BBM samples)

2,119 variants

Missing MAF values were manually retrieve from ExAC database. Variants with MAF >1% were removed. Variants with MAF ≤ 1% or no reported on ExAC were retained.

```
                        ┌─────────────────────────────────────┐
                        │           1127 variants             │
                        └─────────────────────────────────────┘
                                         │
Excluding individual variants present              Retain recurrent genes
in more than 3 BBM samples.                        │
                                         ▼
                        ┌─────────────────────────────────────┐
                        │            985 variants             │
                        │ Missense (846), Exon InDels (102),  │
                        │ nonsense (21), Splice–Site (16)     │
                        └─────────────────────────────────────┘
                                         │
                                         │    Pathogenic prediction of 846
                                         │    missense variants by either SIFT
                                         │    or Polyphen-2. Variants predicted
                                         │    to be benign by either tool were
                                         │    excluded.
                                         ▼
                        ┌─────────────────────────────────────┐
                        │      593 Pathogenic variants        │
                        │ (454 Missense pathogenic, 102 Exon  │
                        │ InDels, 21 nonsense, 16 splice-Site)│
                        └─────────────────────────────────────┘
                                         │
Retain variants with MAF either                 Retain genes with 3 or
≤ 0.1% or no reported                            more damaging variants
                                         ▼
                        ┌─────────────────────────────────────┐
                        │    286 variants = 63 candidate genes│
                        └─────────────────────────────────────┘
```

Appendix D- Bioinformatic Filtering of nonsynonymous variants

Appendix D.1- Total Nonsynonymous variants

https://livewlvac-my.sharepoint.com/:x:/g/personal/i_olivares_wlv_ac_uk/Ebr6ZuMl1dhHtN0882CrfesBAPx2rrYl_W_9zQwenAbsMQ?e=g5pswa

Appendix D.2- Variants with MAF ≤1% or no reported on ExAC

https://livewlvac-my.sharepoint.com/:x:/g/personal/i_olivares_wlv_ac_uk/EREgM8Ock0pFhtAriqZi_ZABzus8ccBXXuV94o4c_zRNyg?e=Myg28U

Appendix D.3- Recurrent genes: more than 5 variants with MAF<1% or no reported value on the data set. MAF annotation from the ExAC for the missing values

https://livewlvac-my.sharepoint.com/:x:/g/personal/i_olivares_wlv_ac_uk/EW-kshSPpRNDv-ePlE7wUNYB0-TS0qBZDyFJFsXdGYYuXQ?e=jbOuMv

Appendix D.4:

C.4.1: Variants with MAF ≤1% or not found reported on ExAC.

C.4.2: Remove variants with more than 3 samples per variant.

C.4.3: Retain recurrent genes (5 or more variants)

https://livewlvac-my.sharepoint.com/:x:/g/personal/i_olivares_wlv_ac_uk/EcqZHV2dmYtLjPFRHxeT5DoBwl96OTqthuw3nbUGRN9PKQ?e=XtYzr6

Appendix D5:

D.5.1- Recurrent variants

D.5.2- Pathogenic prediction of Missense variants

D.5.3- Missense predicted pathogenic variants

D.5. 4- All damaging Mutations

https://livewlvac-my.sharepoint.com/:x:/g/personal/i_olivares_wlv_ac_uk/EbeGoVw96RNBrEatnQ9JGdMBQVmeLmWhB7kNY7fuUKqhOg?e=K52n7n
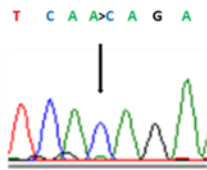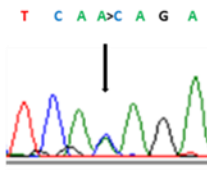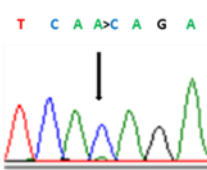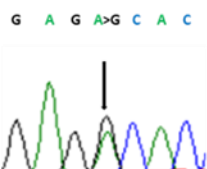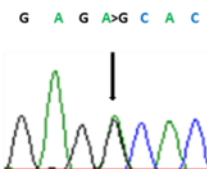
Appendix D6
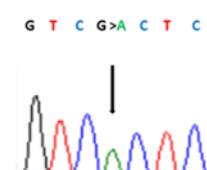
D.6.1- Retain genes having 3 or more damaging variants

D.6.2- Candidate gene tables
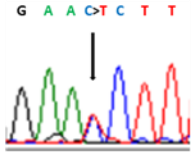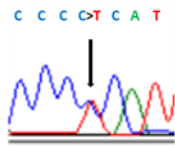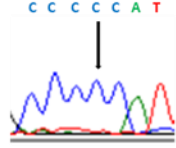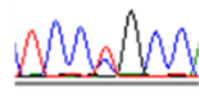
https://livewlvac-
my.sharepoint.com/:x:/g/personal/i_olivares_wlv_ac_uk/EePrqq_7pJpPkNVGo8FHFV4BqEU
cIF43gV59iKb1Hdzt3w?e=2v9zV1

Appendix E- Sanger sequencing electropherogram of gene mutations identified by Whole Exome Sequencing
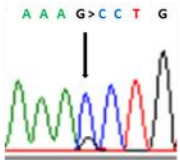
Appendix E.1- Sanger sequencing electropherogram results showing mutations identified by Whole Exome Sequencing in HERC1 gene

Gene	Nucleotide and protein change	Tumour	Matched blood	Matched Primary tumour
HERC1	c.A3113C p.K1038T			
	c.A5983G p.T1995A			Unavailable
	c.G8618A p. R2873H		No available	No available
	c.A12982G; p.T4328A		No available	No available
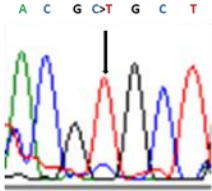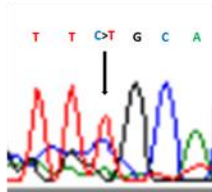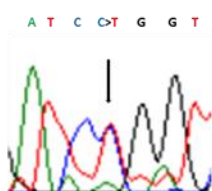
Appendix E.2- Sanger sequencing electropherogram results showing mutations identified by whole exome sequencing in KMT2D gene.

Gene	Nucleotide and protein change	Tumour	Matched blood	Matched Primary
KMT2D	c.G6154A (C>T reverse) p.V2052I		Unavailable	Unavailable
	c.C7144T p.P2382S			Unavailable
	c.C7670T P.P2557L		Unavailable	Unavailable
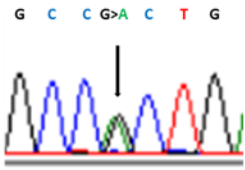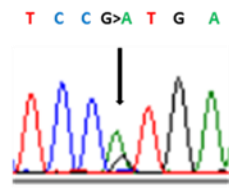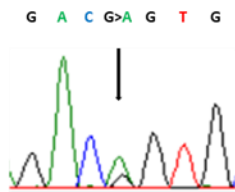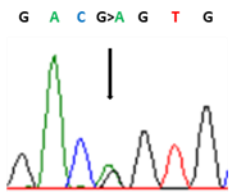
Appendix E.3- Sanger sequencing electropherogram showing mutations identified by whole exome sequencing in BRCA2 gene.

Gene	Nucleotide and protrein change	Tumour	Matched blood	Matched Primary tumour
BRCA2	c. G3709C p. A1237P	AAA G>CCT G	Unavailable	Unavailable
	c.8933C>CA	AAAAAAA AAAAAAAA TTC>CAAG	AAAAAAAGA AAAAGATTCAG	Unavailable
	c.C3262T; p. P1088S	ACCC>TCTCA	Unavailable	Unavailable
	c.G8090A p.S2697N	TGAG>ACGC	Unavailable	Unavailable
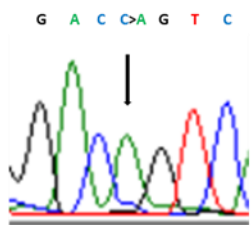
Appendix E.4- Sanger sequencing electropherogram showing mutations identified by whole exome sequencing in COL6A3 gene.

Gene	Nucleotide and protrein change	Tumour	Matched blood	Matched Primary tumour
COL6A3	c.C2657T p.A919V		Unavailable	Unavailable
	c.C3754T p.R12552C		Unavailable	Unavailable
	c.C7375T p.R2459W		Unavailable	Unavailable

Appendix E.5- Sanger sequencing electropherograms showing mutations identified by whole exome sequencing in MATN2 gene.

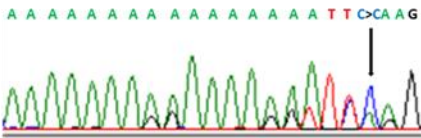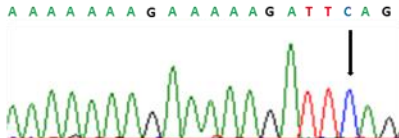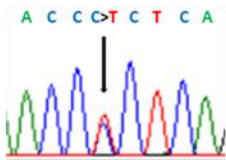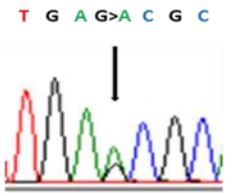| Gene	Nucleotide and protrein change	Tumour	Matched blood	Matched Primary tumour
MATN2	c.G944T (Reverse DNA strand) p.T238M		Unavailable	Unavailable
	c.G1904T p.R558H		Unavailable	Unavailable
	c.G2165A p. R645Q		Unavailable	

Appendix E.6-Sanger sequencing electropherograms showing mutations identified by whole exome sequencing in NEFH gene.

Gene	Nucleotide and protrein change	Tumour	Matched blood	Matched Primary tumour
NEFH	c.C1054A p.R352S		Unavailable	Unavailable

Appendix F – Mutations Mapped onto Protein Domains

Appendix F.1- BRCA2
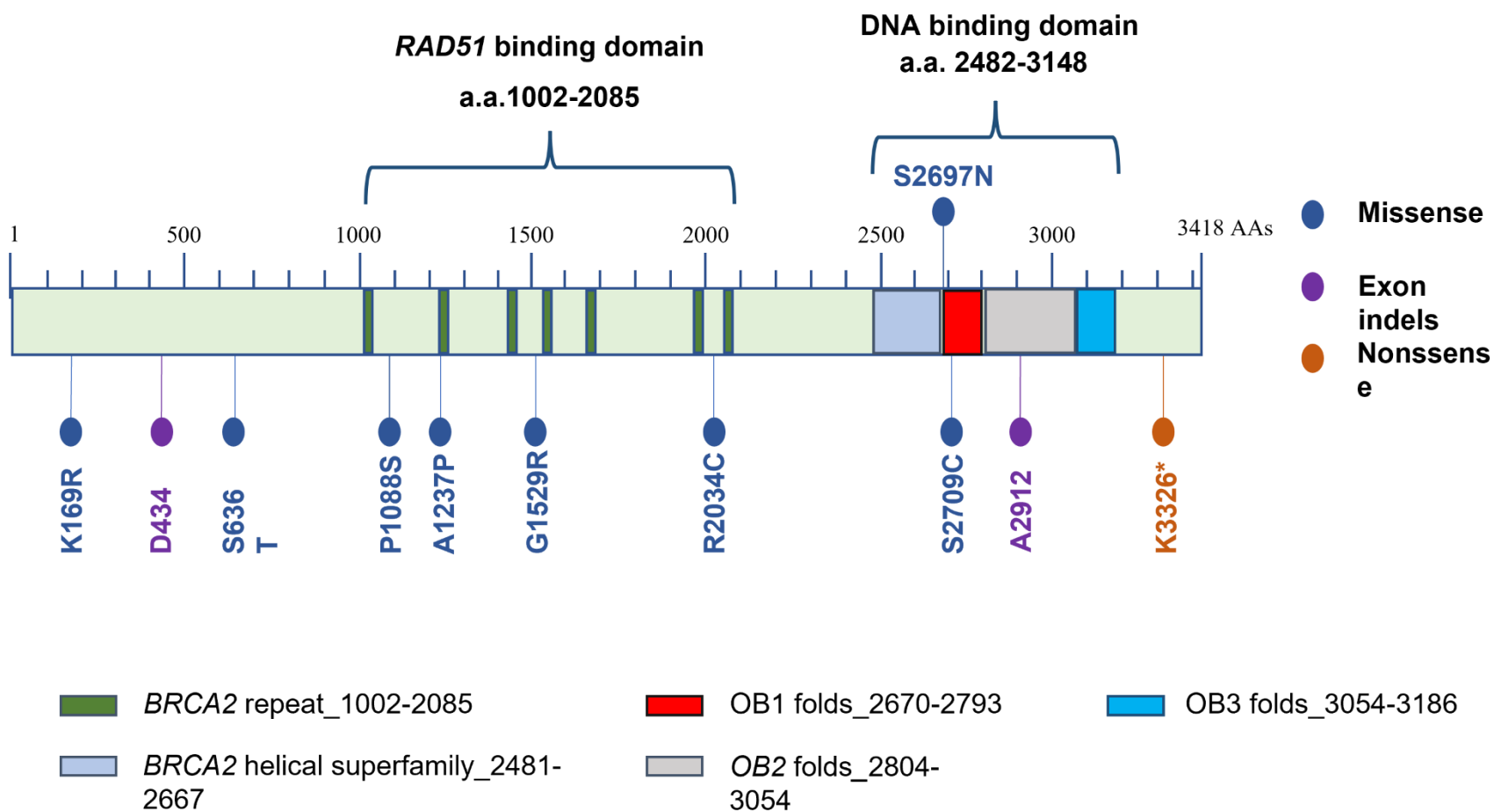
BRCA2

Lung
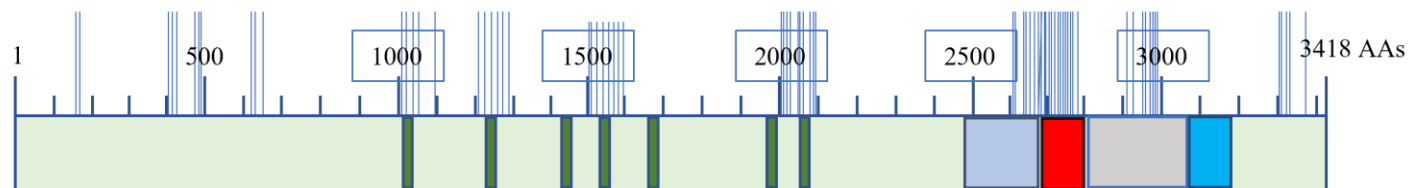
BRCA2

Breast
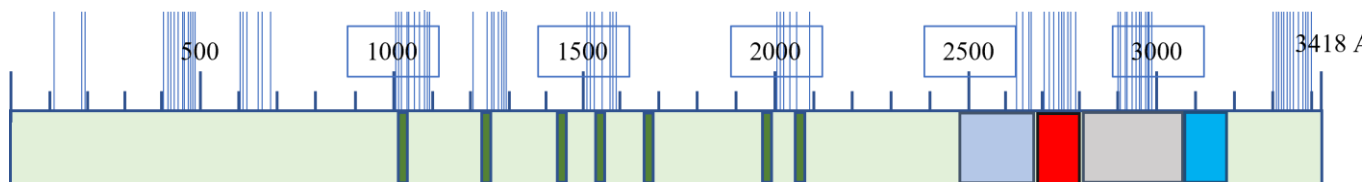
BRCA2

Skin

APPENDIX F.1: Schematic representation of BRCA2 mutations distribution in within the protein domains. BRCA2 protein is formed by 3418 amino acid residues and five conserved domains. A) Amino acid changes are shown in blue, purple, and orange corresponding to missense, frameshift, and nonsense mutations, respectively. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.

Appendix F.2: RP1L1
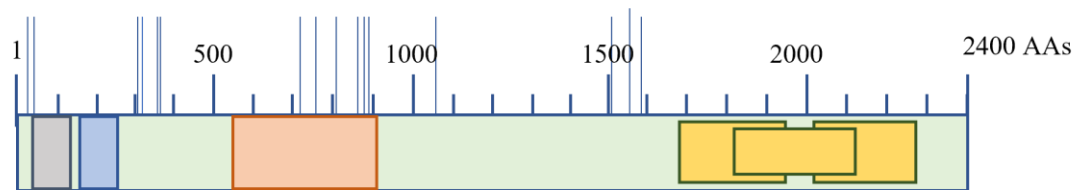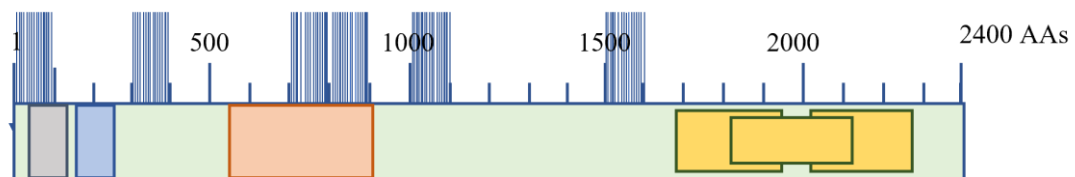
APPENDIX F.2: Schematic representation of RP1L1 mutations distribution in within the protein domains. RP1L1 is protein is formed by 2400 amino acid residues and four conserved domains. A) Amino acid changes are shown in blue, purple, and orange corresponding to missense, frameshift, and nonsense mutations, respectively. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.

Appendix F.3: COL6A3
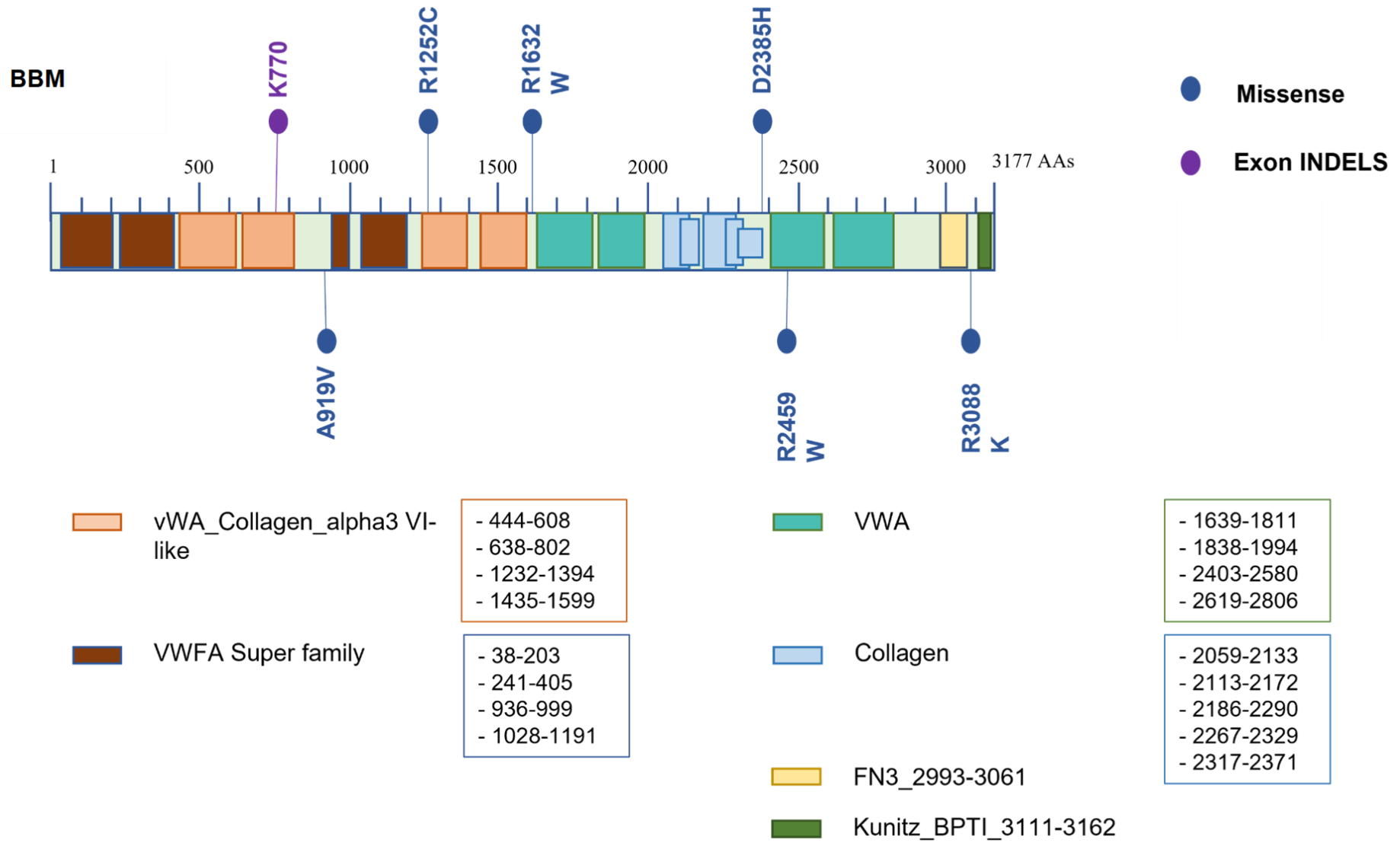
COL6A3

Lung
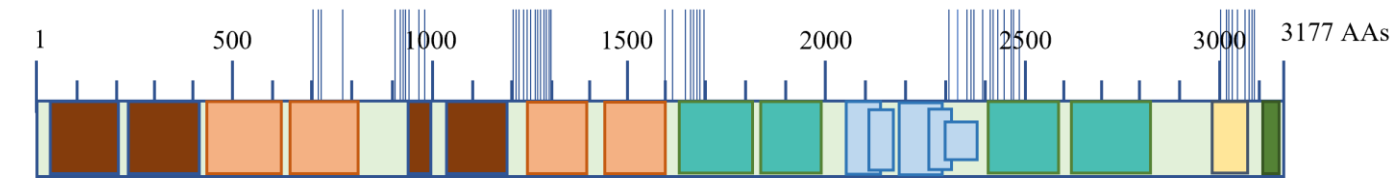
COL6A3

Breast
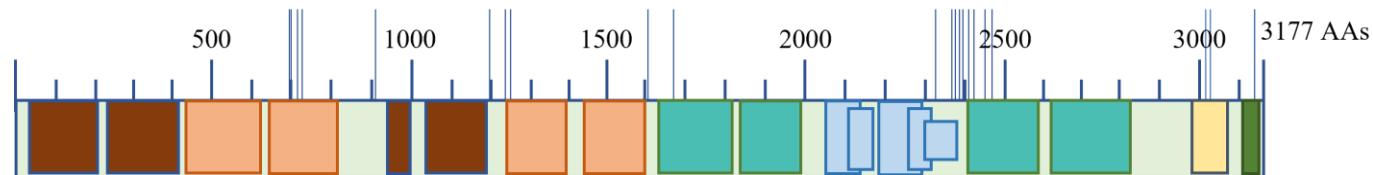
COL6A3

Skin
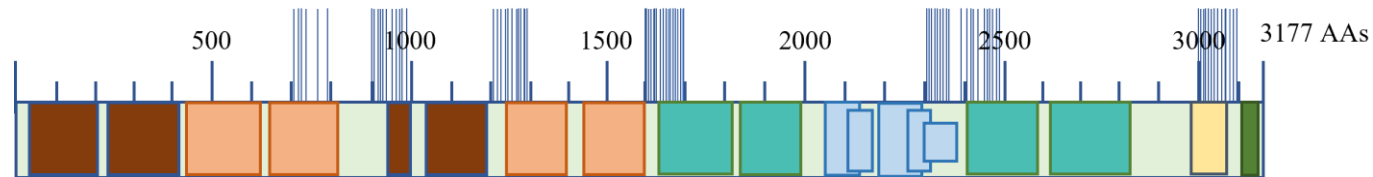
APPENDIX F.3: Schematic representation of COL6A3 mutations distribution in within the protein domains. COL6A3 is protein is formed by 3177 amino acid residues and five conserved domains. A) Amino acid changes are shown in blue and purple, corresponding to missense and frameshift and mutations, respectively. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.

BBM

W444*
S673F
P1019L
K1296Q
R1919C
G2032A
R2977W
V3760M
A4071V

1   500   1000   1500   2000   2500   3000   3500   4000   4391 AAs

T361N
R1090Q
R2004H
Q2618L
G2950R
A3396V
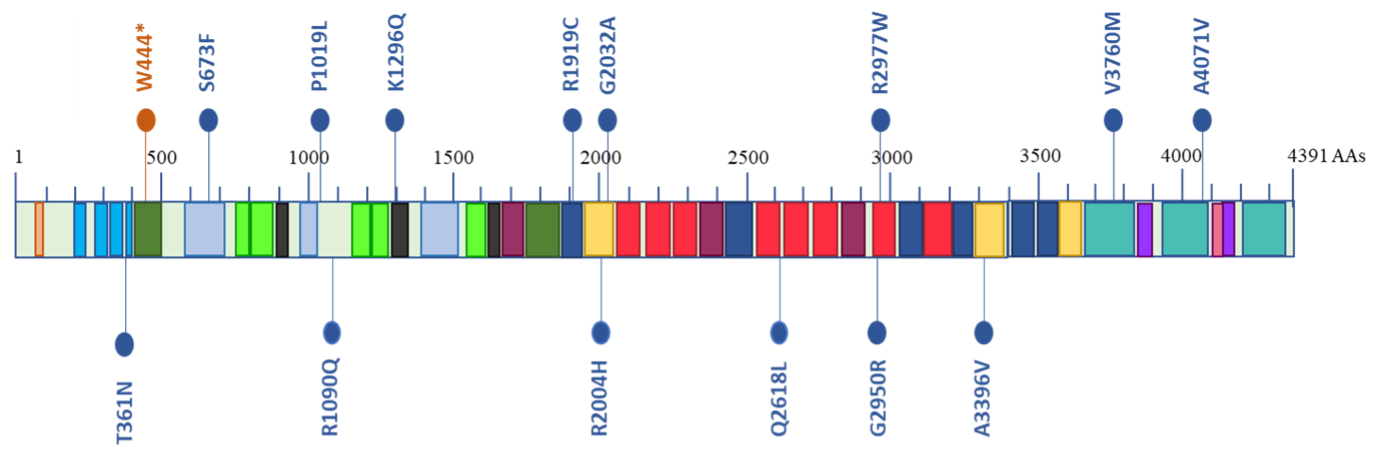
	SEA_ 80-94			I-set		- 1956-2042
						- 3299-3382
						- 3581-3658

	Lam B	- 590-716
		- 985-1112
		- 1391-1516

	IG like	- 2057-2134 - 2548-2617 - 2932-3009
		- 2158-2235 - 2636-2713 - 3120-3202
		- 2251-2325 - 2733-2810

| | Igl_Perlecan like | - 421-498 |
| | | - 1771-1856 |

	LamG	- 3665-3825
		- 3934-4083
		- 4203-4362

	Ig Super family	- 1866-1950 - 3221-3295
		- 2437-2520 - 3405-3483
		- 3022-3108 - 3493-3572

Appendix F.4: HSPG2

208

APPENDIX F.4: Schematic representation of HSPG2 mutations distribution in within the protein domains. HSPG2 is protein is formed by 4391 amino acid residues and thirteen conserved domains. A) Amino acid changes are shown in blue and orange corresponding to missense and nonsense mutations, respectively. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.
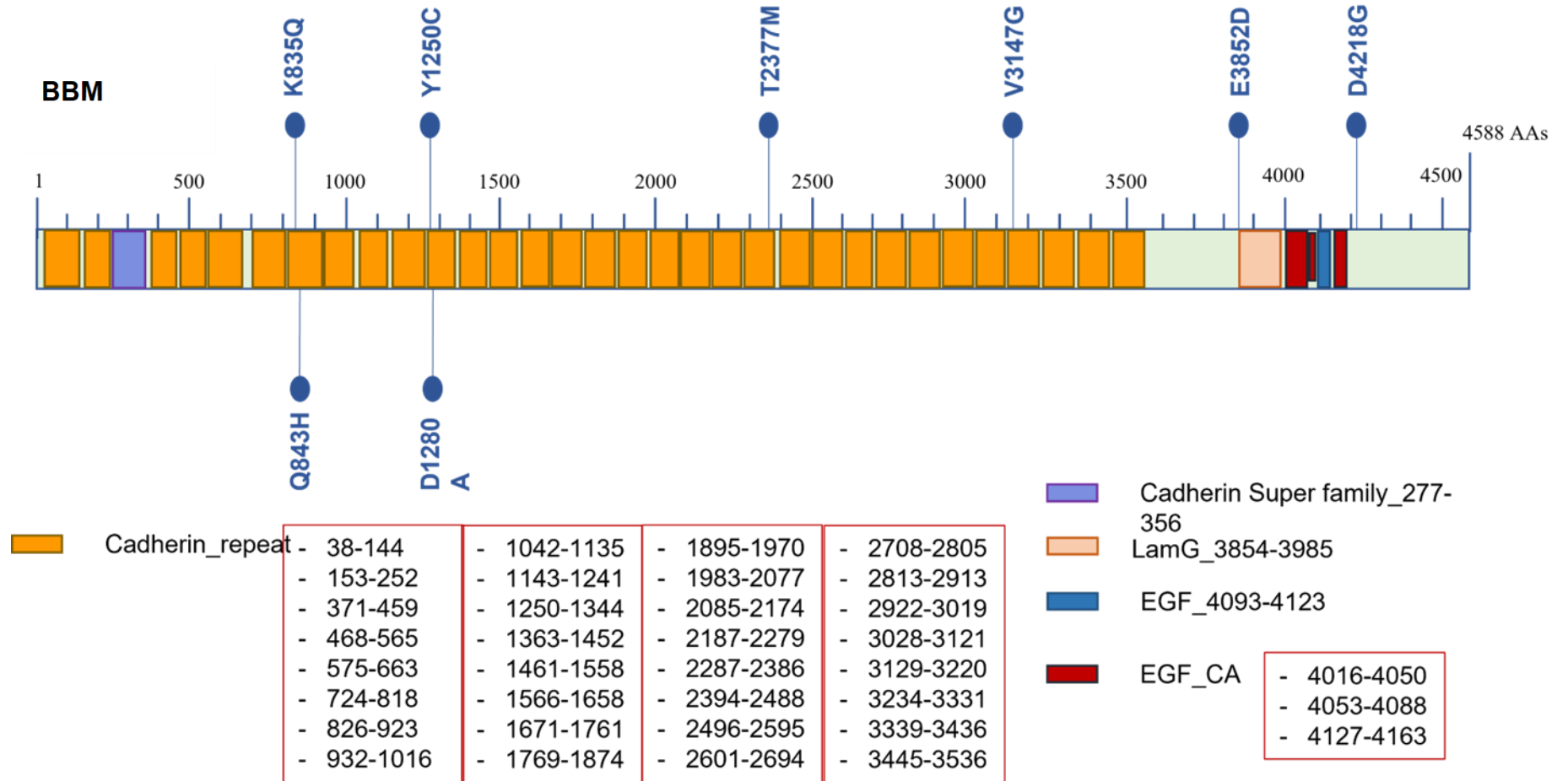
Lung
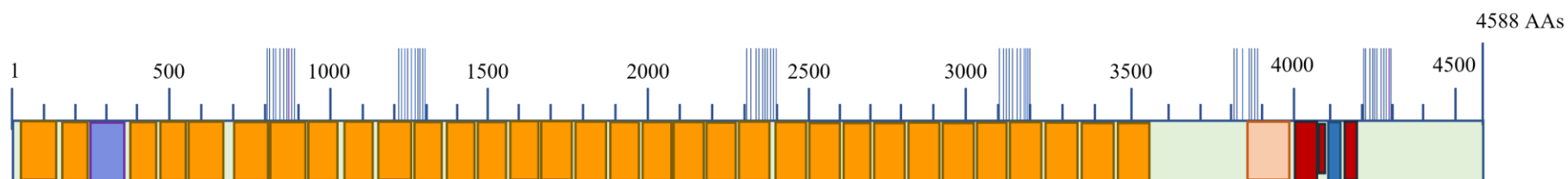
Breast
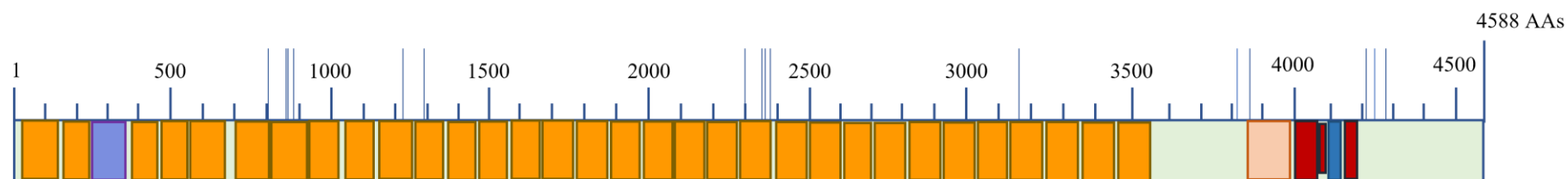
Skin
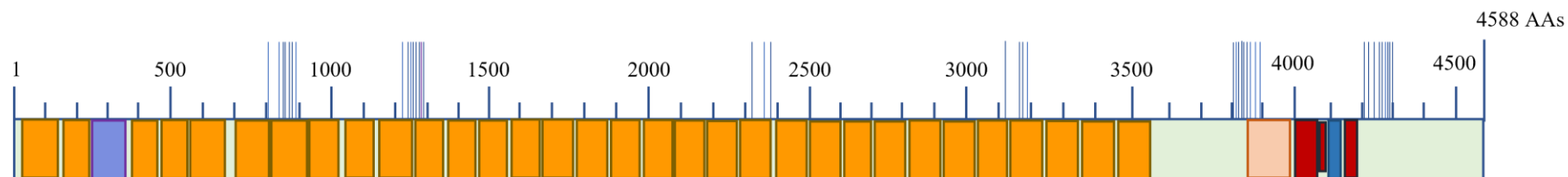
APPENDIX F.5: Schematic representation of FAT1 mutations distribution in within the protein domains. FAT1 is protein is formed by 4588 amino acid residues and five conserved domains. A) Amino acid changes are shown in blue corresponding to missense mutationsB) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.

Appendix F.6: RYR1

Lung

5038 AAs

Breast

5038 AAs

Skin

5038 AAs
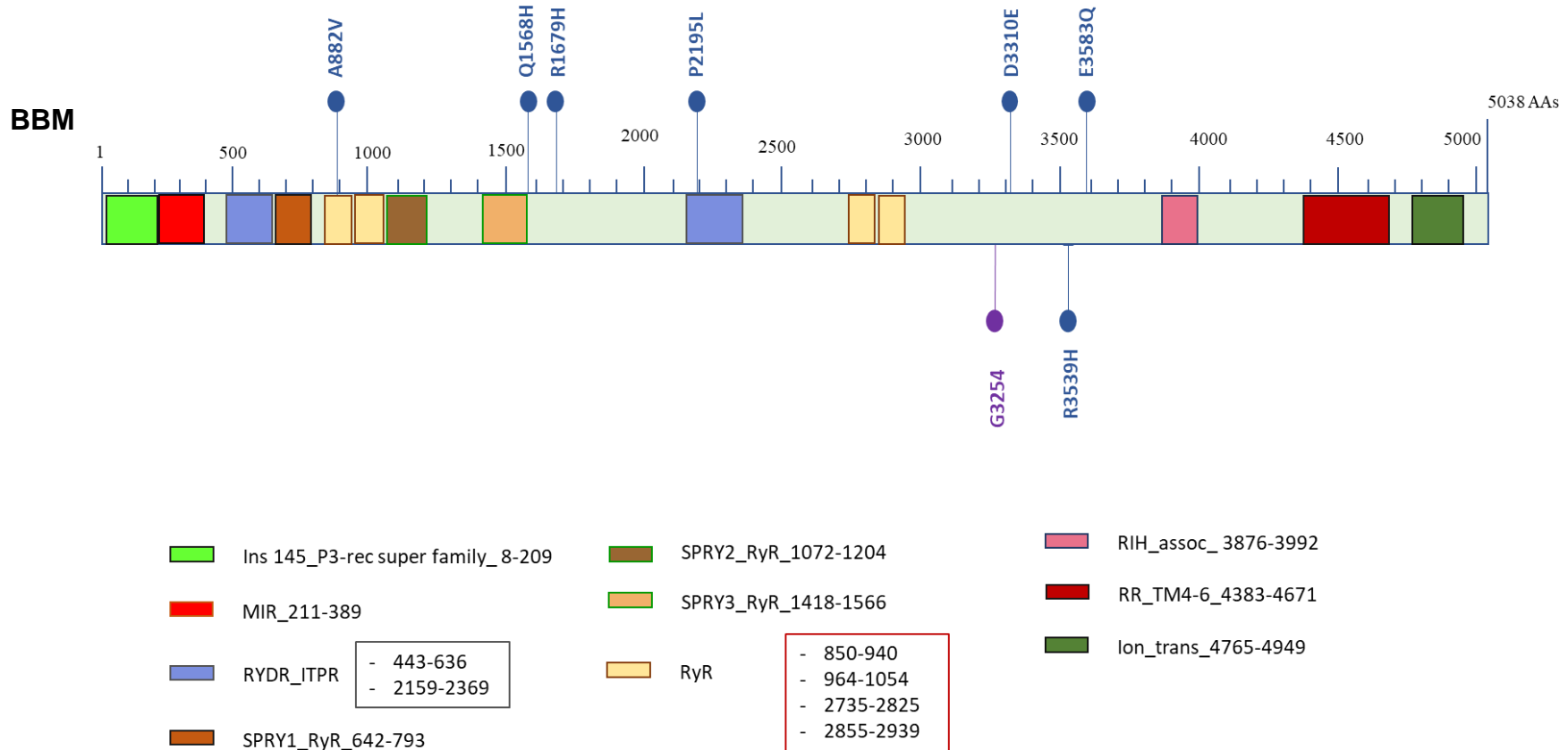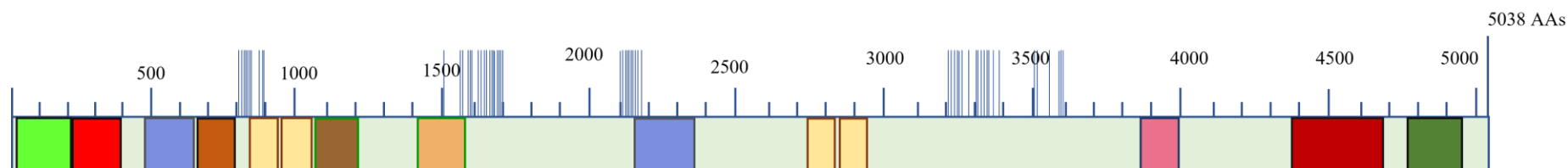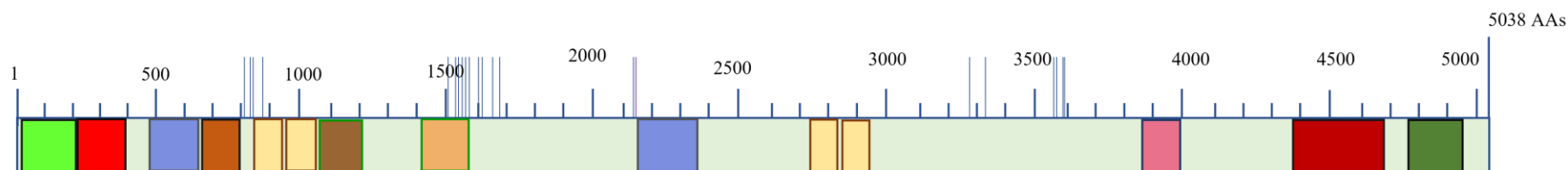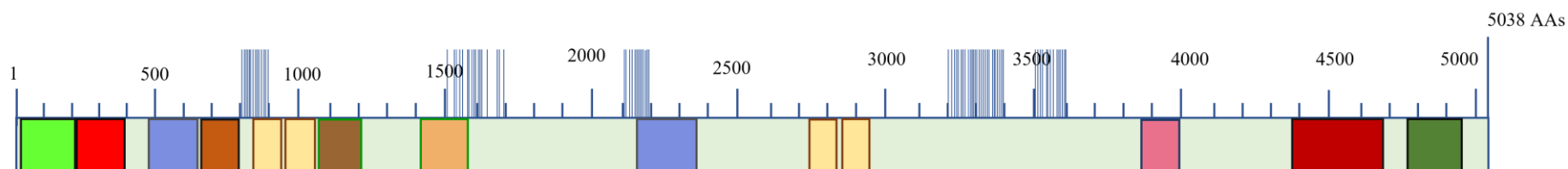
Appendix F.6: Schematic representation of RYR1 mutations distribution in within the protein domains. RYR1 is protein is formed by 5038 amino acid residues and ten conserved domains. A) Amino acid changes are shown in blue and purple, and corresponding to missense and frameshift mutations, respectively. * Amino acid changes predicted to be benign by either SIFT or Polyphen-2. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.

Appendix F.7: RYR3

BBM

S554Y
E3119K
R4376Q

T2508M
R3194H

1   500   1000   1500   2000   2500   3000   3500   4000   4500   4870 AAs

Ins 145_P3-rec super family_ 12-212

MIR_216-397

RYDR_ITPR
- 440-621
- 2021-2233

SPRY1_RyR_641-792

RyR
- 849-937
- 963-1053
- 2598-2668

SPRY2_RyR_1072-1204

SPRY3_RyR_1324-1462

RyR Super family_2716-2800

RIH_assoc_ 3727-3844

EF_hand 7_3928-3983
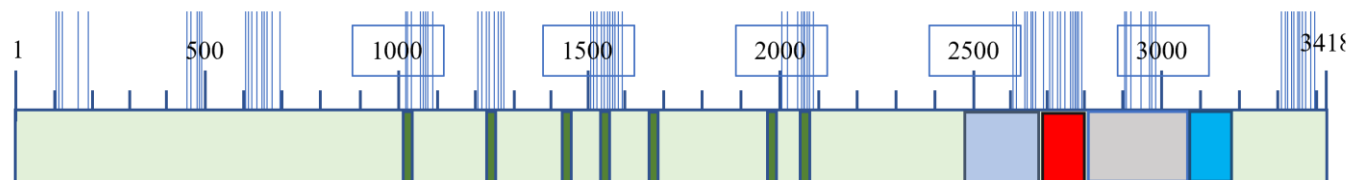
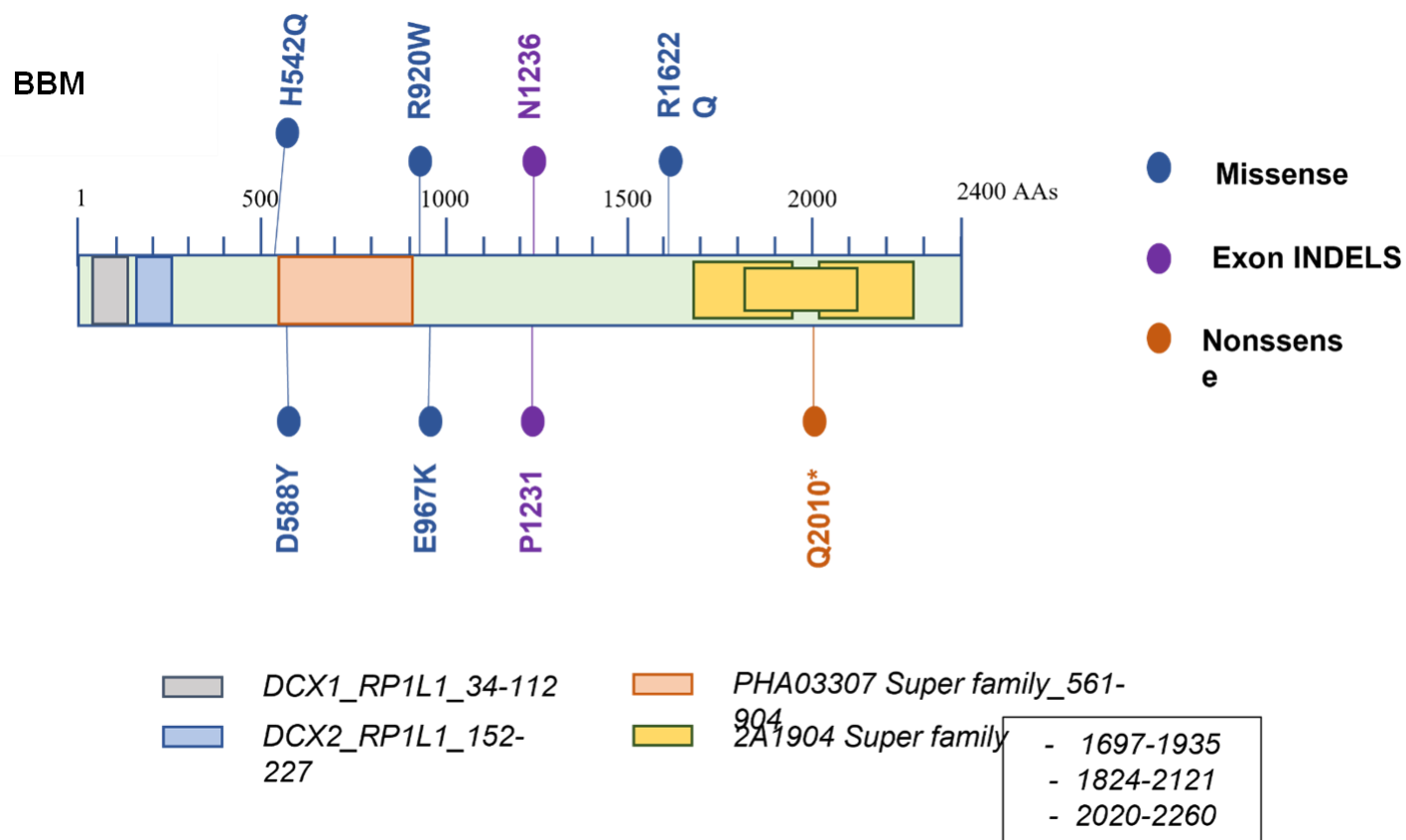RR_TM4-6_4235-4504

Ion_trans_4621-4781
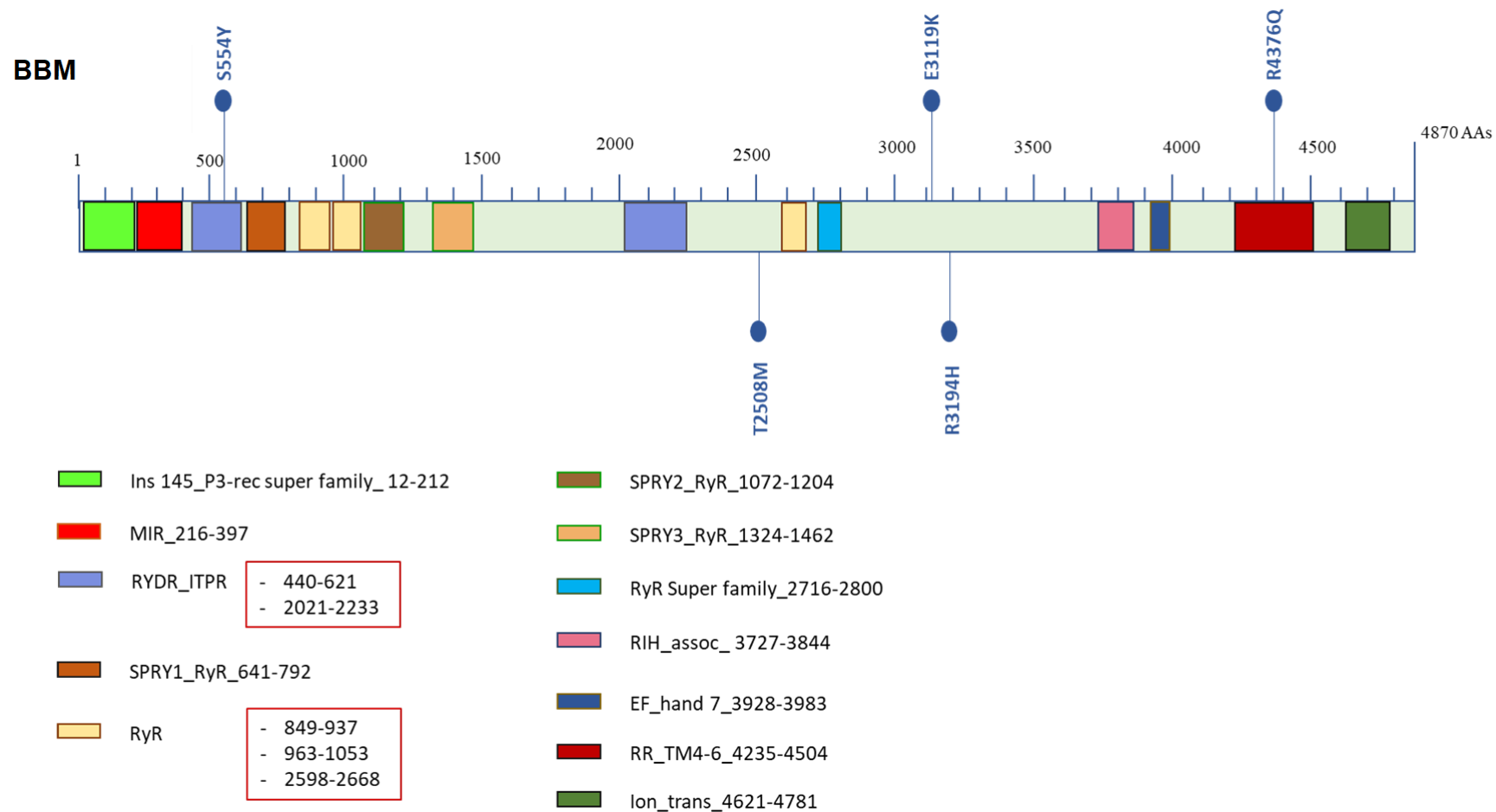
Lung
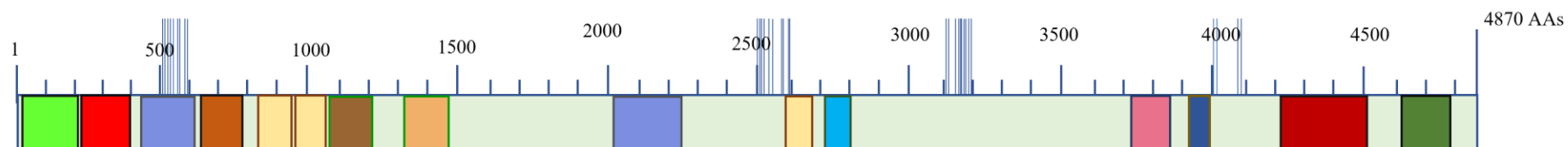
Breast
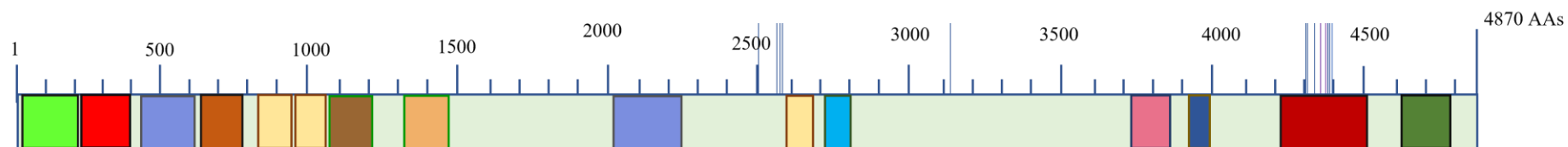
Skin
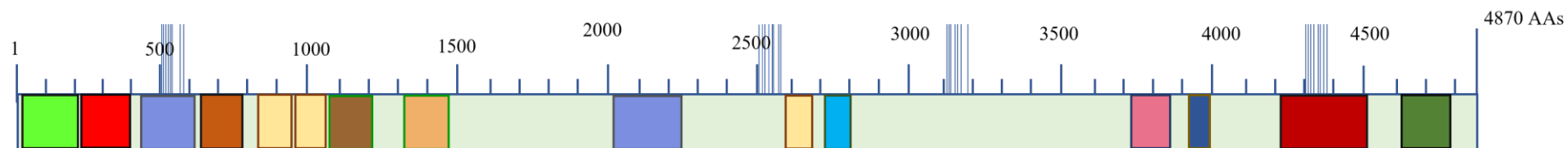
APPENDIX F.7: Schematic representation of RYR3 mutations distribution in within the protein domains. RYR3 is protein is formed by 4870 amino acid residues and twelve conserved domains. A) Amino acid changes are shown in blue, corresponding to missense mutations. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.

Appendix F.8: KMT2D

BBM
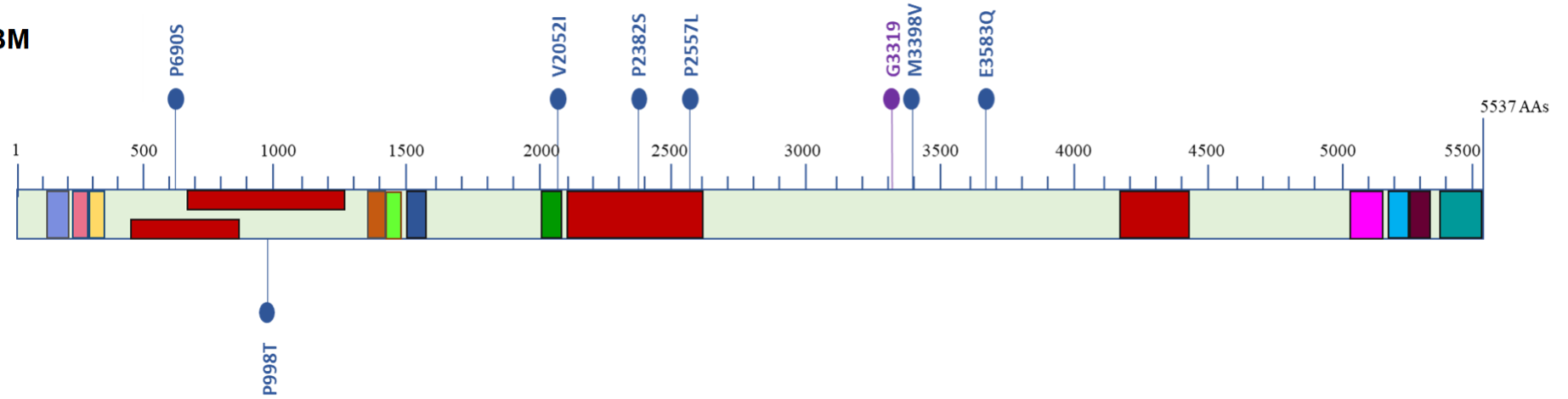
Legend:
- ePHD1_KMT2D_134-217
- PDH1_KMT2C_Like_228-273
- PHD_SF Super family_275-320
- PHA03247 Super family
  - 469-860
  - 690-1272
  - 2108-2600
  - 4162-4426
- PHD3_KMT2D_1378-11428
- PDH5_KMT2C_Like_1429-1475
- PHD5_KMT2D_1506-1556
- HMG_2021-2072
- ePHD2_KMT2D_5032-5138
- FYRN_5181-5232
- FYRC_5240-5327
- SET_KMT2D_5382-5536

Mutations: P690S, P998T, V2052I, P2382S, P2557L, G3319, M3398V, E3583Q

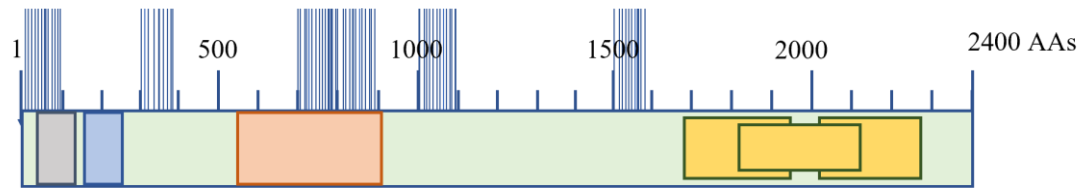5537 AAs

Lung
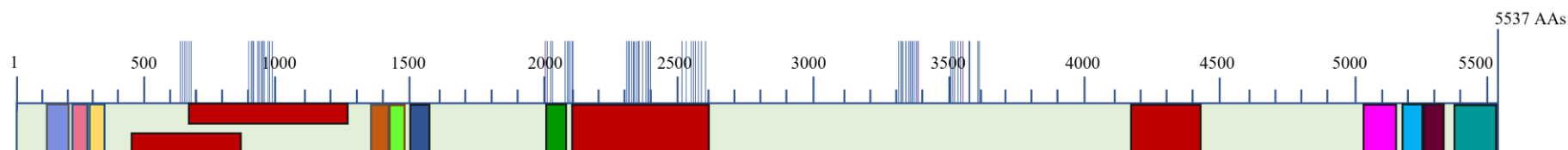
Breast
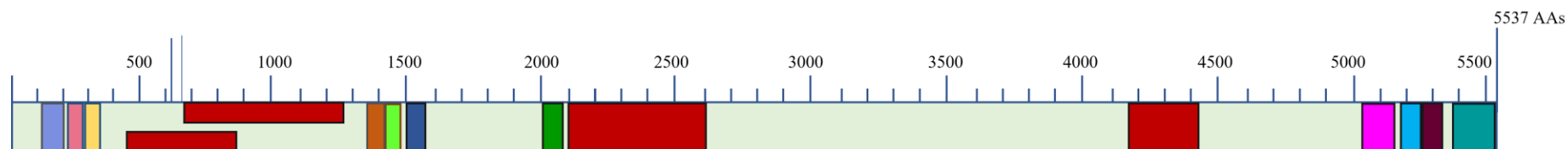
Skin
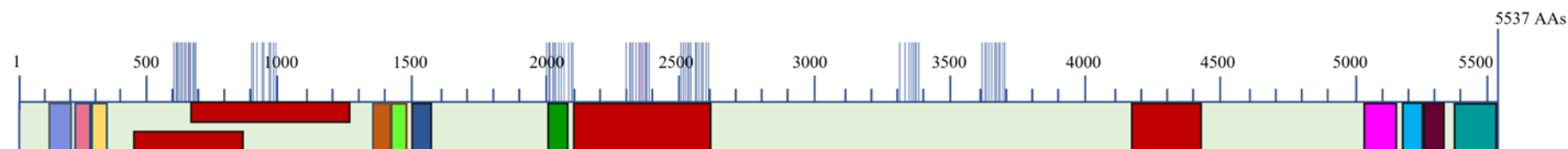
APPENDIX F.8: Schematic representation of KMT2D mutations distribution in within the protein domains. KMT2D is protein is formed by 5537 amino acid residues and twelve conserved domains. A) Amino acid changes are shown in blue and purple, corresponding to missense and frameshift mutations, respectively. B) Mutation distribution within the same area in primary tumour data obtained from CoSMIC database.

Reference list

Abbott, N.J., Rönnbäck, L. and Hansson, E. (2006). Astrocyte–endothelial interactions at the blood–brain barrier. *Nature Reviews Neuroscience*, [online] 7(1), pp.41–53. doi:10.1038/nrn1824.

Abhinand, C.S., Raju, R., Soumya, S.J., Arya, P.S. and Sudhakaran, P.R. (2016). VEGF-A/VEGFR2 signaling network in endothelial cells relevant to angiogenesis. *Journal of Cell Communication and Signaling*, 10(4), pp.347–354. doi:10.1007/s12079-016-0352-8.

Abou Ziki, M.D., Seidelmann, S.B., Smith, E., Atteya, G., Jiang, Y., Fernandes, R.G., Marieb, M.A., Akar, J.G. and Mani, A. (2017). Deleterious protein-altering mutations in the *SCN10A* voltage-gated sodium channel gene are associated with prolonged QT. *Clinical Genetics*, 93(4), pp.741–751. doi:10.1111/cge.13036.

Adamo, B., Deal, A.M., Burrows, E., Geradts, J., Hamilton, E., Blackwell, K.L., Livasy, C., Fritchie, K., Prat, A., Harrell, J.C., Ewend, M.G., Carey, L.A., Miller, C.R. and Anders, C.K. (2011). Phosphatidylinositol 3-kinase pathway activation in breast cancer brain metastases. *Breast Cancer Research*, 13(6). doi:10.1186/bcr3071.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), pp.248–249. doi:10.1038/nmeth0410-248.

Aganezov, S., Yan, S.M., Soto, D.C., Kirsche, M., Zarate, S., Avdeyev, P., Taylor, D.J., Shafin, K., Shumate, A., Xiao, C., Wagner, J., McDaniel, J., Olson, N.D., Sauria, M.E.G., Vollger, M.R., Rhie, A., Meredith, M., Martin, S., Lee, J. and Koren, S. (2022). A complete reference genome improves analysis of human genetic variation. *Science*, 376(6588). doi:10.1126/science.abl3533.

Aguilar-Cazares, D., Chavez-Dominguez, R., Carlos-Reyes, A., Lopez-Camarillo, C., Hernadez de la Cruz, O.N. and Lopez-Gonzalez, J.S. (2019). Contribution of Angiogenesis to Inflammation and Cancer. *Frontiers in Oncology*, 9. doi:10.3389/fonc.2019.01399.

Ahmad, A. (2019). Breast Cancer Statistics: Recent Trends. *Advances in Experimental Medicine and Biology*, 1152, pp.1–7. doi:10.1007/978-3-030-20301-6_1.

Ahmad, A.S., Ormiston-Smith, N. and Sasieni, P.D. (2015). Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960. *British Journal of Cancer*, [online] 112(5), pp.943–947. doi:10.1038/bjc.2014.606.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A.P., Caldas, C., Davies, H.R., Desmedt, C., Eils, R., Eyfjörd, J.E., Foekens, J.A. and Greaves, M. (2013). Signatures of mutational processes in human cancer. *Nature*, [online] 500(7463), pp.415–21. doi:10.1038/nature12477.

Alholle, A., Brini, A.T., Gharanei, S., Vaiyapuri, S., Arrigoni, E., Dallol, A., Gentle, D., Kishida, T., Hiruma, T., Avigad, S., Grimer, R., Maher, E.R. and Latif, F. (2013). Functional epigenetic approach identifies frequently methylated genes in Ewing sarcoma. *Epigenetics*, 8(11), pp.1198–1204. doi:10.4161/epi.26266.

Ali, S., Górska, Z., Duchnowska, R. and Jassem, J. (2021). Molecular Profiles of Brain Metastases: A Focus on Heterogeneity. *Cancers*, 13(11), p.2645. doi:10.3390/cancers13112645.

Allen, J.L. and Mellor, H. (2014). The Coculture Organotypic Assay of Angiogenesis. *Methods in Molecular Biology*, pp.265–270. doi:10.1007/978-1-4939-1462-3_17.

Allgöwer, C., Kretz, A.-L., von Karstedt, S., Wittau, M., Henne-Bruns, D. and Lemke, J. (2020). Friend or Foe: S100 Proteins in Cancer. *Cancers*, 12(8), p.2037. doi:10.3390/cancers12082037.

Altaleb, A., 2021. Metastases: A Visual Guide. In *Surgical Pathology* (pp. 131-136). Springer, Cham.

Álvaro-Espinosa, L., de Pablos-Aragoneses, A., Valiente, M. and Priego, N., 2021. Brain Microenvironment Heterogeneity: Potential Value for Brain Tumors. *Frontiers in Oncology*, *11*, p.714428.

Anaya-Ruiz, M., Bandala, C., Martinez-Morales, P., Landeta, G., Martinez-Contreras, R.D., Martinez-Montiel, N. and Perez-Santos, M. (2018). Emerging Drugs for the Treatment of Breast Cancer Brain Metastasis: A Review of Patent Literature. *Recent Patents on Anti-Cancer Drug Discovery*, 13(3), pp.348–359. doi:10.2174/1574892813666180430113605.

Andersen, A.P., Moreira, J.M.A. and Pedersen, S.F. (2014). Interactions of ion transporters and channels with cancer cell metabolism and the tumour microenvironment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1638), p.20130098. doi:10.1098/rstb.2013.0098.

Andreassen, P.R., Seo, J., Wiek, C. and Hanenberg, H. (2021). Understanding BRCA2 Function as a Tumor Suppressor Based on Domain-Specific Activities in DNA Damage Responses. *Genes*, [online] 12(7), p.1034. doi:10.3390/genes12071034.

Angus, L., Smid, M., Wilting, S.M., van Riet, J., Van Hoeck, A., Nguyen, L., Nik-Zainal, S., Steenbruggen, T.G., Tjan-Heijnen, V.C.G., Labots, M., van Riel, J.M.G.H., Bloemendal, H.J., Steeghs, N., Lolkema, M.P., Voest, E.E., van de Werken, H.J.G., Jager, A., Cuppen, E., Sleijfer, S. and Martens, J.W.M. (2019). The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nature Genetics*, 51(10), pp.1450–1458. doi:10.1038/s41588-019-0507-7.

Anjanappa, M., Hao, Y., Simpson, E.R., Bhat-Nakshatri, P., Nelson, J.B., Tersey, S.A., Mirmira, R.G., Cohen-Gadol, A.A., Saadatzadeh, M.R., Li, L., Fang, F., Nephew, K.P., Miller, K.D., Liu, Y. and Nakshatri, H. (2017). A system for detecting high impact-low frequency mutations in primary tumors and metastases. *Oncogene*, 37(2), pp.185–196. doi:10.1038/onc.2017.322.

Anoosha, P., Sakthivel, R. and Michael Gromiha, M. (2016). Exploring preferred amino acid mutations in cancer genes: Applications to identify potential drug targets. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1862(2), pp.155–165. doi:10.1016/j.bbadis.2015.11.006.

Arafeh, R. and Samuels, Y. (2019). PIK3CA in cancer: The past 30 years. *Seminars in Cancer Biology*, 59, pp.36–49. doi:10.1016/j.semcancer.2019.02.002.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G. and Muzio, L.L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *International Journal of Molecular Medicine*, [online] 40(2), pp.271–280. doi:10.3892/ijmm.2017.3036.

Arvanitis, C.D., Ferraro, G.B. and Jain, R.K. (2019). The blood–brain barrier and blood–tumour barrier in brain tumours and metastases. *Nature Reviews Cancer*. [online] doi:10.1038/s41568-019-0205-x.

Auton, A., Brooks, L., Durbin, R., Garrison, E., Kang, H., Korbel, J., Marchini, J., McCarthy, S., McVean, G. and Abecasis, G. (2015). The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, [online] (526), pp.68–74. doi:10.1038/nature15393.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., Ng, P.K.-S., Jeong, K.J., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E.M., Mularoni, L. and Rubio-Perez, C. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, [online] 173(2), pp.371-385.e18. doi:10.1016/j.cell.2018.02.060.

Bailleux, C., Eberst, L. and Bachelot, T. (2020). Treatment strategies for breast cancer brain metastases. *British Journal of Cancer*, 124(1), pp.142–155. doi:10.1038/s41416-020-01175-y.

Bailleux, C., Eberst, L. and Bachelot, T., 2021. Treatment strategies for breast cancer brain metastases. *British journal of cancer*, *124*(1), pp.142-155.

Bartsch, R., Berghoff, A.S., Furtner, J., Marhold, M., Bergen, E.S., Roider-Schur, S., Starzer, A.M., Forstner, H., Rottenmanner, B., Dieckmann, K. and Bago-Horvath, Z., 2022. Trastuzumab deruxtecan in HER2-positive breast cancer with brain metastases: a single-arm, phase 2 trial. *Nature medicine*, *28*(9), pp.1840-1847.

Baschieri, F., Uetz-von Allmen, E., Legler, D.F. and Farhan, H. (2015). Loss of GM130 in breast cancer cells and its effects on cell migration, invasion and polarity. *Cell Cycle*, 14(8), pp.1139–1147. doi:10.1080/15384101.2015.1007771.

Basu, M., Chakraborty, B., Ghosh, S., Samadder, S., Dutta, S., Roy, A., Pal, D.K., Ghosh, A. and Panda, C.K. (2020). Divergent molecular profile of PIK3CA gene in arsenic-associated bladder carcinoma. *Mutagenesis*, 35(6), pp.499–508. doi:10.1093/mutage/geaa031.

Basu, M.K., Poliakov, E. and Rogozin, I.B. (2009). Domain mobility in proteins: functional and evolutionary implications. *Briefings in Bioinformatics*, [online] 10(3), pp.205–216. doi:10.1093/bib/bbn057.

Batalini, F., Moulder, S.L., Winer, E.P., Rugo, H.S., Lin, N.U. and Wulf, G.M. (2020). Response of Brain Metastases From *PIK3CA*-Mutant Breast Cancer to Alpelisib. *JCO Precision Oncology*, (4), pp.572–578. doi:10.1200/po.19.00403.

Bates, J.P., Derakhshandeh, R., Jones, L. and Webb, T.J. (2018). Mechanisms of immune evasion in breast cancer. *BMC Cancer*, 18(1). doi:10.1186/s12885-018-4441-3.

Beck, T.F., Mullikin, J.C. and Biesecker, L.G. (2016). Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clinical Chemistry*, 62(4), pp.647–654. doi:10.1373/clinchem.2015.249623.

Bertucci, F., Ng, C.K.Y., Patsouris, A., Droin, N., Piscuoglio, S., Carbuccia, N., Soria, J.C., Dien, A.T., Adnani, Y., Kamal, M., Garnier, S., Meurice, G., Jimenez, M., Dogan, S., Verret, B., Chaffanet, M., Bachelot, T., Campone, M., Lefeuvre, C. and Bonnefoi, H. (2019). Genomic characterization of metastatic breast cancers. *Nature*, [online] 569(7757), pp.560–564. doi:10.1038/s41586-019-1056-z.

Bhoumik, A., Takahashi, S., Breitweiser, W., Shiloh, Y., Jones, N. and Ronai, Z. (2005). ATM-Dependent Phosphorylation of ATF2 Is Required for the DNA Damage Response. *Molecular Cell*, 18(5), pp.577–587. doi:10.1016/j.molcel.2005.04.015.

Blazquez, R., Wlochowitz, D., Wolff, A., Seitz, S., Wachter, A., Perera-Bel, J., Bleckmann, A., Beißbarth, T., Salinas, G., Riemenschneider, M.J. and Proescholdt, M., 2018. PI3K: A master regulator of brain metastasis-promoting macrophages/microglia. *Glia, 66*(11), pp.2438-2455.

Blokzijl, F., Janssen, R., van Boxtel, R. and Cuppen, E. (2018). MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Medicine*, [online] 10(1). doi:10.1186/s13073-018-0539-0.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A. and Mi, H. (2020). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, 49(D1), pp.D344–D354. doi:10.1093/nar/gkaa977.

Boire, A., Brastianos, P.K., Garzia, L. and Valiente, M. (2019). Brain metastasis. *Nature Reviews Cancer*, 20(1), pp.4–11. doi:10.1038/s41568-019-0220-y.

Bonam, S.R., Wang, F. and Muller, S. (2019). Lysosomes as a therapeutic target. *Nature Reviews Drug Discovery*, 18(12), pp.923–948. doi:10.1038/s41573-019-0036-1.

Bos, P.D., Zhang, X.H.-F. ., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A. and Massagué, J. (2009). Genes that mediate breast cancer metastasis to the brain. *Nature*, [online] 459(7249), pp.1005–1009. doi:10.1038/nature08021.

Boumahdi, S. and de Sauvage, F.J. (2019). The great escape: tumour cell plasticity in resistance to targeted therapy. *Nature Reviews Drug Discovery*, [online] 19(1), pp.39–56. doi:10.1038/s41573-019-0044-1.

Brastianos, P., Carter, S., Santagata, S., Cahill, D., Taylor-Weiner, A., Jones, R., Ligon, K., Tabernero, J., Seoane, J., Martinez Saez, E., Johnson, B., Choueiri, T., Stemmer-Rachamimov, A., Lin, N., Beroukhim, R., Batchelor, T., Baselga, J., Louis, D., Hahn, W. and Getz, G. (2015). 2905 Genomic characterization of brain

metastases and paired primary tumors reveals branched evolution and potential therapeutic targets. *European Journal of Cancer*, 51, p.S586. doi:10.1016/s0959-8049(16)31622-7.

Brooks, A.J. and Putoczki, T. (2020). JAK-STAT Signalling Pathway in Cancer. *Cancers*, 12(7), p.1971. doi:10.3390/cancers12071971.

Brosnan, E.M. and Anders, C.K. (2018). Understanding patterns of brain metastasis in breast cancer and designing rational therapeutic strategies. *Annals of Translational Medicine*, [online] 6(9), pp.163–163. doi:10.21037/atm.2018.04.35.

Brücher, B.L.D.M. and Jamall, I.S. (2016). Somatic Mutation Theory - Why it's Wrong for Most Cancers. *Cellular Physiology and Biochemistry*, [online] 38(5), pp.1663–1680. doi:10.1159/000443106.

Brzozowski, J.S. and Skelding, K.A., 2019. The multi-functional calcium/calmodulin stimulated protein kinase (CaMK) family: emerging targets for anti-cancer therapeutic intervention. *Pharmaceuticals*, *12*(1), p.8.

Bus, P., Gerrits, T., Heemskerk, S.A.C., Zandbergen, M., Wolterbeek, R., Bruijn, J.A., Baelde, H.J. and Scharpfenecker, M. (2018). Endoglin Mediates Vascular Endothelial Growth Factor-A–Induced Endothelial Cell Activation by Regulating Akt Signaling. *The American Journal of Pathology*, 188(12), pp.2924–2935. doi:10.1016/j.ajpath.2018.08.005.

Cacho-Díaz, B., García-Botello, D.R., Wegman-Ostrosky, T., Reyes-Soto, G., Ortiz-Sánchez, E. and Herrera-Montalvo, L.A., 2020. Tumor microenvironment differences between primary tumor and brain metastases. *Journal of Translational Medicine*, *18*(1), pp.1-12.

Caiado, F., Silva-Santos, B. and Norell, H. (2016). Intra-tumour heterogeneity - going beyond genetics. *The FEBS Journal*, 283(12), pp.2245–2258. doi:10.1111/febs.13705.

Calmon, M.F., Jeschke, J., Zhang, W., Dhir, M., Siebenkäs, C., Herrera, A., Tsai, H.-C., O'Hagan, H.M., Pappou, E.P., Hooker, C.M., Fu, T., Schuebel, K.E., Gabrielson, E., Rahal, P., Herman, J.G., Baylin, S.B. and Ahuja, N. (2015).

Epigenetic silencing of neurofilament genes promotes an aggressive phenotype in breast cancer. *Epigenetics*, 10(7), pp.622–632. doi:10.1080/15592294.2015.1050173.

cancer.sanger.ac.uk. (n.d.). *Cancer Browser*. [online] Available at: https://cancer.sanger.ac.uk/cosmic/browse/tissue?hn=carcinoma&in=t&sn=breast&ss=all [Accessed 19 Jan. 2020].

Cao, S., Zhou, D.C., Oh, C., Jayasinghe, R.G., Zhao, Y., Yoon, C.J., Wyczalkowski, M.A., Bailey, M.H., Tsou, T., Gao, Q., Malone, A., Reynolds, S., Shmulevich, I., Wendl, M.C., Chen, F. and Ding, L. (2020). Discovery of driver non-coding splice-site-creating mutations in cancer. *Nature Communications*, 11(1). doi:10.1038/s41467-020-19307-6.

Cardenas, A.E., Anderson, C.M., Elber, R. and Webb, L.J. (2019). Partition of Positively and Negatively Charged Tryptophan Ions in Membranes with Inverted Phospholipid Heads: Simulations and Experiments. *The Journal of Physical Chemistry B*, 123(15), pp.3272–3281. doi:10.1021/acs.jpcb.9b00754.

Caterino, M., Ruoppolo, M., Mandola, A., Costanzo, M., Orrù, S. and Imperlini, E. (2017). Protein–protein interaction networks as a new perspective to evaluate distinct functional roles of voltage-dependent anion channel isoforms. *Molecular BioSystems*, 13(12), pp.2466–2476. doi:10.1039/c7mb00434f.

Celià-Terrassa, T., Liu, D.D., Choudhury, A., Hang, X., Wei, Y., Zamalloa, J., Alfaro-Aco, R., Chakrabarti, R., Jiang, Y.-Z., Koh, B.I., Smith, H.A., DeCoste, C., Li, J.-J., Shao, Z.-M. and Kang, Y. (2017). Normal and cancerous mammary stem cells evade interferon-induced constraint through the miR-199a–LCOR axis. *Nature Cell Biology*, 19(6), pp.711–723. doi:10.1038/ncb3533.

Chaffer, C.L. and Weinberg, R.A. (2011). A Perspective on Cancer Cell Metastasis. *Science*, 331(6024), pp.1559–1564. doi:10.1126/science.1203543.

Chaffer, C.L., Thompson, E.W. and Williams, E.D. (2007). Mesenchymal to Epithelial Transition in Development and Disease. *Cells Tissues Organs*, 185(1-3), pp.7–19. doi:10.1159/000101298.

Chandler, M.R., Bilgili, E.P. and Merner, N.D. (2016). A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. *Human Mutation*, 37(9), pp.835–846. doi:10.1002/humu.23017.

Chang, Y.-S., Chang, C.-M., Lin, C.-Y., Chao, D.-S., Huang, H.-Y. and Chang, J.-G. (2020). Pathway Mutations in Breast Cancer Using Whole-Exome Sequencing. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics*, 28(2), pp.107–116. doi:10.3727/096504019x15698362825407.

Chang, Y.-S., Huang, H.-D., Yeh, K.-T. and Chang, J.-G. (2017). Identification of novel mutations in endometrial cancer patients by whole-exome sequencing. *International Journal of Oncology*, 50(5), pp.1778–1784. doi:10.3892/ijo.2017.3919.

Chen, C., Köberle, B., Kaufmann, A.M. and Albers, A.E. (2010). A Quest for Initiating Cells of Head and Neck Cancer and Their Treatment. *Cancers*, 2(3), pp.1528–1554. doi:10.3390/cancers2031528.

Chen, C.-L., Hsu, S.-C., Ann, D.K., Yen, Y. and Kung, H.-J. (2021). Arginine Signaling and Cancer Metabolism. *Cancers*, 13(14), p.3541. doi:10.3390/cancers13143541.

Cheng, H., Schaeffer, R.D., Liao, Y., Kinch, L.N., Pei, J., Shi, S., Kim, B.-H. and Grishin, N.V. (2014). ECOD: An Evolutionary Classification of Protein Domains. *PLoS Computational Biology*, 10(12), p.e1003926. doi:10.1371/journal.pcbi.1003926.

Chen, W., Hoffmann, A.D., Liu, H. and Liu, X., 2018. Organotropism: new insights into molecular mechanisms of breast cancer metastasis. *NPJ precision oncology*, *2*(1), pp.1-12.

Chen, Z., Wang, J., Cai, L., Zhong, B., Luo, H., Hao, Y., Yu, W., Wang, B., Su, C., Lei, Y., Bella, A.E., Xiang, A.P. and Wang, T. (2014). Role of the Stem Cell-Associated Intermediate Filament Nestin in Malignant Proliferation of Non-Small Cell Lung Cancer. *PLoS ONE*, 9(2), p.e85584. doi:10.1371/journal.pone.0085584.

Chigira, T., Nagatoishi, S., Takeda, H., Yoshimaru, T., Katagiri, T. and Tsumoto, K. (2019). Biophysical characterization of the breast cancer-related BIG3-PHB2 interaction: Effect of non-conserved loop region of BIG3 on the structure and the interaction. *Biochemical and Biophysical Research Communications*, 518(1), pp.183–189. doi:10.1016/j.bbrc.2019.08.028.

Chin, A.R. and Wang, S.E. (2016). Cancer Tills the Premetastatic Field: Mechanistic Basis and Clinical Implications. *Clinical Cancer Research*, 22(15), pp.3725–3733. doi:10.1158/1078-0432.ccr-16-0028.

Cho, E.S., Kang, H.E., Kim, N.H. and Yook, J.I. (2019). Therapeutic implications of cancer epithelial-mesenchymal transition (EMT). *Archives of Pharmacal Research*, 42(1), pp.14–24. doi:10.1007/s12272-018-01108-7.

Choong, G.M., Cullen, G.D. and O'Sullivan, C.C. (2020). Evolving standards of care and new challenges in the management of HER2-positive breast cancer. *CA: A Cancer Journal for Clinicians*, 70(5), pp.355–374. doi:10.3322/caac.21634.

Chung, W., Eum, H.H., Lee, H.-O., Lee, K.-M., Lee, H.-B., Kim, K.-T., Ryu, H.S., Kim, S., Lee, J.E., Park, Y.H., Kan, Z., Han, W. and Park, W.-Y. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nature Communications*, [online] 8(1). doi:10.1038/ncomms15081.

Ciriello, G., Gatza, M.L., Beck, A.H. and Wilkerson, M.D. (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Science Direct*, [online] 163(2), pp.506–519. doi:https://doi.org/10.1016/j.cell.2015.09.033.

Claxton, S. and Fruttiger, M. (2004). Periodic Delta-like 4 expression in developing retinal arteries. *Gene Expression Patterns*, 5(1), pp.123–127. doi:10.1016/j.modgep.2004.05.004.

Corti, C., Antonarelli, G., Criscitiello, C., Lin, N.U., Carey, L.A., Cortés, J., Poortmans, P. and Curigliano, G. (2022). Targeting brain metastases in breast cancer. *Cancer Treatment Reviews*, 103, p.102324. doi:10.1016/j.ctrv.2021.102324.

Cosgrove, N., Varešlija, D., Keelan, S., Elangovan, A., Atkinson, J.M., Cocchiglia, S., Bane, F.T., Singh, V., Furney, S., Hu, C., Carter, J.M., Hart, S.N., Yadav, S., Goetz, M.P., Hill, A.D.K., Oesterreich, S., Lee, A.V., Couch, F.J. and Young, L.S. (2022). Mapping molecular subtype specific alterations in breast cancer brain metastases identifies clinically relevant vulnerabilities. *Nature Communications*, 13(1). doi:10.1038/s41467-022-27987-5.

Cotto, K., Danos, A., Lesurf, R., Park, M., Griffith, M. and Griffith, O. (2019). 24. Identification of recurrent, non-coding mutations across breast cancer molecular subtypes. *Cancer Genetics*, 233-234, p.S10. doi:10.1016/j.cancergen.2019.04.030.

Craig, A.J., von Felden, J., Garcia-Lezana, T., Sarcognato, S. and Villanueva, A. (2020). Tumour evolution in hepatocellular carcinoma. *Nature Reviews Gastroenterology & Hepatology*, [online] 17(3), pp.139–152. doi:10.1038/s41575-019-0229-4.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerød, A., Green, A., Provenzano, E. and Wishart, G. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403), pp.346–352. doi:10.1038/nature10983.

Dai, Z., Ramesh, V. and Locasale, J.W. (2020). The evolving metabolic landscape of chromatin biology and epigenetics. *Nature Reviews Genetics*, 21(12), pp.737–753. doi:10.1038/s41576-020-0270-8.

Damaghi, M. and Gillies, R. (2017). Phenotypic changes of acid-adapted cancer cells push them toward aggressiveness in their evolution in the tumor microenvironment. *Cell Cycle*, 16(19), pp.1739–1743. doi:10.1080/15384101.2016.1231284.

Damaghi, M., West, J., Robertson-Tessi, M., Xu, L., Ferrall-Fairbanks, M.C., Stewart, P.A., Persi, E., Fridley, B.L., Altrock, P.M., Gatenby, R.A., Sims, P.A., Anderson, A.R.A. and Gillies, R.J. (2021). The harsh microenvironment in early breast

cancer selects for a Warburg phenotype. *Proceedings of the National Academy of Sciences*, 118(3), p.e2011342118. doi:10.1073/pnas.2011342118.

Darwiche, N. (2020). Epigenetic mechanisms and the hallmarks of cancer: an intimate affair. *American Journal of Cancer Research*, [online] 10(7), pp.1954–1978. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7407342/#b9 [Accessed 8 Nov. 2022].

Das, R. and Ghosh, S.K. (2017). Genetic variants of the DNA repair genes from Exome Aggregation Consortium (EXAC) database: significance in cancer. *DNA Repair*, 52, pp.92–102. doi:10.1016/j.dnarep.2017.02.013.

Dayem Ullah, A.Z., Oscanoa, J., Wang, J., Nagano, A., Lemoine, N.R. and Chelala, C. (2018). SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Research*, 46(W1), pp.W109–W113. doi:10.1093/nar/gky399.

De Cario, R., Kura, A., Suraci, S., Magi, A., Volta, A., Marcucci, R., Gori, A.M., Pepe, G., Giusti, B. and Sticchi, E. (2020). Sanger Validation of High-Throughput Sequencing in Genetic Diagnosis: Still the Best Practice? *Frontiers in Genetics*, 11. doi:10.3389/fgene.2020.592588.

De Jonghe, P., Mersivanova, I., Nelis, E., Del Favero, J., Martin, J.-J., Van Broeckhoven, C., Evgrafov, O. and Timmerman, V. (2001). Further evidence that neurofilament light chain gene mutations can cause Charcot-Marie-Tooth disease type 2E. *Annals of Neurology*, 49(2), pp.245–249. doi:3.0.co;2-a">10.1002/1531-8249(20010201)49:2<245::aid-ana45>3.0.co;2-a.

De Mattos-Arruda, L., Ng, C.K., Piscuoglio, S., Gonzalez-Cao, M., Lim, R.S., De Filippo, M.R., Fusco, N., Schultheis, A.M., Ortiz, C., Viteri, S. and Arias, A., 2018. Genetic heterogeneity and actionable mutations in HER2-positive primary breast cancers and their brain metastases. *Oncotarget*, *9*(29), p.20617.Vancouver

De Palma, M., Biziato, D. and Petrova, T.V. (2017). Microenvironmental regulation of tumour angiogenesis. *Nature Reviews Cancer*, 17(8), pp.457–474. doi:10.1038/nrc.2017.51.

Dhanasekaran, K., Kumari, S. and Kanduri, C. (2012). Noncoding RNAs in Chromatin Organization and Transcription Regulation: An Epigenetic View. *Subcellular Biochemistry*, pp.343–372. doi:10.1007/978-94-007-4525-4_15.

Dietlein, F., Weghorn, D., Taylor-Weiner, A., Richters, A., Reardon, B., Liu, D., Lander, E.S., Van Allen, E.M. and Sunyaev, S.R. (2020). Identification of cancer driver genes based on nucleotide context. *Nature Genetics*, 52(2), pp.208–218. doi:10.1038/s41588-019-0572-y.

Di Resta, C., Galbiati, S., Carrera, P.F. and Ferrari, M. (2018). Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities. *EJIFCC*, [online] 29(1), pp.4–14. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5949614/.

Dirican, E., Akkiprik, M. and Özer, A. (2016). Mutation distributions and clinical correlations of PIK3CA gene mutations in breast cancer. *Tumor Biology*, 37(6), pp.7033–7045. doi:10.1007/s13277-016-4924-2.

Li, S., Chen, X., Mao, L., Zahid, K.R., Wen, J., Zhang, L., Zhang, M., Duan, J., Duan, J., Yin, X. and Wang, Y., 2018. Histone deacetylase 1 promotes glioblastoma cell proliferation and invasion via activation of PI3K/AKT and MEK/ERK signaling pathways. *Brain research*, *1692*, pp.154-162.

Nicolas, C.S., Amici, M., Bortolotto, Z.A., Doherty, A., Csaba, Z., Fafouri, A., Dournaud, P., Gressens, P., Collingridge, G.L. and Peineau, S., 2013. The role of JAK-STAT signaling within the CNS. *Jak-stat, 2*(1), p.e22925.

Dolznig, H., Rupp, C., Puri, C., Haslinger, C., Schweifer, N., Wieser, E., Kerjaschki, D. and Garin-Chesa, P. (2011). Modeling Colon Adenocarcinomas in Vitro. *The American Journal of Pathology*, 179(1), pp.487–501. doi:10.1016/j.ajpath.2011.03.015.

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. and Liu, X. (2014). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular Genetics*, [online] 24(8), pp.2125–2137. doi:10.1093/hmg/ddu733.

Dongre, A. and Weinberg, R.A. (2018). New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nature Reviews Molecular Cell Biology*, 20(2), pp.69–84. doi:10.1038/s41580-018-0080-4.

Dong, S., Li, R., Cheng, Z., Liu, D., Xia, J., Xu, J., Li, S., Wang, J., Yue, Y., Fan, Y., Cao, Y., Dai, L., Wang, J., Zhao, P., Wang, X., Xiao, Z., Qiu, C., Wang, G. and Zou, C. (2021). Mutational Pattern in Multiple Pulmonary Nodules Are Associated With Early Stage Lung Adenocarcinoma. *Frontiers in Oncology*, 10. doi:10.3389/fonc.2020.571521.

Drapela, S. and Gomes, A.P. (2021). Metabolic requirements of the metastatic cascade. *Current Opinion in Systems Biology*, 28, p.100381. doi:10.1016/j.coisb.2021.100381.

Dravis, C., Chung, C.-Y., Lytle, N.K., Herrera-Valdez, J., Luna, G., Trejo, C.L., Reya, T. and Wahl, G.M. (2018). Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. *SSRN Electronic Journal*. doi:10.2139/ssrn.3155661.

Dresang, L.R., Van Scoyk, C.A., Kuehn, K.J., Tauber, T.A., Tondin, A.R., Broske, M.A. and Schreiner, C.J. (2022). Comparing transcriptomic profiles from seven cell lines to elucidate liver metastatic potential. *Advances in Cancer Biology - Metastasis*, 4, p.100018. doi:10.1016/j.adcanc.2021.100018.

Drubin, D.G. and Nelson, W.James. (1996). Origins of Cell Polarity. *Cell*, [online] 84(3), pp.335–344. doi:10.1016/s0092-8674(00)81278-7.

Duchnowska, R., Jassem, J., Goswami, C.P., Dundar, M., Gökmen-Polar, Y., Li, L., Woditschka, S., Biernat, W., Sosińska-Mielcarek, K., Czartoryska-Arłukowicz, B., Radecka, B., Tomasevic, Z., Stępniak, P., Wojdan, K., Sledge, G.W., Steeg, P.S. and Badve, S. (2015). Predicting early brain metastases based on clinicopathological factors and gene expression analysis in advanced HER2-positive breast cancer patients. *Journal of Neuro-Oncology*, 122(1), pp.205–216. doi:10.1007/s11060-014-1704-y.

Duijf, P.H.G., Nanayakkara, D., Nones, K., Srihari, S., Kalimutho, M. and Khanna, K.K. (2019). Mechanisms of Genomic Instability in Breast Cancer. *Trends in Molecular Medicine*, [online] 25(7), pp.595–611. doi:10.1016/j.molmed.2019.04.004.

Dujon, A.M., Capp, J.-P., Brown, J.S., Pujol, P., Gatenby, R.A., Ujvari, B., Alix-Panabières, C. and Thomas, F. (2021). Is There One Key Step in the Metastatic Cascade? *Cancers*, 13(15), p.3693. doi:10.3390/cancers13153693.

Du, S., Sung, Y.-S., Wey, M., Wang, Y., Alatrash, N., Berthod, A., MacDonnell, F.M. and Armstrong, D.W. (2020). Roles of N-methyl-d-aspartate receptors and d-amino acids in cancer cell viability. *Molecular Biology Reports*, 47(9), pp.6749–6758. doi:10.1007/s11033-020-05733-8.

Eelen, G., Treps, L., Li, X. and Carmeliet, P. (2020). Basic and Therapeutic Aspects of Angiogenesis Updated. *Circulation Research*, 127(2), pp.310–329. doi:10.1161/circresaha.120.316851.

Engin, H.B., Guney, E., Keskin, O., Oliva, B. and Gursoy, A. (2013). Integrating Structure to Protein-Protein Interaction Networks That Drive Metastasis to Brain and Lung in Breast Cancer. *PLoS ONE*, 8(11), p.e81035. doi:10.1371/journal.pone.0081035.

Ensembl.org. (2014). *Pathogenicity predictions*. [online] Available at: https://www.ensembl.org/info/genome/variation/prediction/protein_function.html.

Ensembl.org. (2020). *Ensembl*. [online] Available at: https://www.ensembl.org/.

Ernst, C., Hahnen, E., Engel, C., Nothnagel, M., Weber, J., Schmutzler, R.K. and Hauke, J. (2018). Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Medical Genomics*, 11(1). doi:10.1186/s12920-018-0353-y.

Fagan, R.J. and Dingwall, A.K. (2019). COMPASS Ascending: Emerging clues regarding the roles of MLL3/KMT2C and MLL2/KMT2D proteins in cancer. *Cancer Letters*, 458, pp.56–65. doi:10.1016/j.canlet.2019.05.024.

Falzone, L., Salomone, S. and Libra, M. (2018). Evolution of Cancer Pharmacological Treatments at the Turn of the Third Millennium. *Frontiers in Pharmacology*, [online] 9. doi:10.3389/fphar.2018.01300.

Fearnley, G.W., Abdul-Zani, I., Latham, A.M., Hollstein, M.C., Ladbury, J.E., Wheatcroft, S.B., Odell, A.F. and Ponnambalam, S. (2019). Tpl2 is required for VEGF-A-stimulated signal transduction and endothelial cell function. *Biology Open*. doi:10.1242/bio.034215.

Fearnley, G.W., Odell, A.F., Latham, A.M., Mughal, N.A., Bruns, A.F., Burgoyne, N.J., Homer-Vanniasinkam, S., Zachary, I.C., Hollstein, M.C., Wheatcroft, S.B. and Ponnambalam, S. (2014). VEGF-A isoforms differentially regulate ATF-2–dependent VCAM-1 gene expression and endothelial–leukocyte interactions. *Molecular Biology of the Cell*, 25(16), pp.2509–2521. doi:10.1091/mbc.e14-05-0962.

Fecci, P.E., Champion, C.D., Hoj, J., McKernan, C.M., Goodwin, C.R., Kirkpatrick, J.P., Anders, C.K., Pendergast, A.M. and Sampson, J.H. (2019). The Evolving Modern Management of Brain Metastasis. *Clinical Cancer Research*, 25(22), pp.6570–6580. doi:10.1158/1078-0432.ccr-18-1624.

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D.M., Piñeros, M., Znaor, A. and Bray, F. (2021). Cancer statistics for the year 2020: an overview. *International Journal of Cancer*, 149(4). doi:10.1002/ijc.33588.

Fernández-Rozadilla, C., Álvarez-Barona, M., Quintana, I., López-Novo, A., Amigo, J., Cameselle-Teijeiro, J.M., Roman, E., Gonzalez, D., Llor, X., Bujanda, L., Bessa, X., Jover, R., Balaguer, F., Castells, A., Castellví-Bel, S., Capellá, G., Carracedo, A., Valle, L. and Ruiz-Ponte, C. (2021). Exome sequencing of early-onset patients supports genetic heterogeneity in colorectal cancer. *Scientific Reports*, 11(1). doi:10.1038/s41598-021-90590-z.

Ferreira, H.J. and Esteller, M. (2018). Non-coding RNAs, epigenetics, and cancer: tying it all together. *Cancer and Metastasis Reviews*, 37(1), pp.55–73. doi:10.1007/s10555-017-9715-8.

Ferronika, P., Hof, J., Kats-Ugurlu, G., Sijmons, R.H., Terpstra, M.M., de Lange, K., Leliveld-Kors, A., Westers, H. and Kok, K. (2019). Comprehensive Profiling of Primary and Metastatic ccRCC Reveals a High Homology of the Metastases to

a Subregion of the Primary Tumour. *Cancers*, 11(6), p.812. doi:10.3390/cancers11060812.

Fitzgerald, D.M., Muzikansky, A., Pinto, C., Henderson, L., Walmsley, C., Allen, R., Ferraro, G.B., Isakoff, S., Moy, B., Oh, K., Shih, H.A., Dias-Santagata, D., Iafrate, A.J., Bardia, A., Brastianos, P.K. and Juric, D. (2019). Association between PIK3CA mutation status and development of brain metastases in HR+/HER2- metastatic breast cancer. *Annals of Oncology*, 30, p.v110. doi:10.1093/annonc/mdz242.013.

Flanagan, S.E., Patch, A.-M. and Ellard, S. (2010). Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations. *Genetic Testing and Molecular Biomarkers*, 14(4), pp.533–537. doi:10.1089/gtmb.2010.0036.

Flavahan, W.A., Gaskell, E. and Bernstein, B.E. (2017). Epigenetic plasticity and the hallmarks of cancer. *Science*, 357(6348), p.eaal2380. doi:10.1126/science.aal2380.

Frisk, G., Svensson, T., Bäcklund, L.M., Lidbrink, E., Blomqvist, P. and Smedby, K.E. (2012). Incidence and time trends of brain metastases admissions among breast cancer patients in Sweden. *British Journal of Cancer*, [online] 106(11), pp.1850–1853. doi:10.1038/bjc.2012.163.

Frisk, G., Tinge, B., Ekberg, S., Eloranta, S., Bäcklund, L.M., Lidbrink, E. and Smedby, K.E. (2017). Survival and level of care among breast cancer patients with brain metastases treated with whole brain radiotherapy. *Breast Cancer Research and Treatment*, 166(3), pp.887–896. doi:10.1007/s10549-017-4466-3.

Froeling, F.E.M., Marshall, J.F. and Kocher, H.M. (2010). Pancreatic cancer organotypic cultures. *Journal of Biotechnology*, 148(1), pp.16–23. doi:10.1016/j.jbiotec.2010.01.008.

FUJII, T., TOKUDA, S., NAKAZAWA, Y., KUROZUMI, S., OBAYASHI, S., YAJIMA, R. and SHIRABE, K. (2020). Eribulin Suppresses New Metastases in Patients With Metastatic Breast Cancer. *In Vivo*, 34(2), pp.917–921. doi:10.21873/invivo.11858.

Fumagalli, C., Ranghiero, A., Gandini, S., Corso, F., Taormina, S., De Camilli, E., Rappa, A., Vacirca, D., Viale, G., Guerini-Rocco, E. and Barberis, M. (2020). Inter-tumor genomic heterogeneity of breast cancers: comprehensive genomic profile of primary early breast cancers and relapses. *Breast Cancer Research*, 22(1). doi:10.1186/s13058-020-01345-z.

Fusco, N., Malapelle, U., Fassan, M., Marchiò, C., Buglioni, S., Zupo, S., Criscitiello, C., Vigneri, P., Dei Tos, A.P., Maiorano, E. and Viale, G., (2021). PIK3CA mutations as a molecular target for hormone receptor-positive, HER2-negative metastatic breast cancer. *Frontiers in Oncology*, *11*, p.644737.

Gaggioli, C., Hooper, S., Hidalgo-Carcedo, C., Grosse, R., Marshall, J.F., Harrington, K. and Sahai, E. (2007). Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. *Nature Cell Biology,* [online] 9(12), pp.1392–1400. doi:10.1038/ncb1658.

Gallardo, A., Molina, A., Asenjo, H.G., Lopez-Onieva, L., Martorell-Marugán, J., Espinosa-Martinez, M., Griñan-Lison, C., Alvarez-Perez, J.C., Cara, F.E., Navarro-Marchal, S.A., Carmona-Sáez, P., Medina, P.P., Marchal, J.A., Granados-Principal, S., Sánchez-Pozo, A. and Landeira, D. (2022). EZH2 endorses cell plasticity to non-small cell lung cancer cells facilitating mesenchymal to epithelial transition and tumour colonization. *Oncogene*, 41(28), pp.3611–3624. doi:10.1038/s41388-022-02375-x.

Gambardella, V., Tarazona, N., Cejalvo, J.M., Lombardi, P., Huerta, M., Roselló, S., Fleitas, T., Roda, D. and Cervantes, A. (2020). Personalized Medicine: Recent Progress in Cancer Therapy. *Cancers*, [online] 12(4). doi:10.3390/cancers12041009.

Ganesh, K. and Massagué, J. (2021). Targeting metastatic cancer. *Nature Medicine*, [online] 27(1), pp.34–44. doi:10.1038/s41591-020-01195-4.

Gao, Y., Bado, I., Wang, H., Zhang, W., Rosen, J.M. and Zhang, X.H.-F. . (2019). Metastasis Organotropism: Redefining the Congenial Soil. *Developmental Cell*, 49(3), pp.375–391. doi:10.1016/j.devcel.2019.04.012.

García-Gómez, P., Priego, N., Álvaro-Espinosa, L. and Valiente, M., 2019. 5.1 Brain Vasculature: Blood-Brain Barrier, Vascular Co-option, and Angiogenesis ST6GALNAC5. *Central Nervous System Metastases*, p.59.

Gauthier, N.P., Reznik, E., Gao, J., Sumer, S.O., Schultz, N., Sander, C. and Miller, M.L. (2015). MutationAligner: a resource of recurrent mutation hotspots in protein domains in cancer. *Nucleic Acids Research*, 44(D1), pp.D986–D991. doi:10.1093/nar/gkv1132.

Giannoudis, A., Sartori, A., Eastoe, L., Zakaria, R., Charlton, C., Hickson, N., Platt-Higgins, A., Rudland, P.S., Irwin, D., Jenkinson, M.D. and Palmieri, C., 2021. Genomic profiling using the UltraSEEK panel identifies discordancy between paired primary and breast cancer brain metastases and an association with brain metastasis-free survival. *Breast Cancer Research and Treatment*, *190*(2), pp.241-253.

Gisselsson, D. (2008). Classification of chromosome segregation errors in cancer. *Chromosoma*, 117(6), pp.511–519. doi:10.1007/s00412-008-0169-1.

Gnad, F., Baucom, A., Mukhyala, K., Manning, G. and Zhang, Z. (2013). Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics*, 14(S3). doi:10.1186/1471-2164-14-s3-s7.

gnomad.broadinstitute.org. (n.d.). *gnomAD*. [online] Available at: https://gnomad.broadinstitute.org/about.

Gomis, R.R. and Gawrzak, S. (2017). Tumor cell dormancy. *Molecular Oncology*, [online] 11(1), pp.62–78. doi:10.1016/j.molonc.2016.09.009.

González-Pérez, A. and López-Bigas, N. (2011). Improving the Assessment of the Outcome of Nonsynonymous SNVs with a Consensus Deleteriousness Score, Condel. *The American Journal of Human Genetics*, [online] 88(4), pp.440–449. doi:10.1016/j.ajhg.2011.03.004.

Goradel, N.H., Mohammadi, N., Haghi-Aminjan, H., Farhood, B., Negahdari, B. and Sahebkar, A. (2018). Regulation of tumor angiogenesis by microRNAs: State of

the art. *Journal of Cellular Physiology*, 234(2), pp.1099–1110. doi:10.1002/jcp.27051.

Gozdecka, M., Lyons, S., Kondo, S., Taylor, J., Li, Y., Walczynski, J., Thiel, G., Breitwieser, W. and Jones, N. (2014). JNK Suppresses Tumor Formation via a Gene-Expression Program Mediated by ATF2. *Cell Reports*, 9(4), pp.1361–1374. doi:10.1016/j.celrep.2014.10.043.

Grote, I., Bartels, S., Kandt, L., Bollmann, L., Christgen, H., Gronewold, M., Raap, M., Lehmann, U., Gluz, O., Nitz, U., Kuemmel, S., Eulenburg, C., Braun, M., Aktas, B., Grischke, E., Schumacher, C., Luedtke-Heckenkamp, K., Kates, R., Wuerstlein, R. and Graeser, M. (2021). *TP53* mutations are associated with primary endocrine resistance in luminal early breast cancer. *Cancer Medicine*, 10(23), pp.8581–8594. doi:10.1002/cam4.4376.

Guan, X. (2015). Cancer metastases: challenges and opportunities. *Acta Pharmaceutica Sinica B*, 5(5), pp.402–418. doi:10.1016/j.apsb.2015.07.005.

Gucalp, A., Traina, T.A., Eisner, J.R., Parker, J.S., Selitsky, S.R., Park, B.H., Elias, A.D., Baskin-Bey, E.S. and Cardoso, F. (2018). Male breast cancer: a disease distinct from female breast cancer. *Breast Cancer Research and Treatment*, 173(1), pp.37–48. doi:10.1007/s10549-018-4921-9.

Gu, G., Dustin, D. and Fuqua, S.A. (2016). Targeted therapy for breast cancer and molecular mechanisms of resistance to treatment. *Current Opinion in Pharmacology*, 31, pp.97–103. doi:10.1016/j.coph.2016.11.005.

Guo, M.H., Plummer, L., Chan, Y.-M., Hirschhorn, J.N. and Lippincott, M.F. (2018). Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. *The American Journal of Human Genetics*, 103(4), pp.522–534. doi:10.1016/j.ajhg.2018.08.016.

Gu, Y., Ji, Y., Jiang, H. and Qiu, G. (2020). Clinical Effect of Driver Mutations of *KRAS, CDKN2A/P16, TP53*, and *SMAD4* in Pancreatic Cancer: A Meta-Analysis. *Genetic Testing and Molecular Biomarkers*, 24(12), pp.777–788. doi:10.1089/gtmb.2020.0078.

Habbous, S., Forster, K., Darling, G., Jerzak, K., Holloway, C.M.B., Sahgal, A. and Das, S. (2020). Incidence and real-world burden of brain metastases from solid tumors and hematologic malignancies in Ontario: a population-based study. *Neuro-Oncology Advances*, 3(1). doi:10.1093/noajnl/vdaa178.

Hahn, J., Wang, X. and Margeta, M. (2015). Astrocytes increase the activity of synaptic GluN2B NMDA receptors. *Frontiers in Cellular Neuroscience*, 9. doi:10.3389/fncel.2015.00117.

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, 12(1), pp.31–46. doi:10.1158/2159-8290.cd-21-1059.

Han, J., Liu, Y., Yang, S., Wu, X., Li, H. and Wang, Q. (2021). MEK inhibitors for the treatment of non-small cell lung cancer. *Journal of Hematology & Oncology*, 14(1). doi:10.1186/s13045-020-01025-7.

Hao, Y., Baker, D. and Ten Dijke, P., (2019). TGF-β-mediated epithelial-mesenchymal transition and cancer metastasis. *International journal of molecular sciences*, *20*(11), p.2767.

Hapach, L.A., Mosier, J.A., Wang, W. and Reinhart-King, C.A. (2019). Engineered models to parse apart the metastatic cascade. *npj Precision Oncology*, 3(1). doi:10.1038/s41698-019-0092-3.

Harvard.edu. (2019). *PolyPhen-2: prediction of functional effects of human nsSNPs*. [online] Available at: http://genetics.bwh.harvard.edu/pph2/.

Hashemi, S., Nowzari Dalini, A., Jalali, A., Banaei-Moghaddam, A.M. and Razaghi-Moghadam, Z. (2017). Cancerouspdomains: comprehensive analysis of cancer type-specific recurrent somatic mutations in proteins and domains. *BMC Bioinformatics*, 18(1). doi:10.1186/s12859-017-1779-5.

Hausser, J. and Alon, U. (2020). Tumour heterogeneity and the evolutionary trade-offs of cancer. *Nature Reviews Cancer*. doi:10.1038/s41568-020-0241-6.

Hellström, M., Phng, L.-K., Hofmann, J.J., Wallgard, E., Coultas, L., Lindblom, P., Alva, J., Nilsson, A.-K., Karlsson, L., Gaiano, N., Yoon, K., Rossant, J., Iruela-Arispe,

M.L., Kalén, M., Gerhardt, H. and Betsholtz, C. (2007). Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nature*, 445(7129), pp.776–780. doi:10.1038/nature05571.

He, M.M., Li, Q., Yan, M., Cao, H., Hu, Y., He, K.Y., Cao, K., Li, M.M. and Wang, K. (2019). Variant Interpretation for Cancer (VIC): a computational tool for assessing clinical impacts of somatic variants. *Genome Medicine*, 11(1). doi:10.1186/s13073-019-0664-4.

Henrik Heiland, D., Ravi, V.M., Behringer, S.P., Frenking, J.H., Wurm, J., Joseph, K., Garrelfs, N.W.C., Strähle, J., Heynckes, S., Grauvogel, J., Franco, P., Mader, I., Schneider, M., Potthoff, A.-L., Delev, D., Hofmann, U.G., Fung, C., Beck, J., Sankowski, R. and Prinz, M. (2019). Tumor-associated reactive astrocytes aid the evolution of immunosuppressive environment in glioblastoma. *Nature communications*, [online] 10(1), p.2541. doi:10.1038/s41467-019-10493-6.

Hetheridge, C., Mavria, G. and Mellor, H. (2011). Uses of the *in vitro* endothelial–fibroblast organotypic co-culture assay in angiogenesis research. *Biochemical Society Transactions*, 39(6), pp.1597–1600. doi:10.1042/bst20110738.

Hillen, F. and Griffioen, A.W. (2007). Tumour vascularization: sprouting angiogenesis and beyond. *Cancer and Metastasis Reviews*, 26(3-4), pp.489–502. doi:10.1007/s10555-007-9094-7.

Hodgkinson, A., Casals, F., Idaghdour, Y., Grenier, J.-C., Hernandez, R.D. and Awadalla, P. (2013). Selective constraint, background selection, and mutation accumulation variability within and between human populations. *BMC Genomics*, 14(1). doi:10.1186/1471-2164-14-495.

Hofmann, J.J. and Iruela-Arispe, M.L. (2007). Notch Signaling in Blood Vessels. *Circulation Research*, 100(11), pp.1556–1568. doi:10.1161/01.res.0000266408.42939.e4.

Hongu, T. and Oskarsson, T. (2020). Addicted to Acidic Microenvironment. *Developmental Cell*, 55(4), pp.381–382. doi:10.1016/j.devcel.2020.11.004.

Horgan, D., Curigliano, G., Rieß, O., Hofman, P., Büttner, R., Conte, P., Cufer, T., Gallagher, W.M., Georges, N., Kerr, K., Penault-Llorca, F., Mastris, K., Pinto, C., Van Meerbeeck, J., Munzone, E., Thomas, M., Ujupan, S., Vainer, G.W., Velthaus, J.-L. and André, F. (2022). Identifying the Steps Required to Effectively Implement Next-Generation Sequencing in Oncology at a National Level in Europe. *Journal of Personalized Medicine*, 12(1), p.72. doi:10.3390/jpm12010072.

Hosonaga, M., Saya, H. and Arima, Y. (2020). Molecular and cellular mechanisms underlying brain metastasis of breast cancer. *Cancer and Metastasis Reviews*. doi:10.1007/s10555-020-09881-y.

Hou, J., He, Z., Liu, T., Chen, D., Wang, B., Wen, Q. and Zheng, X. (2022). Evolution of Molecular Targeted Cancer Therapy: Mechanisms of Drug Resistance and Novel Opportunities Identified by CRISPR-Cas9 Screening. *Frontiers in Oncology*, 12. doi:10.3389/fonc.2022.755053.

Howe, G.A., Kazda, K. and Addison, C.L. (2017). MicroRNA-30b controls endothelial cell capillary morphogenesis through regulation of transforming growth factor beta 2. *PLOS ONE*, 12(10), p.e0185619. doi:10.1371/journal.pone.0185619.

Huang, Z., Zhao, B., Qin, Z., Li, Y., Wang, T., Zhou, W., Zheng, J., Yang, S., Shi, Y., Fan, Y. and Xiang, R. (2019). Novel dual inhibitors targeting CDK4 and VEGFR2 synergistically suppressed cancer progression and angiogenesis. *European Journal of Medicinal Chemistry*, 181, p.111541. doi:10.1016/j.ejmech.2019.07.044.

Huang, Z., Zhuo, Y., Shen, Z., Wang, Y., Wang, L., Li, H., Chen, J. and Chen, W. (2013). The role of NEFL in cell growth and invasion in head and neck squamous cell carcinoma cell lines. *Journal of Oral Pathology & Medicine*, 43(3), pp.191–198. doi:10.1111/jop.12109.

Huebner, K., Procházka, J., Monteiro, A.C., Mahadevan, V. and Schneider-Stock, R. (2019). The activating transcription factor 2: an influencer of cancer progression. *Mutagenesis*, 34(5-6), pp.375–389. doi:10.1093/mutage/gez041.

Hutchinson, L. (2015). Understanding metastasis. *Nature Reviews Clinical Oncology*, 12(5), pp.247–247. doi:10.1038/nrclinonc.2015.71.

Hu, T., Chitnis, N., Monos, D. and Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Elsevier*, [online] 82(11). doi:https://doi.org/10.1016/j.humimm.2021.02.012.

Hutter, C. and Zenklusen, J.C. (2018). The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell*, 173(2), pp.283–285. doi:10.1016/j.cell.2018.03.042.

Hu, X., Fu, M., Zhao, X. and Wang, W., 2021. The JAK/STAT signaling pathway: From bench to clinic. *Signal Transduction and Targeted Therapy*, *6*(1), pp.1-33.

In, G.K., Poorman, K., Saul, M., O'Day, S., Farma, J.M., Olszanski, A.J., Gordon, M.S., Thomas, J.S., Eisenberg, B., Flaherty, L. and Weise, A., 2020. Molecular profiling of melanoma brain metastases compared to primary cutaneous melanoma and to extracranial metastases. *Oncotarget*, *11*(33), p.3118.

Internationalgenome.org. (2012). *Phase 3 | 1000 Genomes*. [online] Available at: https://www.internationalgenome.org/category/phase-3/ [Accessed 12 Nov. 2022].

InterPro EMBL-EBI (2019). *InterPro protein sequence analysis & classification < InterPro < EMBL-EBI*. [online] Ebi.ac.uk. Available at: https://www.ebi.ac.uk/interpro/.

Iwamoto, T., Niikura, N., Ogiya, R., Yasojima, H., Watanabe, K., Kanbayashi, C., Tsuneizumi, M., Matsui, A., Fujisawa, T., Iwasa, T., Shien, T., Saji, S., Masuda, N. and Iwata, H. (2019). Distinct gene expression profiles between primary breast cancers and brain metastases from pair-matched samples. *Scientific Reports*, 9(1). doi:10.1038/s41598-019-50099-y.

Izraely, S. and Witz, I.P. (2020). Site-specific metastasis: A cooperation between cancer cells and the metastatic microenvironment. *International Journal of Cancer*, 148(6), pp.1308–1322. doi:10.1002/ijc.33247.

Jalali Sefid Dashti, M. and Gamieldien, J. (2017). A practical guide to filtering and prioritizing genetic variants. *BioTechniques*, [online] 62(1), pp.18–30. doi:10.2144/000114492.

Jandrig, B., Seitz, S., Hinzmann, B., Arnold, W., Micheel, B., Koelble, K., Siebert, R., Schwartz, A., Ruecker, K., Schlag, P.M., Scherneck, S. and Rosenthal, A. (2004). ST18 is a breast cancer tumor suppressor gene at human chromosome 8q11.2. *Oncogene*, 23(57), pp.9295–9302. doi:10.1038/sj.onc.1208131.

Jiang, S.-H., Hu, L.-P., Wang, X., Li, J. and Zhang, Z.-G. (2019). Neurotransmitters: emerging targets in cancer. *Oncogene*, 39(3), pp.503–515. doi:10.1038/s41388-019-1006-0.

Jin, C., Li, H., Murata, T., Sun, K., Horikoshi, M., Chiu, R. and Yokoyama, K.K. (2002). JDP2, a Repressor of AP-1, Recruits a Histone Deacetylase 3 Complex To Inhibit the Retinoic Acid-Induced Differentiation of F9 Cells. *Molecular and Cellular Biology*, 22(13), pp.4815–4826. doi:10.1128/mcb.22.13.4815-4826.2002.

Jin, J., Shi, Y., Zhang, S. and Yang, S. (2019a). PIK3CA mutation and clinicopathological features of colorectal cancer: a systematic review and Meta-Analysis. *Acta Oncologica*, 59(1), pp.66–74. doi:10.1080/0284186x.2019.1664764.

Jin, Y., Zhang, Y., Li, B., Zhang, J., Dong, Z., Hu, X. and Wan, Y. (2019b). TRIM21 mediates ubiquitination of Snail and modulates epithelial to mesenchymal transition in breast cancer cells. *International Journal of Biological Macromolecules*, 124, pp.846–853. doi:10.1016/j.ijbiomac.2018.11.269.

Joosse, S.A., Hannemann, J., Spötter, J., Bauche, A., Andreas, A., Müller, V. and Pantel, K. (2012). Changes in Keratin Expression during Metastatic Progression of Breast Cancer: Impact on the Detection of Circulating Tumor Cells. *Clinical Cancer Research*, 18(4), pp.993–1003. doi:10.1158/1078-0432.ccr-11-2100.

Kang, E., Seo, J., Yoon, H. and Cho, S. (2021). The Post-Translational Regulation of Epithelial–Mesenchymal Transition-Inducing Transcription Factors in Cancer

Metastasis. *International Journal of Molecular Sciences*, 22(7), p.3591. doi:10.3390/ijms22073591.

Karakas, B., Bachman, K.E. and Park, B.H. (2006). Mutation of the PIK3CA oncogene in human cancers. *British Journal of Cancer*, [online] 94(4), pp.455–459. doi:10.1038/sj.bjc.6602970.

Khanna, S., Komati, R., Eichenbaum, D.A., Hariprasad, I., Ciulla, T.A. and Hariprasad, S.M. (2019). Current and upcoming anti-VEGF therapies and dosing strategies for the treatment of neovascular AMD: a comparative review. *BMJ Open Ophthalmology*, [online] 4(1). doi:10.1136/bmjophth-2019-000398.

Kim, J.-W., Akiyama, M., Park, J.-H., Lin, M.-L., Shimo, A., Ueki, T., Daigo, Y., Tsunoda, T., Nishidate, T., Nakamura, Y. and Katagiri, T. (2009). Activation of an estrogen/estrogen receptor signaling by BIG3 through its inhibitory effect on nuclear transport of PHB2/REA in breast cancer. *Cancer Science*, 100(8), pp.1468–1478. doi:10.1111/j.1349-7006.2009.01209.x.

King, H.W., Michael, M.Z. and Gleadle, J.M. (2012). Hypoxic enhancement of exosome release by breast cancer cells. *BMC cancer*, [online] 12, p.421. doi:10.1186/1471-2407-12-421.

Kobayashi, Y., Yang, S., Nykamp, K., Garcia, J., Lincoln, S.E. and Topper, S.E. (2017). Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. *Genome Medicine*, 9(1). doi:10.1186/s13073-017-0403-7.

Koboldt, D.C. (2020). Best practices for variant calling in clinical sequencing. *Genome Medicine*, 12(1). doi:10.1186/s13073-020-00791-w.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., Wilson, R.K., Ally, A., Balasundaram, M., Butterfield, Y.S.N., Carlsen, R., Carter, C., Chu, A., Chuah, E., Chun, H.-J.E. and Coope, R.J.N. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), pp.61–70. doi:10.1038/nature11412.

Koch, S., Tugues, S., Li, X., Gualandi, L. and Claesson-Welsh, L. (2011). Signal transduction by vascular endothelial growth factor receptors. *Biochemical Journal*, 437(2), pp.169–183. doi:10.1042/bj20110301.

Koniali, L., Hadjisavvas, A., Constantinidou, A., Christodoulou, K., Christou, Y., Demetriou, C., Panayides, A.S., Pitris, C., Pattichis, C.S., Zamba-Papanicolaou, E. and Kyriacou, K., 2020. Risk factors for breast cancer brain metastases: a systematic review. *Oncotarget*, *11*(6), p.650.

Kornreich, M., Avinery, R., Malka-Gibor, E., Laser-Azogui, A. and Beck, R. (2015). Order and disorder in intermediate filament proteins. *FEBS Letters*, 589(19PartA), pp.2464–2476. doi:10.1016/j.febslet.2015.07.024.

Kotecki, N., Lefranc, F., Devriendt, D. and Awada, A. (2018). Therapy of breast cancer brain metastases: challenges, emerging treatments and perspectives. *Therapeutic Advances in Medical Oncology*, 10, p.175883591878031. doi:10.1177/1758835918780312.

Koudonas, A., Papaioannou, M., Kampantais, S., Anastasiadis, A., Hatzimouratidis, K. and Dimitriadis, G. (2022). Methylation of PCDH17 and NEFH as prognostic biomarker for nonmetastatic RCC: A cohort study. *Medicine*, 101(28), p.e29599. doi:10.1097/md.0000000000029599.

Krebs, L.T. (2004). Haploinsufficient lethality and formation of arteriovenous malformations in Notch pathway mutants. *Genes & Development*, 18(20), pp.2469–2473. doi:10.1101/gad.1239204.

Krishna Priya, S., Nagare, R.P., Sneha, V.S., Sidhanth, C., Bindhya, S., Manasa, P. and Ganesan, T.S. (2016). Tumour angiogenesis-Origin of blood vessels. *International Journal of Cancer*, 139(4), pp.729–735. doi:10.1002/ijc.30067.

Krøigård, A.B., Larsen, M.J., Brasch-Andersen, C., Lænkholm, A.-V., Knoop, A.S., Jensen, J.D., Bak, M., Mollenhauer, J., Thomassen, M. and Kruse, T.A. (2017). Genomic Analyses of Breast Cancer Progression Reveal Distinct Routes of Metastasis Emergence. *Scientific Reports*, [online] 7(1). doi:10.1038/srep43813.

Kuhnert, F., Kirshner, J.R. and Thurston, G. (2011). Dll4-Notch signaling as a therapeutic target in tumor angiogenesis. *Vascular Cell*, 3(1), p.20. doi:10.1186/2045-824x-3-20.

Kumar, P., Henikoff, S. and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7), pp.1073–1081. doi:10.1038/nprot.2009.86.

Kumar, S., Srivastav, R.K., Wilkes, D.W., Ross, T., Kim, S., Kowalski, J., Chatla, S., Zhang, Q., Nayak, A., Guha, M., Fuchs, S.Y., Thomas, C. and Chakrabarti, R. (2019). Estrogen-dependent DLL1-mediated Notch signaling promotes luminal breast cancer. *Oncogene*, [online] 38(12), pp.2092–2107. doi:10.1038/s41388-018-0562-z.

Kuol, N., Stojanovska, L., Apostolopoulos, V. and Nurgali, K. (2018). Role of the nervous system in cancer metastasis. *Journal of Experimental & Clinical Cancer Research*, [online] 37(1). doi:10.1186/s13046-018-0674-x.

Lai, W.-A., Liu, C.-Y., Lin, S.-Y., Chen, C.-C. and Hang, J.-F. (2020). Characterization of Driver Mutations in Anaplastic Thyroid Carcinoma Identifies RAS and PIK3CA Mutations as Negative Survival Predictors. *Cancers*, 12(7), p.1973. doi:10.3390/cancers12071973.

Lamouille, S., Xu, J. and Derynck, R. (2014). Molecular mechanisms of epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology*, 15(3), pp.178–196. doi:10.1038/nrm3758.

Langley, R.R. and Fidler, I.J. (2011). The seed and soil hypothesis revisited-The role of tumor-stroma interactions in metastasis to different organs. *International Journal of Cancer*, 128(11), pp.2527–2535. doi:10.1002/ijc.26031.

Laskar, S., Das, R., Kundu, S., Saha, A., Nandi, N., Choudhury, Y. and Kumar Ghosh, S. (2022). Whole exome sequencing identifies the potential role of genes involved in p53 pathway in Nasopharyngeal Carcinoma from Northeast India. *Gene*, 812, p.146099. doi:10.1016/j.gene.2021.146099.

Lasky-Su, J. (2016). *Clinical and Translational Science Principles of Human Research*. Second ed. Elsevier Inc., pp.347–362.

Lau, E. and Ronai, Z.A. (2012). ATF2 – at the crossroad of nuclear and cytosolic functions. *Journal of Cell Science*. doi:10.1242/jcs.095000.

Lau, E., Sedy, J., Sander, C., Shaw, M.A., Feng, Y., Scortegagna, M., Claps, G., Robinson, S., Cheng, P., Srivas, R., Soonthornvacharin, S., Ideker, T., Bosenberg, M., Gonzalez, R., Robinson, W., Chanda, S.K., Ware, C., Dummer, R., Hoon, D. and Kirkwood, J.M. (2015). Transcriptional repression of IFNβ1 by ATF2 confers melanoma resistance to therapy. *Oncogene*, 34(46), pp.5739–5748. doi:10.1038/onc.2015.22.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A. and Kiezun, A., 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, *499*(7457), pp.214-218.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., Kiezun, A., Hammerman, P.S., McKenna, A., Drier, Y., Zou, L., Ramos, A.H., Pugh, T.J., Stransky, N., Helman, E. and Kim, J. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457), pp.214–218. doi:10.1038/nature12213.

Lee, J.Y., Park, K., Lim, S.H., Kim, H.S., Yoo, K.H., Jung, K.S., Song, H.-N., Hong, M., Do, I.-G., Ahn, T., Lee, S.K., Bae, S.Y., Kim, S.W., Lee, J.E., Nam, S.J., Kim, D.-H., Jung, H.H., Kim, J.-Y., Ahn, J.S. and Im, Y.-H. (2015). Mutational profiling of brain metastasis from breast cancer: matched pair analysis of targeted sequencing between brain metastasis and primary breast cancer. *Oncotarget*, 6(41), pp.43731–43742. doi:10.18632/oncotarget.6192.

Lee, Y.T., Tan, Y.J. and Oon, C.E. (2018). Molecular targeted therapy: Treating cancer with specificity. *European Journal of Pharmacology*, 834, pp.188–196. doi:10.1016/j.ejphar.2018.07.034.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J. and Cooper, D.N. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, [online] 536(7616), pp.285–291. doi:10.1038/nature19057.

Liberti, M.V. and Locasale, J.W. (2016). Correction to: 'The Warburg Effect: How Does it Benefit Cancer Cells?' *Trends in Biochemical Sciences*, 41(3), p.287. doi:10.1016/j.tibs.2016.01.004.

Li, H., Wei, S., Cheng, K., Gounko, N.V., Ericksen, R.E., Xu, A., Hong, W. and Han, W., 2014. BIG 3 inhibits insulin granule biogenesis and insulin secretion. *EMBO reports*, *15*(6), pp.714-722.

Li, J., Xu, Y., Zhu, H., Wang, Y., Li, P. and Wang, D. (2022). The dark side of synaptic proteins in tumours. *British Journal of Cancer*, 127(7), pp.1184–1192. doi:10.1038/s41416-022-01863-x.

Li, L. and Hanahan, D. (2013). Hijacking the Neuronal NMDAR Signaling Circuit to Promote Tumor Growth and Invasion. *Cell*, 153(1), pp.86–100. doi:10.1016/j.cell.2013.02.051.

Limbourg, A., Ploom, M., Elligsen, D., SörensenI., Ziegelhoeffer, T., Gossler, A., Drexler, H. and Limbourg, F.P. (2007). Notch Ligand Delta-Like 1 Is Essential for Postnatal Arteriogenesis. *Circulation Research*, 100(3), pp.363–371. doi:10.1161/01.res.0000258174.77370.2c.

Lim, Z.-F. and Ma, P.C. (2019). Emerging insights of tumor heterogeneity and drug resistance mechanisms in lung cancer targeted therapy. *Journal of Hematology & Oncology*, [online] 12(1). doi:10.1186/s13045-019-0818-2.

Liu, B.-H., Xu, Y., Yuan, F.-E. and Chen, Q.-X. (2018). Molecular mechanisms involved in angiogenesis and potential target of antiangiogenesis in human glioblastomas. *Glioma*, 1(2), p.35. doi:10.4103/glioma.glioma_10_17.

Liu, J., Huang, Y., Li, T., Jiang, Z., Zeng, L. and Hu, Z. (2021a). The role of the Golgi apparatus in disease (Review). *International Journal of Molecular Medicine*, 47(4). doi:10.3892/ijmm.2021.4871.

Liu, J., Smith, S. and Wang, C. (2022). Reversing the Epithelial–Mesenchymal Transition in Metastatic Cancer Cells Using CD146-Targeted Black Phosphorus Nanosheets and a Mild Photothermal Treatment. *ACS Nano*, 16(2), pp.3208–3220. doi:10.1021/acsnano.1c11070.

Liu, Q., Zhang, H., Jiang, X., Qian, C., Liu, Z. and Luo, D. (2017). Factors involved in cancer metastasis: a better understanding to 'seed and soil' hypothesis. *Molecular Cancer*, 16(1). doi:10.1186/s12943-017-0742-4.

Liu, T., Li, H., Gounko, N.V., Zhou, Z., Xu, A., Hong, W. and Han, W., 2014. Detection of insulin granule exocytosis by an electrophysiology method with high temporal resolution reveals enlarged insulin granule pool in BIG3-knockout mice. *American Journal of Physiology-Endocrinology and Metabolism*, *307*(7), pp.E611-E618.

Liu, T., Li, H., Hong, W. and Han, W. (2016a). Brefeldin A-inhibited guanine nucleotide exchange protein 3 is localized in lysosomes and regulates GABA signaling in hippocampal neurons. *Journal of Neurochemistry*, 139(5), pp.748–756. doi:10.1111/jnc.13859.

Liu, Y., Maekawa, T., Yoshida, K., Furuse, T., Kaneda, H., Wakana, S. and Ishii, S. (2016b). ATF7 ablation prevents diet-induced obesity and insulin resistance. *Biochemical and Biophysical Research Communications*, 478(2), pp.696–702. doi:10.1016/j.bbrc.2016.08.009.

Liu, Y., White, K.A. and Barber, D.L. (2020). Intracellular pH Regulates Cancer and Stem Cell Behaviors: A Protein Dynamics Perspective. *Frontiers in Oncology*, 10. doi:10.3389/fonc.2020.01401.

Liu, Z., Zheng, M., Lei, B., Zhou, Z., Huang, Y., Li, W., Chen, Q., Li, P. and Deng, Y. (2021b). Whole-exome sequencing identifies somatic mutations associated with

lung cancer metastasis to the brain. *Annals of Translational Medicine*, 9(8), pp.694–694. doi:10.21037/atm-21-1555.

Li, Z., Seehawer, M. and Polyak, K. (2022). Untangling the web of intratumour heterogeneity. *Nature Cell Biology*, 24(8), pp.1192–1201. doi:10.1038/s41556-022-00969-x.

Loh, Chai, Tang, Wong, Sethi, Shanmugam, Chong and Looi (2019). The E-Cadherin and N-Cadherin Switch in Epithelial-to-Mesenchymal Transition: Signaling, Therapeutic Implications, and Challenges. *Cells*, [online] 8(10), p.1118. doi:10.3390/cells8101118.

Lopez-Bergami, P., Lau, E. and Ronai, Z. (2010). Emerging roles of ATF2 and the dynamic AP1 network in cancer. *Nature Reviews Cancer*, 10(1), pp.65–76. doi:10.1038/nrc2681.

Louie, E., Chen, X.F., Coomes, A., Ji, K., Tsirka, S. and Chen, E.I., 2013. Neurotrophin-3 modulates breast cancer cells and the microenvironment to promote the growth of breast cancer brain metastasis. *Oncogene*, *32*(35), pp.4064-4077.

Lowery, F.J. and Yu, D., 2017. Brain metastasis: Unique challenges and open opportunities. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *1867*(1), pp.49-57.

Lugano, R., Ramachandran, M. and Dimberg, A. (2019). Tumor angiogenesis: causes, consequences, challenges and opportunities. *Cellular and Molecular Life Sciences*, 77(9), pp.1745–1770. doi:10.1007/s00018-019-03351-7.

Luthra, R., Chen, H., Roy-Chowdhuri, S. and Singh, R. (2015). Next-Generation Sequencing in Clinical Molecular Diagnostics of Cancer: Advantages and Challenges. *Cancers*, 7(4), pp.2023–2036. doi:10.3390/cancers7040874.

Lu, Z., Zou, J., Li, S., Topper, M.J., Tao, Y., Zhang, H., Jiao, X., Xie, W., Kong, X., Vaz, M., Li, H., Cai, Y., Xia, L., Huang, P., Rodgers, K., Lee, B., Riemer, J.B., Day, C.-P., Yen, R.-W.C. and Cui, Y. (2020). Epigenetic therapy inhibits

metastases by disrupting premetastatic niches. *Nature*, 579(7798), pp.284–290. doi:10.1038/s41586-020-2054-x.

Lv, Y., Ma, X., Du, Y. and Feng, J. (2021). Understanding Patterns of Brain Metastasis in Triple-Negative Breast Cancer and Exploring Potential Therapeutic Targets. *OncoTargets and Therapy*, Volume 14, pp.589–607. doi:10.2147/ott.s293685.

Macedo, F., Ladeira, K., Longatto-Filho, A. and Martins, S.F. (2017). Gastric Cancer and Angiogenesis: Is VEGF a Useful Biomarker to Assess Progression and Remission? *Journal of Gastric Cancer*, 17(1), p.1. doi:10.5230/jgc.2017.17.e1.

Manjunath, G.P., Ramanujam, P.L. and Galande, S. (2017). Structure function relations in PDZ-domain-containing proteins: Implications for protein networks in cellular signalling. *Journal of Biosciences*, 43(1), pp.155–171. doi:10.1007/s12038-017-9727-0.

Martinelli, E., Morgillo, F., Troiani, T. and Ciardiello, F. (2017). Cancer resistance to therapies against the EGFR-RAS-RAF pathway: The role of MEK. *Cancer Treatment Reviews*, 53, pp.61–69. doi:10.1016/j.ctrv.2016.12.001.

Martínez-Jiménez, F., Muiños, F., Sentís, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., Mularoni, L., Pich, O., Bonet, J., Kranas, H., Gonzalez-Perez, A. and Lopez-Bigas, N. (2020). A compendium of mutational cancer driver genes. *Nature Reviews Cancer*, [online] 20(10), pp.555–572. doi:10.1038/s41568-020-0290-x.

Martínez-Pizarro, A., Dembic, M., Pérez, B., Andresen, B.S. and Desviat, L.R. (2018). Intronic PAH gene mutations cause a splicing defect by a novel mechanism involving U1snRNP binding downstream of the 5' splice site. *PLOS Genetics*, 14(4), p.e1007360. doi:10.1371/journal.pgen.1007360.

Matityahu, A. and Onn, I. (2017). A new twist in the coil: functions of the coiled-coil domain of structural maintenance of chromosome (SMC) proteins. *Current Genetics*, 64(1), pp.109–116. doi:10.1007/s00294-017-0735-2.

Mattiuzzi, C. and Lippi, G. (2019). Current Cancer Epidemiology. *Journal of Epidemiology and Global Health*, [online] 9(4), pp.217–222. doi:10.2991/jegh.k.191008.001.

McCourt, C.M., McArt, D.G., Mills, K., Catherwood, M.A., Maxwell, P., Waugh, D.J., Hamilton, P., O'Sullivan, J.M. and Salto-Tellez, M. (2013). Validation of Next Generation Sequencing Technologies in Comparison to Current Diagnostic Gold Standards for BRAF, EGFR and KRAS Mutational Analysis. *PLoS ONE*, 8(7), p.e69604. doi:10.1371/journal.pone.0069604.

McMullin, R.P., Wittner, B.S., Yang, C., Denton-Schneider, B.R., Hicks, D., Singavarapu, R., Moulis, S., Lee, J., Akbari, M.R., Narod, S.A., Aldape, K.D., Steeg, P.S., Ramaswamy, S. and Sgroi, D.C. (2014). A BRCA1deficient-like signature is enriched in breast cancer brain metastases and predicts DNA damage-induced poly (ADP-ribose) polymerase inhibitor sensitivity. *Breast Cancer Research*, 16(2). doi:10.1186/bcr3625.

Michiels, T.J.M., Meiring, H.D., Jiskoot, W., Kersten, G.F.A. and Metz, B. (2020). Formaldehyde treatment of proteins enhances proteolytic degradation by the endo-lysosomal protease cathepsin S. *Scientific Reports*, [online] 10, p.11535. doi:10.1038/s41598-020-68248-z.

Miller, K.D., Nogueira, L., Mariotto, A.B., Rowland, J.H., Yabroff, K.R., Alfano, C.M., Jemal, A., Kramer, J.L. and Siegel, R.L. (2019). Cancer treatment and survivorship statistics, 2019. *CA: A Cancer Journal for Clinicians*, [online] 69(5). doi:10.3322/caac.21565.

Miller, Martin L., Reznik, E., Gauthier, Nicholas P., Aksoy, B., Korkut, A., Gao, J., Ciriello, G., Schultz, N. and Sander, C. (2015). Pan-Cancer Analysis of Mutation Hotspots in Protein Domains. *Cell Systems*, [online] 1(3), pp.197–209. doi:10.1016/j.cels.2015.08.014.

Mills, M.N., Figura, N.B., Arrington, J.A., Yu, H.H.M., Etame, A.B., Vogelbaum, M.A., Soliman, H., Czerniecki, B.J., Forsyth, P.A., Han, H.S. and Ahmed, K.A., 2020. Management of brain metastases in breast cancer: a review of current practices

and emerging treatments. *Breast cancer research and treatment*, *180*(2), pp.279-300.

Min, H.-Y. and Lee, H.-Y. (2022). Molecular targeted therapy for anticancer treatment. *Experimental & Molecular Medicine*, 54(10), pp.1670–1694. doi:10.1038/s12276-022-00864-3.

Monroe, J.D., Basheer, F. and Gibert, Y. (2021). Xmrks the Spot: Fish Models for Investigating Epidermal Growth Factor Receptor Signaling in Cancer Research. *Cells, 10(5)*, [online] p.1132. doi:https://doi.org/10.3390/cells10051132.

Morgan, A.J., Giannoudis, A. and Palmieri, C. (2021). The genomic landscape of breast cancer brain metastases: a systematic review. *The Lancet Oncology*, 22(1), pp.e7–e17. doi:10.1016/s1470-2045(20)30556-8.

Mujcic, H., Hill, R.P., Koritzinsky, M. and Wouters, B.G. (2014). Hypoxia Signaling and the Metastatic Phenotype. *Current Molecular Medicine*, 14(5), pp.565–579. doi:10.2174/1566524014666140603115831.

Mukamel, R.E., Handsaker, R.E., Sherman, M.A., Barton, A.R., Zheng, Y., McCarroll, S.A. and Loh, P.-R. (2021). Protein-coding repeat polymorphisms strongly shape diverse human phenotypes. *Science*, 373(6562), pp.1499–1505. doi:10.1126/science.abg8289.

National Center for Biotechnology Information (2017). *NCBI Conserved Domain Search*. [online] Nih.gov. Available at: https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi.

Naxerova, K. and Jain, R.K. (2015). Using tumour phylogenetics to identify the roots of metastasis in humans. *Nature Reviews Clinical Oncology*, 12(5), pp.258–272. doi:10.1038/nrclinonc.2014.238.

Negrini, S., Gorgoulis, V.G. and Halazonetis, T.D. (2010). Genomic instability — an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology*, [online] 11(3), pp.220–228. doi:10.1038/nrm2858.

Nelakurti, D.D., Rossetti, T., Husbands, A.Y. and Petreaca, R.C. (2021). Arginine Depletion in Human Cancers. *Cancers*, 13(24), p.6274. doi:10.3390/cancers13246274.

Neman, J., Termini, J., Wilczynski, S., Vaidehi, N., Choy, C., Kowolik, C.M., Li, H., Hambrecht, A.C., Roberts, E. and Jandial, R. (2014). Human breast cancer metastases to the brain display GABAergic properties in the neural niche. *Proceedings of the National Academy of Sciences of the United States of America*, [online] 111(3), pp.984–989. doi:10.1073/pnas.1322098111.

Ng, P.C. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research*, 31(13), pp.3812–3814. doi:10.1093/nar/gkg509.

Niculescu, V.F. (2019). The reproductive life cycle of cancer: Hypotheses of cell of origin, TP53 drivers and stem cell conversions in the light of the atavistic cancer cell theory. *Medical Hypotheses*, 123, pp.19–23. doi:10.1016/j.mehy.2018.12.006.

Niroula, A. and Vihinen, M. (2019). How good are pathogenicity predictors in detecting benign variants? *PLOS Computational Biology*, 15(2), p.e1006481. doi:10.1371/journal.pcbi.1006481.

Noguera-Troise, I., Daly, C., Papadopoulos, N.J., Coetzee, S., Boland, P., Gale, N.W., Chieh Lin, H., Yancopoulos, G.D. and Thurston, G. (2006). Blockade of Dll4 inhibits tumour growth by promoting non-productive angiogenesis. *Nature*, 444(7122), pp.1032–1037. doi:10.1038/nature05355.

Ortega-Molina, A., Boss, I.W., Canela, A., Pan, H., Jiang, Y., Zhao, C., Jiang, M., Hu, D., Agirre, X., Niesvizky, I., Lee, J.-E., Chen, H.-T., Ennishi, D., Scott, D.W., Mottok, A., Hother, C., Liu, S., Cao, X.-J., Tam, W. and Shaknovich, R. (2015). The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nature Medicine*, [online] 21(10), pp.1199–1208. doi:10.1038/nm.3943.

Ou, A., Ott, M., Fang, D. and Heimberger, A.B. (2021). The Role and Therapeutic Targeting of JAK/STAT Signaling in Glioblastoma. *Cancers*, 13(3), p.437. doi:10.3390/cancers13030437.

Owusu-Akyaw, A., Krishnamoorthy, K., Goldsmith, L.T. and Morelli, S.S. (2018). The role of mesenchymal–epithelial transition in endometrial function. *Human Reproduction Update*, 25(1), pp.114–133. doi:10.1093/humupd/dmy035.

Padmanaban, V., Grasset, E.M., Neumann, N.M., Fraser, A.K., Henriet, E., Matsui, W., Tran, P.T., Cheung, K.J., Georgess, D. and Ewald, A.J. (2020). Organotypic culture assays for murine and human primary and metastatic-site tumors. *Nature Protocols*, 15(8), pp.2413–2442. doi:10.1038/s41596-020-0335-3.

Palego, L., Betti, L., Rossi, A. and Giannaccini, G. (2016). Tryptophan Biochemistry: Structural, Nutritional, Metabolic, and Medical Aspects in Humans. *Journal of Amino Acids*, 2016, pp.1–13. doi:10.1155/2016/8952520.

Panagiotou, O.A., Evangelou, E. and Ioannidis, J.P.A. (2010). Genome-wide Significant Associations for Variants With Minor Allele Frequency of 5% or Less—An Overview: A HuGE Review. *American Journal of Epidemiology*, 172(8), pp.869–889. doi:10.1093/aje/kwq234.

Paoli, C. and Carrer, A. (2020). Organotypic Culture of Acinar Cells for the Study of Pancreatic Cancer Initiation. *Cancers*, 12(9), p.2606. doi:10.3390/cancers12092606.

Park, S.H., Kim, M.J., Jung, H.H., Chang, W.S., Choi, H.S., Rachmilevitch, I., Zadicario, E. and Chang, J.W., 2020. Safety and feasibility of multiple blood-brain barrier disruptions for the treatment of glioblastoma in patients undergoing standard adjuvant chemotherapy. *Journal of neurosurgery*, *134*(2), pp.475-483.

Parlani, M., Jorgez, C. and Friedl, P. (2022). Plasticity of cancer invasion and energy metabolism. *Trends in Cell Biology*. doi:10.1016/j.tcb.2022.09.009.

Pascale, R.M., Calvisi, D.F., Simile, M.M., Feo, C.F. and Feo, F. (2020). The Warburg Effect 97 Years after Its Discovery. *Cancers*, [online] 12(10). doi:10.3390/cancers12102819.

Pascual, G., Domínguez, D. and Benitah, S.A. (2018). The contributions of cancer cell metabolism to metastasis. *Disease Models & Mechanisms*, 11(8). doi:10.1242/dmm.032920.

Pearson, H.B., Li, J., Meniel, V.S., Fennell, C.M., Waring, P., Montgomery, K.G., Rebello, R.J., Macpherson, A.A., Koushyar, S., Furic, L., Cullinane, C., Clarkson, R.W., Smalley, M.J., Simpson, K.J., Phesse, T.J., Shepherd, P.R., Humbert, P.O., Sansom, O.J. and Phillips, W.A. (2018). Identification of Pik3ca Mutation as a Genetic Driver of Prostate Cancer That Cooperates with Pten Loss to Accelerate Progression and Castration-Resistant Growth. *Cancer Discovery*, 8(6), pp.764–779. doi:10.1158/2159-8290.cd-17-0867.

Pedrosa, R.M.S.M., Mustafa, D.A., Soffietti, R. and Kros, J.M. (2018). Breast cancer brain metastasis: molecular mechanisms and directions for treatment. *Neuro-Oncology*, [online] 20(11), pp.1439–1449. doi:10.1093/neuonc/noy044.

Peinado, H., Lavotshkin, S. and Lyden, D. (2011). The secreted factors responsible for pre-metastatic niche formation: old sayings and new thoughts. *Seminars in Cancer Biology*, [online] 21(2), pp.139–146. doi:10.1016/j.semcancer.2011.01.002.

Peng, L., Jiang, J., Tang, B., Nice, E.C., Zhang, Y.-Y. and Xie, N. (2020). Managing therapeutic resistance in breast cancer: from the lncRNAs perspective. *Theranostics*, [online] 10(23), pp.10360–10377. doi:10.7150/thno.49922.

Peng, W., Wang, J., Cai, J., Chen, L., Li, M. and Wu, F.-X. (2014). Improving protein function prediction using domain and protein complexes in PPI networks. *BMC Systems Biology*, 8(1), p.35. doi:10.1186/1752-0509-8-35.

Pereira, B., Chin, S.-F., Rueda, O.M., Vollan, H.-K.M., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.-J., Tsui, D.W.Y., Liu, B., Dawson, S.-J., Abraham, J., Northen, H., Peden, J.F., Mukherjee, A., Turashvili, G., Green, A.R. and McKinney, S. (2016). Erratum: The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nature Communications*, 7(1). doi:10.1038/ncomms11908.

Perelroizen, R., Philosof, B., Budick-Harmelin, N., Chernobylsky, T., Ron, A., Katzir, R., Shimon, D., Tessler, A., Adir, O., Gaoni-Yogev, A., Meyer, T., Krivitsky, A., Shidlovsky, N., Madi, A., Ruppin, E. and Mayo, L. (2022). Astrocyte immunometabolic regulation of the tumour microenvironment drives glioblastoma pathogenicity. *Brain: A Journal of Neurology*, [online] p.awac222. doi:10.1093/brain/awac222.

Pergialiotis, V., Nikolaou, C., Haidopoulos, D., Frountzas, M., Thomakos, N., Bellos, I., Papapanagiotou, A. and Rodolakis, A. (2020). PIK3CA Mutations and Their Impact on Survival Outcomes of Patients with Cervical Cancer: A Systematic Review. *Acta Cytologica*, 64(6), pp.547–555. doi:10.1159/000509095.

Peterson, T.A., Adadey, A., Santana-Cruz, I., Sun, Y., Winder, A. and Kann, M.G. (2010). DMDM: domain mapping of disease mutations. *Bioinformatics*, 26(19), pp.2458–2459. doi:10.1093/bioinformatics/btq447.

Petrosyan, A. (2015). Onco-Golgi: Is Fragmentation a Gate to Cancer Progression? *Biochemistry & Molecular Biology Journal*, 01(01). doi:10.21767/2471-8084.100006.

Pipaliya, S.V., Schlacht, A., Klinger, C.M., Kahn, R.A. and Dacks, J. (2019). Ancient complement and lineage-specific evolution of the Sec7 ARF GEF proteins in eukaryotes. *Molecular Biology of the Cell*, 30(15), pp.1846–1863. doi:10.1091/mbc.e19-01-0073.

Poon, K.-S. (2021). In silico analysis of BRCA1 and BRCA2 missense variants and the relevance in molecular genetic testing. *Scientific Reports*, 11(1). doi:10.1038/s41598-021-88586-w.

Popper, H. (2020). Primary tumor and metastasis—sectioning the different steps of the metastatic cascade. *Translational Lung Cancer Research*, 9(5), pp.2277–2300. doi:10.21037/tlcr-20-175.

Porta-Pardo, E., Valencia, A. and Godzik, A. (2020). Understanding oncogenicity of cancer driver genes and mutations in the cancer genomics era. *FEBS Letters*. doi:10.1002/1873-3468.13781.

Powell, E., Piwnica-Worms, D. and Piwnica-Worms, H. (2014). Contribution of p53 to Metastasis. *Cancer Discovery*, 4(4), pp.405–414. doi:10.1158/2159-8290.cd-13-0136.

Pożarowska, D. and Pożarowski, P. (2016). The era of anti-vascular endothelial growth factor (VEGF) drugs in ophthalmology, VEGF and anti-VEGF therapy. *Central European Journal of Immunology*, 3, pp.311–316. doi:10.5114/ceji.2016.63132.

Prado, G., Bennett, R.L. and Licht, J.D. (2022). Abstract 3719: Loss of Kmt2c, of the COMPASS-like complex, cooperates with oncogenic signaling in cancer. *Cancer Research*, 82(12_Supplement), pp.3719–3719. doi:10.1158/1538-7445.am2022-3719.

Pretzsch, E., Bösch, F., Neumann, J., Ganschow, P., Bazhin, A., Guba, M., Werner, J. and Angele, M. (2019). Mechanisms of Metastasis in Colorectal Cancer and Metastatic Organotropism: Hematogenous versus Peritoneal Spread. *Journal of Oncology*, [online] 2019, pp.1–13. doi:10.1155/2019/7407190.

Putzke, A.P., Ventura, A.P., Bailey, A.M., Akture, C., Opoku-Ansah, J., Çeliktaş, M., Hwang, M.S., Darling, D.S., Coleman, I.M., Nelson, P.S. and Nguyen, H.M., (2011). Metastatic progression of prostate cancer and e-cadherin: Regulation by Zeb1 and Src family kinases. *The American journal of pathology*, *179*(1), pp.400-410.

Rabbani, B., Tekin, M. and Mahdieh, N. (2013). The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*, 59(1), pp.5–15. doi:10.1038/jhg.2013.114.

Rankin, E.B., Nam, J.-M. and Giaccia, A.J. (2016). Hypoxia: Signaling the Metastatic Cascade. *Trends in Cancer*, [online] 2(6), pp.295–304. doi:10.1016/j.trecan.2016.05.006.

Recasens, A. and Munoz, L. (2019). Targeting Cancer Cell Dormancy. *Trends in Pharmacological Sciences*, [online] 40(2), pp.128–141. doi:10.1016/j.tips.2018.12.004.

Reddy, P., Liu, L., Ren, C., Lindgren, P., Boman, K., Shen, Y., Lundin, E., Ottander, U., Rytinki, M. and Liu, K., (2005). Formation of E-cadherin-mediated cell-cell adhesion activates AKT and mitogen activated protein kinase via phosphatidylinositol 3 kinase and ligand-independent activation of epidermal growth factor receptor in ovarian cancer cells. *Molecular Endocrinology*, *19*(10), pp.2564-2578.

Reinhardt, H.C. and Schumacher, B. (2012). The p53 network: cellular and systemic DNA damage responses in aging and cancer. *Trends in Genetics*, [online] 28(3), pp.128–136. doi:10.1016/j.tig.2011.12.002.

Rentzsch, R. and Orengo, C.A. (2013). Protein function prediction using domain families. *BMC Bioinformatics*, 14(S3). doi:10.1186/1471-2105-14-s3-s5.

Reva, B., Antipin, Y. and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research*, 39(17), pp.e118–e118. doi:10.1093/nar/gkr407.

Ribatti, D., Tamma, R. and Annese, T. (2020). Epithelial-Mesenchymal Transition in Cancer: A Historical Overview. *Translational Oncology*, 13(6), p.100773. doi:10.1016/j.tranon.2020.100773.

Rinaldi, J., Sokol, E.S., Hartmaier, R.J., Trabucco, S.E., Frampton, G.M., Goldberg, M.E., Albacker, L.A., Daemen, A. and Manning, G. (2020). The genomic landscape of metastatic breast cancer: Insights from 11,000 tumors. *PLOS ONE*, 15(5), p.e0231999. doi:10.1371/journal.pone.0231999.

Roca, C. and Adams, R.H. (2007). Regulation of vascular morphogenesis by Notch signaling. *Genes & Development*, 21(20), pp.2511–2524. doi:10.1101/gad.1589207.

Roma-Rodrigues, C., Mendes, R., Baptista, P. and Fernandes, A. (2019). Targeting Tumor Microenvironment for Cancer Therapy. *International Journal of Molecular Sciences*, 20(4), p.840. doi:10.3390/ijms20040840.

Sabit, H., Abdel-Ghany, S., Tombuloglu, H., Cevik, E., Alqosaibi, A., Almulhim, F. and Al-Muhanaa, A. (2021). New insights on CRISPR/Cas9-based therapy for breast Cancer. *Genes and Environment*, 43(1). doi:10.1186/s41021-021-00188-0.

Saftig, P. and Puertollano, R. (2021). How Lysosomes Sense, Integrate, and Cope with Stress. *Trends in Biochemical Sciences*, 46(2), pp.97–112. doi:10.1016/j.tibs.2020.09.004.

Samaržija, I. (2021). Site-Specific and Common Prostate Cancer Metastasis Genes as Suggested by Meta-Analysis of Gene Expression Data. *Life*, 11(7), p.636. doi:10.3390/life11070636.

Sánchez Fernández, I. and Loddenkemper, T. (2014). Subunit Composition of Neurotransmitter Receptors in the Immature and in the Epileptic Brain. *BioMed Research International*, 2014, pp.1–11. doi:10.1155/2014/301950.

Schmeichel, K.L. (2003). Modeling tissue-specific signaling and organ function in three dimensions. *Journal of Cell Science*, 116(12), pp.2377–2388. doi:10.1242/jcs.00503.

Schneider, G., Schmidt-Supprian, M., Rad, R. and Saur, D. (2017). Tissue-specific tumorigenesis – Context matters. *Nature reviews. Cancer*, [online] 17(4), pp.239–253. doi:10.1038/nrc.2017.5.

Schulz, M., Salamero-Boix, A., Niesel, K., Alekseeva, T. and Sevenich, L., 2019. Microenvironmental regulation of tumor progression and therapeutic response in brain metastasis. *Frontiers in immunology*, *10*, p.1713.

Seaby, E.G., Pengelly, R.J. and Ennis, S. (2015). Exome sequencing explained: a practical guide to its clinical application. *Briefings in Functional Genomics*, 15(5), pp.374–384. doi:10.1093/bfgp/elv054.

Sever, R. and Brugge, J.S. (2015). Signal Transduction in Cancer. *Cold Spring Harbor Perspectives in Medicine*, 5(4), pp.a006098–a006098. doi:10.1101/cshperspect.a006098.

Shamir, E.R. and Ewald, A.J. (2014). Three-dimensional organotypic culture: experimental models of mammalian biology and disease. *Nature Reviews Molecular Cell Biology*, 15(10), pp.647–664. doi:10.1038/nrm3873.

Sharma, D., Zagore, L.L., Brister, M.M., Ye, X., Crespo-Hernández, C.E., Licatalosi, D.D. and Jankowsky, E. (2021). The kinetic landscape of an RNA-binding protein in cells. *Nature*, 591(7848), pp.152–156. doi:10.1038/s41586-021-03222-x.

Sharma, P., Alsharif, S., Fallatah, A. and Chung, B.M. (2019a). Intermediate Filaments as Effectors of Cancer Development and Metastasis: A Focus on Keratins, Vimentin, and Nestin. *Cells*, 8(5), p.497. doi:10.3390/cells8050497.

Sharma, Y., Miladi, M., Dukare, S., Boulay, K., Caudron-Herger, M., Groß, M., Backofen, R. and Diederichs, S. (2019b). A pan-cancer analysis of synonymous mutations. *Nature Communications*, [online] 10(1), p.2569. doi:10.1038/s41467-019-10489-2.

Shen, H. and Laird, Peter W. (2013). Interplay between the Cancer Genome and Epigenome. *Cell*, 153(1), pp.38–55. doi:10.1016/j.cell.2013.03.008.

Shi, J., Jiang, D., Yang, S., Sun, Y., Wang, J., Zhang, X., Liu, Y., Lu, Y. and Yang, K. (2020). Molecular profile reveals immune-associated markers of lymphatic invasion in human colon adenocarcinoma. *International Immunopharmacology*, 83, p.106402. doi:10.1016/j.intimp.2020.106402.

Siegel, R.L., Miller, K.D., Fuchs, H.E. and Jemal, A. (2022). Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(1), pp.7–33. doi:10.3322/caac.21708.

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, [online] 40(W1), pp.W452–W457. doi:10.1093/nar/gks539.

Simpkins, F., Jang, K., Yoon, H., Hew, K.E., Kim, M., Azzam, D.J., Sun, J., Zhao, D., Ince, T.A., Liu, W. and Guo, W., 2018. Dual Src and MEK Inhibition Decreases Ovarian Cancer Growth and Targets Tumor Initiating Stem-Like CellsDual Src and Mek Inhibition Targets OVCA Stem Cells. *Clinical cancer research*, *24*(19), pp.4874-4886.

Siveen, K.S., Prabhu, K., Krishnankutty, R., Kuttikrishnan, S., Tsakou, M., Alali, F.Q., Dermime, S., Mohammad, R.M. and Uddin, S. (2017). Vascular Endothelial Growth Factor (VEGF) Signaling in Tumour Vascularization: Potential and Challenges. *Current Vascular Pharmacology*, 15(4). doi:10.2174/1570161115666170105124038.

Soffietti, R., Ahluwalia, M., Lin, N. and Rudà, R. (2020). Management of brain metastases according to molecular subtypes. *Nature Reviews Neurology*, 16(10), pp.557–574. doi:10.1038/s41582-020-0391-x.

Sommariva, M. and Gagliano, N. (2020). E-Cadherin in Pancreatic Ductal Adenocarcinoma: A Multifaceted Actor during EMT. *Cells*, 9(4), p.1040. doi:10.3390/cells9041040.

Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I. and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature Reviews. Cancer*, [online] 18(11), pp.696–705. doi:10.1038/s41568-018-0060-1.

Song, W., Gardner, S.A., Hovhannisyan, H., Natalizio, A., Weymouth, K.S., Chen, W., Thibodeau, I., Bogdanova, E., Letovsky, S., Willis, A. and Nagan, N. (2016). Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification. *Genetics in Medicine*, 18(8), pp.850–854. doi:10.1038/gim.2015.180.

Staub, O. and Rotin, D. (1997). Regulation of ion transport by protein–protein interaction domains. *Current Opinion in Nephrology and Hypertension*, 6(5), pp.447–454. doi:10.1097/00041552-199709000-00007.

Steinbichler, T.B., Dudás, J., Riechelmann, H. and Skvortsova, I.-I. (2017). The role of exosomes in cancer metastasis. *Seminars in Cancer Biology*, 44, pp.170–181. doi:10.1016/j.semcancer.2017.02.006.

Strom, S.P., Lee, H., Das, K., Vilain, E., Nelson, S.F., Grody, W.W. and Deignan, J.L. (2014). Assessing the necessity of confirmatory testing for exome-sequencing

results in a clinical molecular diagnostic laboratory. *Genetics in Medicine*, 16(7), pp.510–515. doi:10.1038/gim.2013.183.

Strouhalova, K., Přechová, M., Gandalovičová, A., Brábek, J., Gregor, M. and Rosel, D. (2020). Vimentin Intermediate Filaments as Potential Target for Cancer Treatment. *Cancers*, 12(1), p.184. doi:10.3390/cancers12010184.

Suchting, S., Freitas, C., Toro, R. del, Noble, F., Benedito, R., Breant, C., Duarte, A. and Eichmann, A. (2007). The Notch ligand Delta-like 4 (Dll4) negatively regulates endothelial tip cell formation and vessel branching. *The FASEB Journal*, 21(5). doi:10.1096/fasebj.21.5.a15-a.

Suh, J.H., Kotecha, R., Chao, S.T., Ahluwalia, M.S., Sahgal, A. and Chang, E.L., 2020. Current approaches to the management of brain metastases. *Nature reviews Clinical oncology*, *17*(5), pp.279-299.

Sveinbjornsson, G., Albrechtsen, A., Zink, F., Gudjonsson, S.A., Oddson, A., Másson, G., Holm, H., Kong, A., Thorsteinsdottir, U., Sulem, P., Gudbjartsson, D.F. and Stefansson, K. (2016). Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nature Genetics*, 48(3), pp.314–317. doi:10.1038/ng.3507.

Szpiech, Z.A., Strauli, N.B., White, K.A., Ruiz, D.G., Jacobson, M.P., Barber, D.L. and Hernandez, R.D. (2017). Prominent features of the amino acid mutation landscape in cancer. *PLOS ONE*, 12(8), p.e0183273. doi:10.1371/journal.pone.0183273.

Tang, D., Yuan, H. and Wang, Y. (2010). The Role of GRASP65 in Golgi Cisternal Stacking and Cell Cycle Progression. *Traffic*, 11(6), pp.827–842. doi:10.1111/j.1600-0854.2010.01055.x.

Tang, J., Kong, D., Cui, Q., Wang, K., Zhang, D., Gong, Y. and Wu, G. (2018). Prognostic Genes of Breast Cancer Identified by Gene Co-expression Network Analysis. *Frontiers in Oncology*, 8. doi:10.3389/fonc.2018.00374.

Tang, M., Xie, Q., Gimple, R.C., Zhong, Z., Tam, T., Tian, J., Kidwell, R.L., Wu, Q., Prager, B.C., Qiu, Z., Yu, A., Zhu, Z., Mesci, P., Jing, H., Schimelman, J., Wang,

P., Lee, D., Lorenzini, M.H., Dixit, D. and Zhao, L. (2020a). Three-dimensional bioprinted glioblastoma microenvironments model cellular dependencies and immune interactions. *Cell Research*, [online] 30(10), pp.833–853. doi:10.1038/s41422-020-0338-1.

Tang, T., Yang, Z., Wang, D., Yang, X., Wang, J., Li, L., Wen, Q., Gao, L., Bian, X. and Yu, S. (2020b). The role of lysosomes in cancer development and progression. *Cell & Bioscience*, 10(1). doi:10.1186/s13578-020-00489-x.

Tan, H., Bao, J. and Zhou, X. (2015). Genome-wide mutational spectra analysis reveals significant cancer-specific heterogeneity. *Scientific Reports*, 5(1). doi:10.1038/srep12566.

Tao, Y., Lei, H., Lee, A.V., Ma, J. and Schwartz, R., 2020. Neural network deconvolution method for resolving pathway-level progression of tumor clonal expression programs with application to breast cancer brain metastases. *Frontiers in physiology*, *11*, p.1055.

Tashima, T., 2022. Brain Cancer Chemotherapy through a Delivery System across the Blood-Brain Barrier into the Brain Based on Receptor-Mediated Transcytosis Using Monoclonal Antibody Conjugates. *Biomedicines*, *10*(7), p.1597.

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S.C., Kok, C.Y., Noble, K., Ponting, L., Ramshaw, C.C., Rye, C.E. and Speedy, H.E. (2018). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, [online] 47(D1), pp.D941–D947. doi:10.1093/nar/gky1015.

Teleanu, R.I., Chircov, C., Grumezescu, A.M. and Teleanu, D.M. (2019). Tumor Angiogenesis and Anti-Angiogenic Strategies for Cancer Treatment. *Journal of Clinical Medicine*, 9(1), p.84. doi:10.3390/jcm9010084.

Tian, W., Liu, S. and Li, B. (2019). Potential Role of Exosomes in Cancer Metastasis. *BioMed Research International*, 2019, pp.1–12. doi:10.1155/2019/4649705.

Tomasini, P., Barlesi, F., Gilles, S., Nanni-Metellus, I., Soffietti, R., Denicolai, E., Pellegrino, E., Bialecki, E., Ouafik, L. and Metellus, P. (2020). Comparative genomic analysis of primary tumors and paired brain metastases in lung cancer patients by whole exome sequencing: a pilot study. *Oncotarget*, 11(50), pp.4648–4654. doi:10.18632/oncotarget.27837.

Truelsen, S.L.B., Mousavi, N., Wei, H., Harvey, L., Stausholm, R., Spillum, E., Hagel, G., Qvortrup, K., Thastrup, O., Harling, H., Mellor, H. and Thastrup, J. (2021). The cancer angiogenesis co-culture assay: In vitro quantification of the angiogenic potential of tumoroids. *PLOS ONE*, 16(7), p.e0253258. doi:10.1371/journal.pone.0253258.

Tsuber, V., Kadamov, Y., Brautigam, L., Berglund, U.W. and Helleday, T. (2017). Mutations in Cancer Cause Gain of Cysteine, Histidine, and Tryptophan at the Expense of a Net Loss of Arginine on the Proteome Level. *Biomolecules*, 7(4), p.49. doi:10.3390/biom7030049.

Turashvili, G. and Brogi, E. (2017). Tumor Heterogeneity in Breast Cancer. *Frontiers in Medicine*, [online] 4. doi:10.3389/fmed.2017.00227.

Turner, K.M., Yeo, S.K., Holm, T.M., Shaughnessy, E. and Guan, J.-L. (2021). Heterogeneity within molecular subtypes of breast cancer. *American Journal of Physiology-Cell Physiology*, 321(2), pp.C343–C354. doi:10.1152/ajpcell.00109.2021.

Tyran, M., Carbuccia, N., Garnier, S., Guille, A., Adelaïde, J., Finetti, P., Touzlian, J., Viens, P., Tallet, A., Goncalves, A., Metellus, P., Birnbaum, D., Chaffanet, M. and Bertucci, F. (2019). A Comparison of DNA Mutation and Copy Number Profiles of Primary Breast Cancers and Paired Brain Metastases for Identifying Clinically Relevant Genetic Alterations in Brain Metastases. *Cancers*, 11(5), p.665. doi:10.3390/cancers11050665.

Uniprot.org. (2019). *UniProt*. [online] Available at: https://www.uniprot.org.

Vander Heiden, M.G., Cantley, L.C. and Thompson, C.B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science*, 324(5930), pp.1029–1033. doi:10.1126/science.1160809.

Van Hoeck, A., Tjoonk, N.H., van Boxtel, R. and Cuppen, E. (2019). Portrait of a cancer: mutational signature analyses for cancer diagnostics. *BMC Cancer*, 19(1). doi:10.1186/s12885-019-5677-2.

Van Roy, F., (2014). Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nature Reviews Cancer, 14*(2), pp.121-134.

Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. and Ng, P.C. (2015). SIFT missense predictions for genomes. *Nature Protocols*, 11(1), pp.1–9. doi:10.1038/nprot.2015.123.

Venkataramani, V., Tanev, D.I., Strahle, C., Studier-Fischer, A., Fankhauser, L., Kessler, T., Körber, C., Kardorff, M., Ratliff, M., Xie, R., Horstmann, H., Messer, M., Paik, S.P., Knabbe, J., Sahm, F., Kurz, F.T., Acikgöz, A.A., Herrmannsdörfer, F., Agarwal, A. and Bergles, D.E. (2019). Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature*, 573(7775), pp.532–538. doi:10.1038/s41586-019-1564-x.

Venkatesh, H. and Monje, M., 2017. Neuronal activity in ontogeny and oncology. *Trends in Cancer*, *3*(2), pp.89-112.

Venkatesh, H.S., Morishita, W., Geraghty, A.C., Silverbush, D., Gillespie, S.M., Arzt, M., Tam, L.T., Espenel, C., Ponnuswami, A., Ni, L., Woo, P.J., Taylor, K.R., Agarwal, A., Regev, A., Brang, D., Vogel, H., Hervey-Jumper, S., Bergles, D.E., Suvà, M.L. and Malenka, R.C. (2019). Electrical and synaptic integration of glioma into neural circuits. *Nature*, [online] 573(7775), pp.539–545. doi:10.1038/s41586-019-1563-y.

Vettore, L., Westbrook, R.L. and Tennant, D.A. (2019). New aspects of amino acid metabolism in cancer. *British Journal of Cancer*, 122(2), pp.150–156. doi:10.1038/s41416-019-0620-5.

Villa, E., Ali, E.S., Sahu, U. and Ben-Sahra, I. (2019). Cancer Cells Tune the Signaling Pathways to Empower de Novo Synthesis of Nucleotides. *Cancers*, [online] 11(5). doi:10.3390/cancers11050688.

Waks, A.G. and Winer, E.P. (2019). Breast Cancer Treatment. *JAMA*, [online] 321(3), p.288. doi:10.1001/jama.2018.19323.

Wallis, C.J.D. and Nam, R.K. (2015). Prostate Cancer Genetics: A Review. *EJIFCC*, [online] 26(2), pp.79–91. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4975354/.

Wang, J., Yang, L., Liang, F., Chen, Y. and Yang, G. (2018). Integrin alpha x stimulates cancer angiogenesis through PI3K/Akt signaling–mediated VEGFR2/VEGF-A overexpression in blood vessel endothelial cells. *Journal of Cellular Biochemistry*, 120(2), pp.1807–1818. doi:10.1002/jcb.27480.

Wang, R., Xu, Y., Niu, C., Gao, X. and Xu, X. (2021a). A Novel Small Peptide H-KI20 Inhibits Retinal Neovascularization Through the JNK/ATF2 Signaling Pathway. *Investigative Opthalmology & Visual Science*, 62(1), p.16. doi:10.1167/iovs.62.1.16.

Wang, Y., Zhang, H., Zhong, H. and Xue, Z. (2021b). Protein domain identification methods and online resources. *Computational and Structural Biotechnology Journal*, [online] 19, pp.1145–1153. doi:10.1016/j.csbj.2021.01.041.

Wang, Y., Zou, S., Zhao, Z., Liu, P., Ke, C. and Xu, S. (2020). New insights into small-cell lung cancer development and therapy. *Cell Biology International*, [online] 44(8), pp.1564–1576. doi:10.1002/cbin.11359.

Ward, C., Meehan, J., Gray, M.E., Murray, A.F., Argyle, D.J., Kunkler, I.H. and Langdon, S.P. (2020). The impact of tumour pH on cancer progression: strategies for clinical intervention. *Exploration of Targeted Anti-tumor Therapy*, 1(2), pp.71–100. doi:10.37349/etat.2020.00005.

Watkins, E.J. (2019). Overview of breast cancer. *Journal of the American Academy of Physician Assistants*, [online] 32(10), p.1. doi:10.1097/01.jaa.0000580524.95733.3d.

Watson, G., Ronai, Z.A. and Lau, E. (2017). ATF2, a paradigm of the multifaceted regulation of transcription factors in biology and disease. *Pharmacological Research*, 119, pp.347–357. doi:10.1016/j.phrs.2017.02.004.

Wei, P., Liu, X. and Fu, Y.-X. (2011). Incorporating predicted functions of nonsynonymous variants into gene-based analysis of exome sequencing data: a comparative study. *BMC Proceedings*, 5(S9). doi:10.1186/1753-6561-5-s9-s20.

White, K.A., Grillo-Hill, B.K. and Barber, D.L. (2017). Cancer cell behaviors mediated by dysregulated pH dynamics at a glance. *Journal of Cell Science*, [online] 130(4), pp.663–669. doi:10.1242/jcs.195297.

White, K.A., Ruiz, D.G., Szpiech, Z.A., Strauli, N.B., Hernandez, R.D., Jacobson, M.P. and Barber, D.L. (2017). Cancer-associated arginine-to-histidine mutations confer a gain in pH sensing to mutant proteins. *Science Signaling*, [online] 10(495), p.eaam9931. doi:10.1126/scisignal.aam9931.

Witzel, I., Oliveira-Ferrer, L., Pantel, K., Müller, V. and Wikman, H. (2016). Breast cancer brain metastases: biology and new clinical perspectives. *Breast Cancer Research*, [online] 18(1). doi:10.1186/s13058-015-0665-1.

Woditschka, S., Evans, L., Duchnowska, R., Reed, L.T., Palmieri, D., Qian, Y., Badve, S., Sledge, G., Gril, B., Aladjem, M.I., Fu, H., Flores, N.M., Gökmen-Polar, Y., Biernat, W., Szutowicz-Zielińska, E., Mandat, T., Trojanowski, T., Och, W., Czartoryska-Arlukowicz, B. and Jassem, J. (2014). DNA Double-Strand Break Repair Genes and Oxidative Damage in Brain Metastasis of Breast Cancer. *JNCI: Journal of the National Cancer Institute*, 106(7). doi:10.1093/jnci/dju145.

World Health Organization (2020). *Cancer today*. [online] Iarc.fr. Available at: https://gco.iarc.fr/today/home.

Wu, J., Gao, F., Wang, C., Qin, M., Han, F., Xu, T., Hu, Z., Long, Y., He, X., Deng, X., Ren, D. and Dai, T. (2019). IL-6 and IL-8 secreted by tumour cells impair the function of NK cells via the STAT3 pathway in oesophageal squamous cell carcinoma. *Journal of Experimental & Clinical Cancer Research*, 38(1). doi:10.1186/s13046-019-1310-0.

Wu, N. and Yu, H. (2012). The Smc complexes in DNA damage response. *Cell & Bioscience*, 2(1), p.5. doi:10.1186/2045-3701-2-5.

Wu, Y., Sarkissyan, M. and Vadgama, J. (2016). Epithelial-Mesenchymal Transition and Breast Cancer. *Journal of Clinical Medicine*, [online] 5(2), p.13. doi:10.3390/jcm5020013.

www.ebi.ac.uk. (n.d.). *InterPro*. [online] Available at: https://www.ebi.ac.uk/interpro/result/InterProScan/iprscan5-R20221020-115708-0181-88349756-p2m/ [Accessed 20 Oct. 2022].

www.mutationaligner.org. (n.d.). *MutataionAligner*. [online] Available at: http://www.mutationaligner.org [Accessed 24 Oct. 2022].

www.ncbi.nlm.nih.gov. (n.d.). *Home - Conserved Domains - NCBI*. [online] Available at: https://www.ncbi.nlm.nih.gov/cdd/.

www.ncbi.nlm.nih.gov. (n.d.). *Welcome to NCBI Domain architecture search*. [online] Available at: https://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=prot&uid=1 813836565 [Accessed 5 May 2022].

Xiao, L., Rao, J.N., Zou, T., Liu, L., Yu, T.-X., Zhu, X.-Y., Donahue, J.M. and Wang, J.-Y. (2010). Induced ATF-2 represses CDK4 transcription through dimerization with JunD inhibiting intestinal epithelial cell growth after polyamine depletion. *American Journal of Physiology-Cell Physiology*, 298(5), pp.C1226–C1234. doi:10.1152/ajpcell.00021.2010.

Xiao, R., Chen, J.-Y., Liang, Z., Luo, D., Chen, G., Lu, Z.J., Chen, Y., Zhou, B., Li, H., Du, X., Yang, Y., San, M., Wei, X., Liu, W., Lécuyer, E., Graveley, B.R., Yeo, G.W., Burge, C.B., Zhang, M.Q. and Zhou, Y. (2019). Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell*, 178(1), pp.107-121.e18. doi:10.1016/j.cell.2019.06.001.

Xie, K., Peng, Y., Zhong, W. and Liu, X. (2022). KMT2C is a Potential Biomarker of Anti-PD-1 Treatment Response in Metastatic Melanoma. *Frontiers in Bioscience-Landmark*, 27(3), p.0103. doi:10.31083/j.fbl2703103.

Yadav, L. (2015). Tumour Angiogenesis and Angiogenic Inhibitors: A Review. *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH.* doi:10.7860/jcdr/2015/12016.6135.

Yakan, S., Sari, E., Erkan, N., Yildirim, M., Vardar, E., Coskun, A., Cetin, D.A. and Eliyatkin, N. (2014). Breast Carcinosarcomas. *The Journal of Breast Health*, 10(3), pp.161–165. doi:10.5152/tjbh.2014.2197.

Yamasaki, J., Hirata, Y., Otsuki, Y., Suina, K., Saito, Y., Masuda, K., Okazaki, S., Ishimoto, T., Saya, H. and Nagano, O., 2022. MEK inhibition suppresses metastatic progression of KRAS-mutated gastric cancer. *Cancer science*, *113*(3), p.916.

Yang, F., Petsalaki, E., Rolland, T., Hill, D.E., Vidal, M. and Roth, F.P. (2015). Protein Domain-Level Landscape of Cancer-Type-Specific Somatic Mutations. *PLoS Computational Biology*, [online] 11(3), p.e1004147. doi:10.1371/journal.pcbi.1004147.

Yang, Q., Zhu, C., Zhang, Y., Wang, Y., Wang, Y., Zhu, L., Yang, X., Li, J., Nie, H., Jiang, S., Zhang, X., Cao, X., Li, Q., Zhang, X., Tian, G., Hu, L., Zhu, L., Zhao, G. and Zhang, Z. (2018). Molecular analysis of gastric cancer identifies genomic markers of drug sensitivity in Asian gastric cancer. *Journal of Cancer*, 9(16), pp.2973–2980. doi:10.7150/jca.25506.

Yang, Z., Chen, Y., Wu, D., Min, Z. and Quan, Y. (2019). Analysis of risk factors for colon cancer progression. *OncoTargets and Therapy*, Volume 12, pp.3991–4000. doi:10.2147/ott.s207390.

Yardley, D.A., Shipley, D., Zubkus, J., Wright, G.L., Ward, P.J., Mani, A., Shastry, M., Finney, L., DeBusk, L. and Hainsworth, J.D. (2019). A Randomized Phase II Study of Eribulin/Cyclophosphamide or Docetaxel/Cyclophosphamide as Neoadjuvant Therapy in Operable HER2-negative Breast Cancer. *Clinical Breast Cancer*, 19(1), pp.1–9. doi:10.1016/j.clbc.2018.08.006.

Yates, L.R., Knappskog, S., Wedge, D., Farmery, J.H.R., Gonzalez, S., Martincorena, I., Alexandrov, L.B., Van Loo, P., Haugland, H.K., Lilleng, P.K., Gundem, G.,

Gerstung, M., Pappaemmanuil, E., Gazinska, P., Bhosle, S.G., Jones, D., Raine, K., Mudie, L. and Campbell, P.J. (2017). Genomic evolution of breast cancer metastasis and relapse. *Cancer Cell*, [online] 32(2), pp.169–184. doi:https://doi.org/10.1016/j.ccell.2017.07.005.

Yoshimaru, T., Komatsu, M., Miyoshi, Y., Honda, J., Sasa, M. and Katagiri, T. (2015). Therapeutic advances in BIG 3- PHB 2 inhibition targeting the crosstalk between estrogen and growth factors in breast cancer. *Cancer Science*, 106(5), pp.550–558. doi:10.1111/cas.12654.

Yoshimaru, T., Komatsu, M., Tashiro, E., Imoto, M., Osada, H., Miyoshi, Y., Honda, J., Sasa, M. and Katagiri, T. (2014). Xanthohumol suppresses oestrogen-signalling in breast cancer through the inhibition of BIG3-PHB2 interactions. *Scientific Reports*, 4(1). doi:10.1038/srep07355.

Yoshimaru, T., Nakamura, Y. and Katagiri, T. (2021). Functional genomics for breast cancer drug target discovery. *Journal of Human Genetics*, 66(9), pp.927–935. doi:10.1038/s10038-021-00962-6.

Zahra, F.T., Sajib, M.S. and Mikelis, C.M. (2021). Role of bFGF in Acquired Resistance upon Anti-VEGF Therapy in Cancer. *Cancers*, [online] 13(6), p.1422. doi:10.3390/cancers13061422.

Zardavas, D., Phillips, W.A. and Loi, S. (2014). PIK3CA mutations in breast cancer: reconciling findings from preclinical and clinical data. *Breast Cancer Research*, 16(1). doi:10.1186/bcr3605.

Zeng, Q., Michael, I.P., Zhang, P., Saghafinia, S., Knott, G., Jiao, W., McCabe, B.D., Galván, J.A., Robinson, H.P.C., Zlobec, I., Ciriello, G. and Hanahan, D. (2019). Synaptic proximity enables NMDAR signalling to promote brain metastasis. *Nature*, [online] 573(7775), pp.526–531. doi:10.1038/s41586-019-1576-6.

Zhang, C.-H., Wang, Z.-B., Quan, S., Huang, X., Tong, J.-S., Ma, J.-Y., Guo, L., Wei, Y., Ouyang, Y.-C., Hou, Y., Xing, F.-Q. and Sun, Q.-Y. (2011). GM130, a cis-Golgi protein, regulates meiotic spindle assembly and asymmetric division in mouse oocyte. *Cell Cycle*, 10(11), pp.1861–1870. doi:10.4161/cc.10.11.15797.

Zhang, C., Zhang, R., Dai, X., Cao, X., Wang, K., Huang, X. and Ren, Q. (2020a). Activating transcription factor 2 (ATF2) negatively regulates the expression of antimicrobial peptide genes through tumor necrosis factor (TNF) in Macrobrachium nipponense. *Fish & Shellfish Immunology*, 107, pp.26–35. doi:10.1016/j.fsi.2020.09.043.

Zhang, J., Dominguez-Sola, D., Hussein, S., Lee, J.-E., Holmes, A.B., Bansal, M., Vlasevska, S., Mo, T., Tang, H., Basso, K., Ge, K., Dalla-Favera, R. and Pasqualucci, L. (2015). Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nature Medicine*, 21(10), pp.1190–1198. doi:10.1038/nm.3940.

Zhang, J., Späth, S.S., Marjani, S.L., Zhang, W. and Pan, X. (2018). Characterization of cancer genomic heterogeneity by next-generation sequencing advances precision medicine in cancer treatment. *Precision Clinical Medicine*, 1(1), pp.29–48. doi:10.1093/pcmedi/pby007.

Zhang, L., Lu, Q. and Chang, C. (2020). Epigenetics in Health and Disease. *Advances in Experimental Medicine and Biology*, pp.3–55. doi:10.1007/978-981-15-3449-2_1.

Zhang, L., Zhang, Q., Hinojosa, D.T., Jiang, K., Pham, Q.-K., Xiao, Z., Colvin, V.L. and Bao, G. (2022). Multifunctional Magnetic Nanoclusters Can Induce Immunogenic Cell Death and Suppress Tumor Recurrence and Metastasis. *ACS Nano.* doi:10.1021/acsnano.2c06776.

Zhang, Y. and Weinberg, R.A. (2018). Epithelial-to-mesenchymal transition in cancer: complexity and opportunities. *Frontiers of Medicine*, [online] 12(4), pp.361–373. doi:10.1007/s11684-018-0656-6.

Zhang, Z., Zhou, C., Li, X., Sawyers, C. and Mu, P. (2020b). Abstract PO-117: CHD1-loss promotes tumor heterogeneity and therapy resistance in prostate cancer. *Cancer Research*, 80(21_Supplement), p.PO-117-PO-117. doi:10.1158/1538-7445.tumhet2020-po-117.

Zheng, B., Song, Z., Chen, Y. and Yan, W. (2021). Genomic Analyses of Metaplastic or Sarcomatoid Carcinomas From Different Organs Revealed Frequent Mutations in KMT2D. *Frontiers in Molecular Biosciences*, 8. doi:10.3389/fmolb.2021.688692.

Zheng, J., Zhang, H., Banerjee, S., Li, Y., Zhou, J., Yang, Q., Tan, X., Han, P., Fu, Q., Cui, X., Yuan, Y., Zhang, M., Shen, R., Song, H., Zhang, X., Zhao, L., Peng, Z., Wang, W. and Yin, Y. (2019a). A comprehensive assessment of Next-Generation Sequencing variants validation using a secondary technology. *Molecular Genetics & Genomic Medicine*, 7(7). doi:10.1002/mgg3.748.

Zheng, S., West, J.J., Yu, C.G. and Harris, T.J.C. (2019b). Arf-GEF localization and function at myosin-rich adherens junctions via coiled-coil heterodimerization with an adaptor protein. *Molecular Biology of the Cell*, 30(26), pp.3090–3103. doi:10.1091/mbc.e19-10-0566.

Zheng, Y. and Cui, Q. (2016). Microscopic mechanisms that govern the titration response and pK_a values of buried residues in staphylococcal nuclease mutants. *Proteins: Structure, Function, and Bioinformatics*, 85(2), pp.268–281. doi:10.1002/prot.25213.

Zimmer, A.S., Van Swearingen, A.E.D. and Anders, C.K. (2020). HER2-positive breast cancer brain metastasis: A new and exciting landscape. *Cancer Reports*. doi:10.1002/cnr2.1274.

List of publications

Conference abstracts

Olivares, I., Kalyanakrishnan, K., Ahmed, S., Murcott, C., Wilkinson, R.N., Cotton, J., Breitwieser, W., Morris, M.R., Armesilla, A. (2019).
"Activating transcription factor ATF2 negatively regulates the expression of endothelial notch ligands." BMJ Journals;105: A167/BS45.

Olivares, I., Pangeni, R.P., Huen, D., Dawson, T.P., Ashton, K., Davis, Charles.H.G., Jenkinson, M.D., Brodbelt, A.R., Wang, W., Darling, J.L., Warr, T.J. and Morris, M.R. (2018).
"Identifying epigenetic changes in breast tumours that metastasise to the brain". European Journal of Surgical Oncology, 44, p.S36. doi:10.1016/j.ejso.2018.01.567.

Second author peer review paper

Pangeni, R.P., **Olivares, I.**, Huen, D., Buzatto, V.C., Dawson, T.P., Ashton, K.M., Davis, C., Brodbelt, A.R., Jenkinson, M.D., Bièche, I., Yang, L., Latif, F., Darling, J.L., Warr, T.J. and Morris, M.R. (2022).
"Genome-wide methylation analyses identifies Non-coding RNA genes dysregulated in breast tumours that metastasise to the brain." *Scientific Reports*, 12(1). doi:10.1038/s41598-022-05050-z.